Applying Bayesian Statistical Methods to Optimise Processes Within Additive Manufacturing

The University Of Sheffield.

Adam Gothorp

This thesis is submitted for the degree of
Doctor of Philosophy

School of Mathematics and Statistics
Faculty of Science
University of Sheffield

April 2023

Acknowledgements

I would firstly like to acknowledge my supervisors Eleanor Stillman, Paul Blackwell and Candice Majewski. They are all very supportive, understanding, and most importantly, very kind. It has been enjoyable to work with you all for the past 4 years. In particular, I would like to thank Eleanor for all her support over my 8(!) years in Sheffield.

I'd also like to mention my housemates: Alan, James and Dan. With half of my PhD taking place during lockdown, I am very grateful for being able to have such lovely people to be stuck with in a house! Cooking Christmas dinners together and watching 12 straight hours of football on the TV made for peak lockdown fun.

I appreciate all of my friends from back home in Lincoln; while there are too many to name, you've all provided endless fun for such a long time (22 years and counting!).

Finally, I want to thank my family: my mum, Tracey, and my brother, Josh, and also my amazing partner, Acushla. I greatly appreciate your support, patience, love, kindness and empathy.

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

It is noted that some of the content of the statistical methods (Chapter 3) had been previously developed by Phillip Paine, a former research assistant at the University of Sheffield. In particular, he provided a transfer of knowledge at the start of the PhD regarding multivariate Bayesian regression (both 'forward' and 'backward' modelling), along with code for the statistical software R and a simulation example.

Moreover, we (the supervisors and I) intend to publish a paper in an Additive manufacturing journal based on this work.

Abstract

Additive Manufacturing (AM), otherwise known as 3D printing, is the process of building 3D objects based on some 3D model, typically provided by some computer-aided design. AM comprises multiple methodologies which can fabricate 3D objects with several different materials. While AM has been a fast-developing field for several years, there are notable gaps in the research for applications of statistically sound methods that are able to answer a wide variety of important and complex questions. Due to savings on cost and time, these methods can be highly beneficial.

This works provides a robust process for estimating relationships between multiple output variables and input variables simultaneously with a 'forward' model, which accounts for measurement error in the data and incorporates expert opinion into the modelling. This expert opinion, alongside the available data, provides a more complete understanding of the relationship to be estimated.

The ultimate aim of this work is the optimisation of the input variables in order to produce desired values for multiple output variables. With the use of Bayesian statistics, these methods provide an intuitive process for this optimisation by inverting the fitted relationship identified in the forward modelling.

Two modelling methods are demonstrated: errors-in-variables Bayesian regression, and errors-in-variables Gaussian processes. The former is a 'parametric' method, which estimates model parameters based on a fixed, predetermined relationship. The latter is a 'nonparametric' method, which does not assume such a relationship, instead assuming a multivariate normal distribution for the output variables, whose covariance is informed by spatial correlation between the input variables.

The modelling process for these statistical methods is developed with simulated data, followed by an application to a real-data example, looking at optimisation of powder properties to provide ideal powder flow and powder bed deposition for improving laser sintering.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The fields of engineering and statistics are becoming more and more intertwined; as more technologies in various engineering fields are developed, more questions are available to be answered by a multitude of statistical methods. More specifically, the field of Additive Manufacturing (AM) remains relatively new and unresearched, leading to numerous possibilities for statistical research to help develop the field. Some examples are:

- quality analysis; the mechanical properties of a 3D-printed object are dependent upon many variables, such as the material properties and the printer settings. Exactly how these variables are related could be found using statistical methods, in order to improve these mechanical properties

- reusing excess powder; after having utilised powder for printing a 3D object, the excess powder is often considered to be unusable and is wasted. Could the properties of a powder, relating to particle size and density (among others), be optimised for the purpose of powder reuse and waste reduction?

The work in this thesis provides a template for approaching questions within AM in a statistically sound manner, to provide answers that account for uncertainty in the estimates of model parameters. Moreover, this work is an example of true interdisciplinary research, with state-of-the-art statistical methods being applied to cutting-edge AM, while also developing the statistical methodology through the exposure to challenging practical problems. The case study to which these methods are applied in this work is a specific example from powdered-polymer AM, where further details are given

in Section 2.2, although the statistical modelling discussed here is applicable across a broad range of applications.

The approach in this work is to answer questions through statistical modelling. The main aim of a statistical model is to relate some response variable to some explanatory variable while describing the variation in the response variable, with appropriate assumptions about how the variables are distributed (in other words, what is the influence of the explanatory variable on the response variable, and how confident is the model in this influence). The two main questions of interest in this work are:

1. Which statistical model best describes the relationship between the response (output) variable(s) and the explanatory (input) variable(s) (i.e., describing the effect of the input variable on the output variable)?

2. Given this statistical model, what values of the explanatory variable(s) are required in order to produce some desired value for the response variable(s)?

Note firstly that, oftentimes there are multiple response variables and multiple explanatory variables considered simultaneously, which provides a high level of complexity to the modelling. Secondly, that the first question implies considering many statistical models, and comparing which of these performs 'best' (best can be defined statistically in many ways, and is discussed further in Chapter 3). The answer to the first question is valuable in helping to understand how these input variable(s) influence the output variable(s).

The second question above is the ultimate goal of this work, and is motivated by a simple example. Suppose there is data available for a response variable $\boldsymbol{Y}$ and explanatory variable x, where the data point x_i represents the laser scanning speed for a Laser Sintering machine for observation i, and Y_i represents the corresponding tensile strength measurements of a 3D-printed object from the Laser Sintering machine. The effect of the laser scanning speed on the tensile strength of the final part can then be estimated with a statistical model, specifically, a *classical* linear regression model. Suppose further that a *simple linear model* is used, which assumes a straight line relationship between the data points of the two variables. Mathematically, this can be written as

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \tag{1.1}$$

where:

- the observed data points for the response variable are denoted by Y_i,

- the observed data points for the explanatory variable are denoted by x_i,
- the term ϵ_i denotes some discrepancy between the observed data point Y_i and the value of the straight line at the observed data point x_i, with the discrepancy assumed to be random error,
- the subscript i denotes the i^{th} observation of the response variable and the explanatory variable, where there are n observed data points $Y_1, \ldots, Y_n$ corresponding to $x_1, \ldots, x_n$.

In this classical setting, the model coefficients β_0 (the intercept of the straight line) and β_1 (the slope) can be estimated using the method of least squares, or using any preferred method. Note here that the model error term ϵ_i is a random variable, thus inducing uncertainty with respect to the response data.

Now suppose a particular desired value of the response variable tensile strength, say Y^*, is of interest, and a corresponding value x^* of the explanatory variable laser scanning speed (i.e., a value of the explanatory variable that can produce the desired response value) is to be found. How should one proceed? Consider the plot in Figure 1.1—this provides some example data of a response variable $\boldsymbol{Y}$ against an explanatory variable $\boldsymbol{x}$, with the 'line of best fit' (i.e., the straight line using the estimates of β_0 and β_1 noted above) from the simple linear model given as the solid black line. The solid red line lies at some desired value of the response, say $Y^* = y^*$. An intuitive thought would be to use the intersection of the desired response value with the line of best fit, and read down to the corresponding value x^* on the x-axis, to provide the corresponding input value. This can be considered a legitimate solution to this problem, but not without complications; how is uncertainty factored into this 'estimate' of the corresponding value? Is this ideologically correct, to predict the behaviour of the explanatory variable given the response variable, having estimated the relationship between the two variables with uncertainty on the response variable and not the input variable? Moreover, an alternative solution can be provided by switching the response variable and explanatory variable and fitting a simple linear model with $\boldsymbol{X}$ as the response variable and $\boldsymbol{y}$ as the explanatory variable (capitalisations reversed here—see Section 3.2). The prediction of the corresponding input value x^* (having treated the observed data $\boldsymbol{X}$ as the 'response' variable) is then given by the *fitted value* (the value of the response variable given an observed data point for the explanatory variable) at the desired response value y^* (having treated the observed data for $\boldsymbol{y}$ as the 'explanatory' vari-

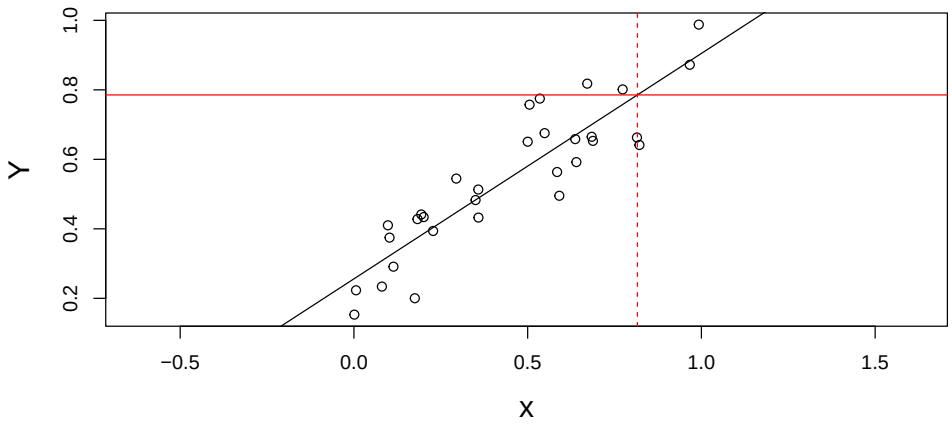

Figure 1.1: Line of best fit relating some response variable and explanatory variable (simulated data), with some desired response $Y^* = y^*$ indicated by the solid red line and a corresponding value of the explanatory variable indicated by the dotted red line $x = x^*$.

able), which is calculated to be the dotted green line in Figure 1.2. Two estimates of the corresponding input value now exist, given the dotted red line and dotted green line—which is most justified? Which is the preferred estimate?

The problem is complicated further by considering more response variables and explanatory variables simultaneously in the modelling, for which a single solution to finding optimal vectors for the corresponding input vector to some desired response vector is not possible in a classical setting. For example, suppose that a linear model with two explanatory variables and one response variable is of interest, e.g.,

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i, \tag{1.2}$$

where $x_{1,i}$ and $x_{2,i}$ represent the i^{th} observation of the first and second explanatory variables respectively. If one estimates the model coefficients, the 'fitted' model is represented by a plane in 3-dimensional space (analogous to the 'line of best fit' from the simple linear model, so we have here a 'plane of best fit'). Suppose again a desired value of the response variable $Y^* = y^*$ is of interest; in this case, the line

$$X_2^* = y^* - \hat{\beta}_0 - \hat{\beta}_1 X_1^*,$$

where X_1^* can take any real number, provides an infinite number of solutions for the two explanatory variables. This work implements Bayesian statistical

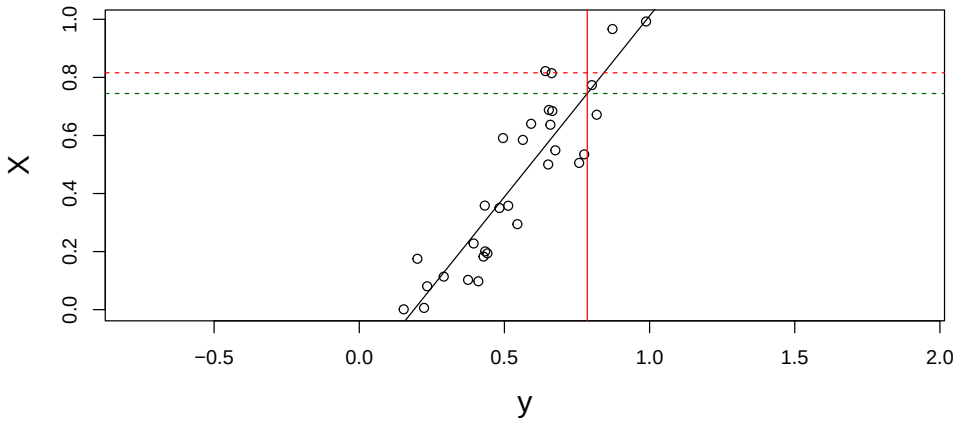

Figure 1.2: Line of best fit for a simple linear model, with the explanatory variable data and response variable data from Figure 1.1 switched around (in terms of both the linear model, and the axes in the plot). This provides a different estimate of $X = x^*$ given the desired 'response' value $y = y^*$.

methods, which allow a single solution to this problem to be found. These methods are discussed in Chapter 3, and implemented on simulated data and real-world data in Chapter 5.

As mentioned above, this work implements a statistical modelling approach, which has many benefits. Firstly, the theory behind the statistical methods is well researched, with many examples of practical applications. Secondly, these methods are versatile, as they are not bound to any specific modelling scenario, and can be appropriately utilised for many applications (with some examples given earlier). The work in this PhD could be transferred over to several aspects of AM, with respect to any AM methodology, or any component within a methodology. This is in contrast to physical models, which can be effective, but are restricted to the application for which they are built.

The main statistical modelling methods applied in this PhD are errors-in-variables Bayesian regression (shortened to EIV BR) and errors-in-variables Gaussian processes (shortened to EIV GP). Both statistical modelling methods account for *measurement error* (synonymous with errors-in-variables) in the explanatory variables, that is, the observations of the explanatory variable are known to be not exact (measurement error is also accounted for on the response variables). In many engineering experiments, it is often the case that the observed data is not measured exactly, for many reasons. For example:

- a set of scales for measuring the mass of a powder is not capable of providing its exact mass,
- pouring powder through a funnel for measuring its angle of repose will lead to randomness, whether it be human error or simply randomness in particle movement.

This is a component of modelling that is very often neglected by those who have not acquired significant training in applying statistical models. Consider again the simple linear model in Equation 1.1. As mentioned in the third bullet point, the model accounts for error in the response variable (i.e., vertical distance between the data points and the fitted model); conversely, the observed data for the explanatory variable is assumed to be exact in this framework. In this case, a violation of the modelling assumptions occurs when the explanatory variable observations x_i are prone to error. An errors-in-variables approach is an extension of standard linear regression modelling that accounts for randomness in the explanatory variable(s), and is considered further in Chapter 3.

The modelling methods mentioned above are considered in a Bayesian framework. *Bayesian* statistics is a branch of statistics that contrasts the more well-known *frequentist* statistics, which is built on the fundamental understanding that the probability of an event occurring is based on the relative frequency of the event occurring given a large number of repeats of an experimental scenario. Conversely, Bayesian statistics often relies on subjectivity to form probability statements (see Section 3.1).

In terms of statistical modelling, a frequentist approach relies heavily on the available data collected in an experiment, and as such the parameters of the model are estimated using just data, whereas parameter estimation in a Bayesian setting considers the available data as well as some prior beliefs about the parameters to be estimated. There are multiple advantages of Bayesian statistical modelling:

- the 'prior beliefs' can be informed by expert opinion to provide some joint understanding of the parameters based on both experts in the field and the data that is available,
- measurement error in the data can be treated naturally,
- a single solution can be produced for the process of optimising the input variable(s) to produce some desired response.

Statistical modelling in a Bayesian framework provides the ability to treat any model parameters, as well as the response and explanatory variables,

as random. Thus, for the second point, the response variable(s) and input variable(s) inherently can account for error by definition (i.e., Bayesian modelling immediately admits measurement error).

To the third point, treating the input variables as random allows for a more justified theoretical approach to estimating the input variable(s) given some desired response value. Using conditional probability, the probability distribution $p(Y|X)$ (that is, the distribution that describes the random behaviour of the response variable given the explanatory variable) can be rearranged with relative ease to find the probability distribution $p(X|Y)$ (describing the random behaviour of the explanatory variable given the response variable). This is achievable because both variables are considered to be random here; in comparison, the classical approach assumes there is no randomness in the explanatory variable, so the probability density function $p(X|Y)$ does not exist, and the problems discussed with respect to Figures 1.1 and 1.2 arise. This is discussed further, with extensions to multivariate regression, in Chapter 3.

The contrasts, both theoretically and practically, between the two methods of EIV BR and EIV GP, are discussed in this work. The main theoretical distinction between the methods is how they approach relating the response variable(s) and the explanatory variable(s); the EIV BR is a *parametric* approach, whereas the EIV GP is a *nonparametric* approach. 'Parametric' in this sense relates to the belief that a 'predetermined' relationship exists between the variables. In the case of the simple linear model discussed briefly above, the model is defined based on the equation of a straight line, which makes the obvious assumption that some straight line determines the relationship between the response variable and the explanatory variable. Hence, the model parameters are estimated with this predetermined assumption. The assumption need not be a straight line; there could be a belief that a quadratic curve relates the two variables, i.e.,

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \epsilon_i,$$

or with the inclusion of multiple explanatory variables, an additive model in the explanatory variables could be the predetermined relationship (such as Equation 1.2). No such predetermined relationship exists when fitting a nonparametric model. The nonparametric approach makes far fewer assumptions about the relationship between the variables. Note that there are still parameters to be estimated in a nonparametric model, and that 'nonparametric' simply refers to the fact that there is no fixed, predetermined relationship between the output(s) and input(s). The Gaussian process,

for example, makes the assumption that the response variable data (and any subset of this data) takes a multivariate normal (Gaussian) distribution. This distribution is then informed further by the assumption that, if two data points of the explanatory variable are close together (in terms of Euclidean distance), it is likely that the corresponding data points of the response variable are close together. These approaches are defined in Chapter 3, and applied and compared in Chapters 4 and 5.

The two modelling approaches of linear regression and Gaussian process regression are covered in both the *forward modelling* and the *backward modelling*. These two terms are linked with the two questions asked at the start of the introduction—the forward modelling corresponds to the estimation of the relationship between the response variable(s) and the explanatory variable(s), and the backward modelling corresponds to the optimisation of the explanatory variables, given some desirable value of the response variables, and given the best-fitting statistical model found in the forward modelling. The terms forward modelling and backward modelling are defined as such because of the interpretation of the variables; the forward model is interested in the outcome of the experiment, and how this is affected by the explanatory variables; the backward model inverts this, and considers how to arrive at a particular outcome, or more exactly, how the outcome affects the explanatory variables. Both the forward modelling and backward modelling are explored theoretically in Chapter 3. The forward modelling of both the statistical methods that feature in this work is implemented in Chapter 4 (with consideration of simulation examples, and real data, which is discussed below), and the backward modelling is demonstrated in Chapter 5.

The subsequent chapters of this thesis are now described briefly below:

- The content of Chapter 2 provides an overall view of AM, with further details describing the process of Laser Sintering, the variables tapped density and angle of repose (with regards to how they were measured for this work), the information that can be acquired from the FT4 powder rheometer, and an overview of the statistical research that currently exists within the AM literature.

- The statistical methods applied in this work are covered in Chapter 3, with an overview of the required foundations (such as the fundamentals of Bayesian statistics and measurement error models), and a detailed discussion of how these models are fitted, assessed and compared in the thesis. This includes the details for both the forward modelling

and the backward modelling, for both the EIV BR models and the EIV GP models.

- Chapter 4 contains a thorough investigation of applying the forward EIV BR and EIV GP models to various sets of simulated data, with the main idea of developing these methods using simulated data to ensure that the model is performing as intended for a variety of scenarios (different linear predictors—the function of the input variable(s) that defines the relationship between the output(s) and input(s)—for example). Subsequently, the methods are applied to the powder flow data (discussed in Section 2.2) to demonstrate the approach and answer the first of the two questions discussed above, that is, to find those explanatory variables from the FT4 powder rheometer that are required to provide the best understanding of the response variables tapped density and angle of repose. The numerical results from applying the EIV BR and the EIV GP models to the real data are also compared.

- The work of Chapter 5 is carried out analogously to that of Chapter 4 (that is, considering sets of simulated data to develop the modelling, followed by applying the models to the real data), with the focus switched from the forward model to the backward model. This chapter provides an answer to the second of the two questions discussed above, that is, what are the optimal values for the vector of explanatory variables that are most likely to produce some desired response vector. As in Chapter 4, the numerical results from applying the EIV BR and the EIV GP models to the real data are again compared.

- The penultimate chapter, Chapter 6, describes aspects of the work that have not been covered in this thesis, that would develop the process and outcomes further.

- Finally, the work is concluded in Chapter 7, describing the results of the work carried out in the thesis.

Chapter 2

Additive Manufacturing

As defined in the annual 'Wohlers Report' (see Wohlers et al. (2019) for the 2019 edition), additive manufacturing is the 'process of joining together materials to make objects from 3D model data, usually layer upon layer' (3D model data referring to CAD files that represent printable 3D objects). Generally speaking, manufacturing methods are often a means of removing materials from starting blocks like metal and plastic and combining them to create an object—this is known as subtractive manufacturing. This distinction is the key difference between additive manufacturing and subtractive manufacturing; with additive manufacturing, the starting point is effectively what the raw material is made of, meaning there is far less waste in additive manufacturing than in subtractive manufacturing. The main advantage of additive manufacturing is the relative ease of creating objects with a high complexity of geometry (Goodridge et al. 2012).

There are several methodologies for additive manufacturing, and this list will continue to grow. The methodologies can often be split into categories depending on its process—these processes are binder jetting (powder materials joined together by liquid bonding agent), sheet lamination (object is bonded by sheets of material), material extrusion (material dispensed through a nozzle), powder bed fusion (the application in this work, where thermal energy is used to fuse sections of a powder bed), material jetting (droplets of material are jetted (spurted out in a rapid stream) onto a platform), vat photopolymerisation (a photopolymer is cured or hardened using light), and directed energy deposition (thermal energy is used to fuse materials by melting through deposition). A more extensive review of the various methodologies can be found in Wong & Hernandez (2012).

A large variety of applications of AM processes exist, particularly in

fields where highly complex object geometries are required. Some examples of 3D-printed objects are teeth retainers, a prosthetic leg, car parts and model replicas. The following subsection details the additive manufacturing process 'Laser Sintering', and some of its key aspects which have been of interest in this work.

2.1 Laser Sintering

AM processes falling under the category 'powder bed fusion' rely on thermal energy to fuse powder together. A variety of materials, such as polymers, metals and ceramics, are suitable for these processes; the work here is specifically focused on polymers, with the statistical modelling also being applicable for a broad range of materials and uses. An image demonstrating the process of Laser Sintering is shown in Figure 2.1 (image produced in Hopkinson et al. (2006)). The materials used for laser sintering are typically nylon-based (polyamide polymer); other materials, such as elastomers (rubber-like material) are becoming more frequently used. The material Polyamide 12, also known as Nylon 12, is by far the most popular powder of choice (Goodridge et al. 2012). Examples of industries which benefit from laser sintering include the aerospace industry and the orthotics industry, where, in both cases, the complexity required for specific designs is well suited to the capabilities of laser sintering.

The laser sintering process, following the depiction in Figure 2.1, is detailed here:

1. Firstly, the 3D-printed object is designed in a Computer-aided design (CAD) package.

2. The CAD file of the object to be printed is delivered to the machine in thin slices.

3. A thin layer (approximately 100 microns) of polymer powder, from which the final 3D-printed object will be constructed, is deposited into the machine.

4. Once 0.1mm of powder has been deposited across the powder bed, the powder bed is heated to a temperature just below the melting point of the powder.

5. The laser is reflected off the scanning mirrors to sinter the powder (i.e., heat it above its melting point), with direction according to the CAD file.

6. Upon completion of the sintering of the layer of powder, the powder bed is dropped, so that another layer of powder is spread onto the powder bed.

7. The steps in (6) and (7) are repeated until every layer of the object has been sintered.

8. The powder bed is dropped so that the 3D-printed object is accessible and is then removed from the machine, ready for post-processing.

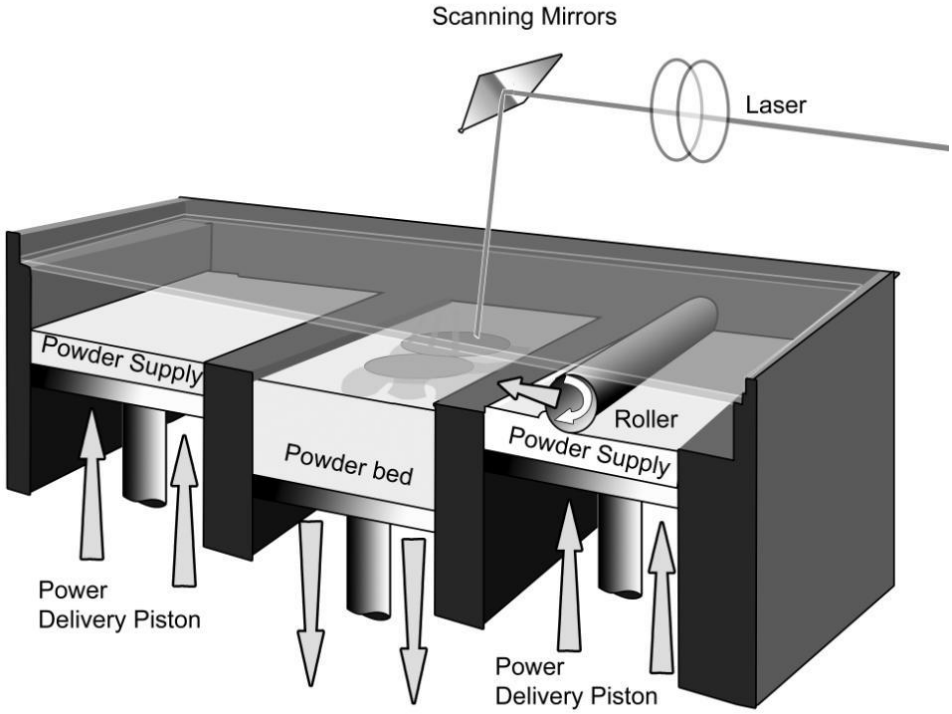

Figure 2.1: A visual representation of the additive manufacturing process laser sintering

The term 'post-processing' refers to the actions carried out on the object once it has been printed. For laser sintering, this involves brushing away any loose powder from the object (any leftover powder from the machine can be recycled, depending on its next use), and then typically the object is 'bead-blasted', that is, the object is put into a bead-blasting machine, where

the user shoots fine glass beads at high pressure at the object, removing any further excess powder from the object as well as improving the surface finish.

Laser sintering is typically an AM methodology of choice when the design of the object is very sophisticated and precision is required. The 3D-printed objects built with laser sintering tend to have relatively high-quality mechanical properties and good geometrical capabilities, but the surface finish of the final part can be lacking, and the speed of the process can be slow if the cross sections require large amounts of scanning.

2.1.1 Powder flow in Laser Sintering

As described above, the process of laser sintering involves a powder being transported through chutes/hoppers in a machine, as well as the powder having to settle on a powder bed for the laser sintering process to continue (Goodridge et al. 2012). There are two key aspects of powder flow which are important in the success of this AM methodology—firstly, the powder must be able to flow at close to an optimal level to get through the chutes/hoppers without either clumping (when the powder is too cohesive) or dissipating (when the powder is too free flowing and 'disappears', or becomes dust). Secondly, the powder must be able to settle evenly on the powder bed to form a flat layer of powder. In this work, two variables are considered which quantify these aspects. The first variable is tapped density, described further in Section 2.1.2, and the second variable is angle of repose, described further in Section 2.1.3.

2.1.2 Tapped density

As a measure of density, the *tapped density* of a powder measures the mass of the powder that can occupy a given volume after it has been 'tapped' in a systematic manner. The 'tapping' of a powder is a method for trying to pack the powder particles as tightly as possible into some volume. Having the powder be tightly packed is beneficial in Laser Sintering; when the powder is deposited onto the powder bed, the tighter it is packed together, the smoother the powder bed, and the less likely geometrical deformities occur in the final part.

The process for measuring tapped density in this work is detailed below:

1. Take a separable metal cylinder where the volume of the base of the cylinder is known (100ml in this case)—see Figure 2.2.

2. With the metal cylinder assembled (both parts are attached to one

another), insert some subsample of the powder, ensuring there is sufficiently more powder in the cylinder than the amount that would just fill the base.

3. Fix the metal cylinder to the platform on the left of the device shown in Figure 2.2.

4. Turn the cog clockwise until 100 full rotations are completed. As the cog is turned, the platform is raised and suddenly dropped, simulating the effect of packing the mass of the powder into a smaller volume. With each full rotation, the platform is raised and dropped 5 times, meaning there are 500 drops throughout the test.

5. After completing the 100 rotations, carefully remove the metal cylinder from the platform, then separate the metal cylinder (taking the top off of the base).

6. If necessary, use a ruler (or something similar) to remove excess powder lying on top of the base.

7. Take some resealable bag and place it on some scales, then re-zero the scales (taking into account the mass of the bag).

8. With the mass of powder that remains in the base of the metal cylinder, carefully place the entirety into the resealable bag and weigh the mass of the powder (in grams).

9. Divide the mass of the powder by the known volume of the base of the cylinder to find the tapped density (in g/ml).

The measurement of tapped density is seen as a good proxy for capturing powder bed deposition, that is, how evenly the powder can lie on a powder bed.

2.1.3 Angle of repose

The 'angle of repose' is defined to be the steepest angle (between the horizontal plane and the slope of some pile of particles) at which the particles can be piled without slumping. The process for measuring angle of repose in this work is detailed below:

1. Insert some subsample (less than 100g) of powder into a beaker.

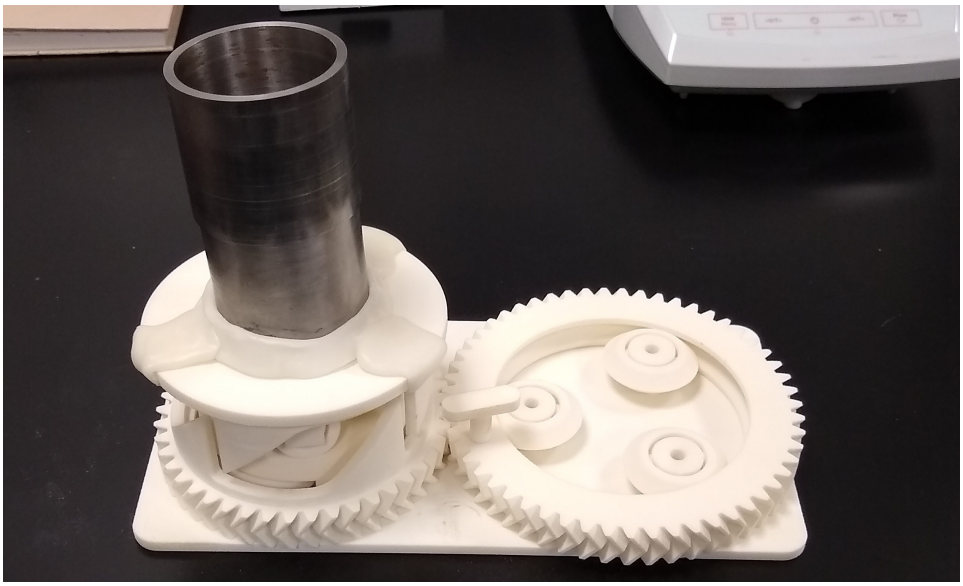

Figure 2.2: Image of the equipment used to measure tapped density

2. Place the tip of the funnel in the centre of some cylinder with known diameter, then transfer the powder from the beaker into the funnel (if the powder is very free flowing, some powder may pour out of the tip and onto the cylinder).

3. Gently raise the funnel to initiate the powder flow through the funnel and onto the cylinder. The funnel may require a slight tap if there is no initial powder flow. As the height of the pile increases, raise the tip of the funnel, ensuring that the tip of the funnel is as close as possible to the summit of the powder pile at all times.

4. Continuously fill the funnel with powder from the beaker until the pile diameter is equal to the diameter of the cylinder.

5. Take the digital callipers and expose the depth gauge to a length greater than the height of the pile, then carefully place the depth gauge in the centre of the pile.

6. Close the callipers until they are in contact with the top of the pile and record the height of the pile.

7. The angle of repose θ is then calculated using

$$\theta = \tan^{-1}\left(\frac{2 \times \text{pile height}}{\text{cylinder diameter}}\right).$$

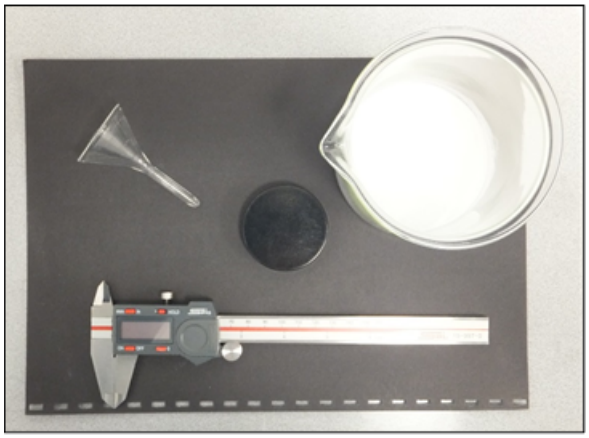

Figure 2.3: Image of the equipment used to measure angle of repose—in the top left is a funnel, in the top right is a beaker filled with powder, in the middle is a cylinder, and at the bottom is a pair of digital callipers.

2.1.4 Powder rheometer

Improving the understanding of the powders being used in Laser Sintering is beneficial due to the influence of powder behaviour on the quality of the final part. Furthermore, acquiring knowledge about the influence of powder properties on the particular aspect of powder flow can help with choice of powder (those which flow 'sufficiently' well) or possibly with the manufacturing of a new powder. A powder rheometer, like the FT4 powder rheometer pictured in Figure 2.4, is a piece of machinery which can perform various tests on a subsample of powder to measure some flow-related characteristics of the powder. An overview of the tests and measurements acquired from the FT4 powder rheometer are presented in Table 2.1; further details relating to the FT4 powder rheometer are found in Freeman Technology (2020).

2.2 Case study – powder behaviour

The specific example considered in this work to demonstrate how to implement the statistical modelling approaches discussed in Chapter 3 looks at

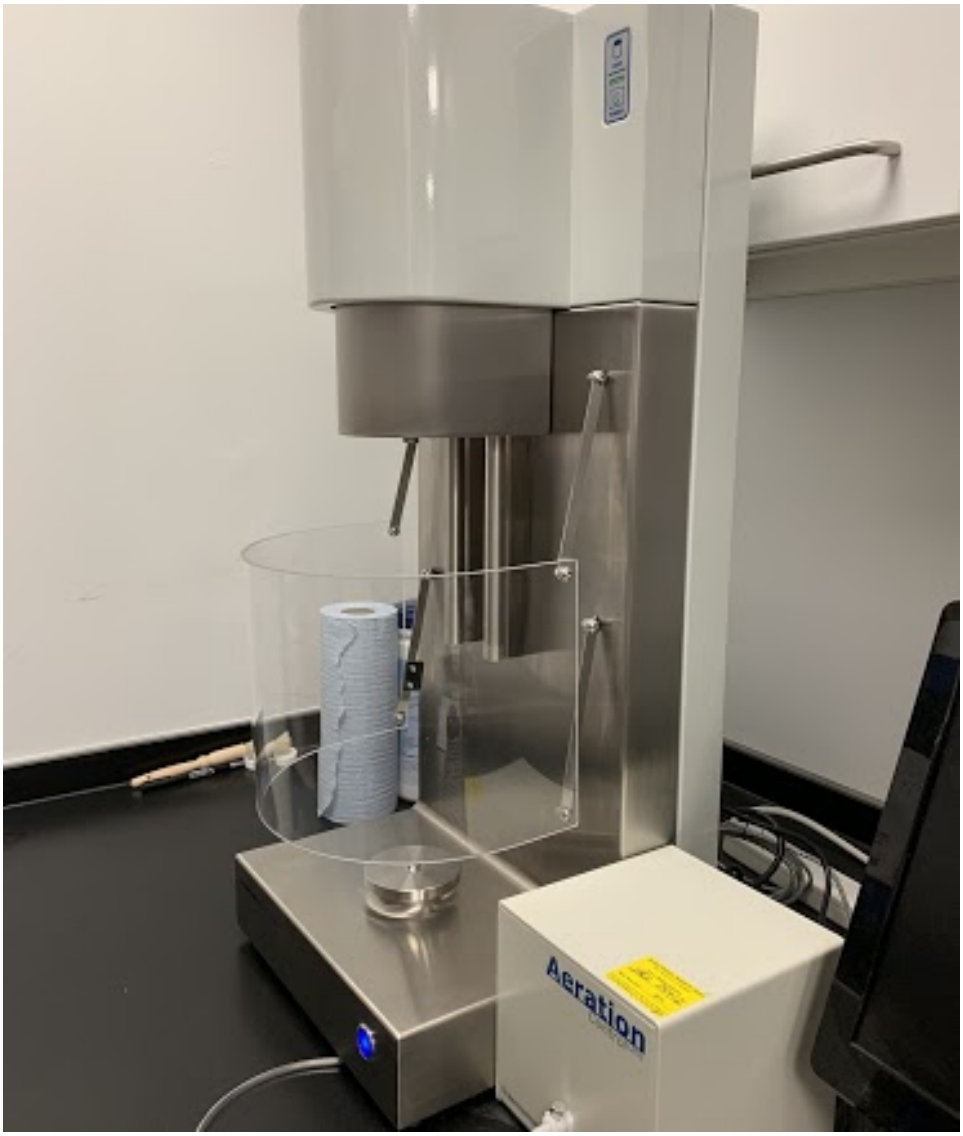

Figure 2.4: Image of Freeman Technology's FT4 Powder Rheometer

Methodology	Measurement	Definition
Aeration	4mm/s test—Total Energy at 2 mm/s (mJ)	Powder is aerated over six tests with the speed of aeration increasing after each test, to a maximum speed of 4mm/s. Total energy consumed is measured during each test
Compressibility	Compressibility percentage at 15kPa Normal Stress (%)	Percentage change in volume after compression. Test completed in nine steps, with increasing pressure from 0.5kPA up to 15kPa
Permeability	Pressure drop at 15kPa Normal Stress	Increasing amounts of normal stress applied to powder, during which air is passed up through vessel. Air pressure drop measured at 15kPa
Shear Cell	Shear Stress (Incipient) (kPa)	Shear head induces vertical and rotational stresses, which are normal and shear stresses respectively. Maximum shear stress occurs (and is measured) immediately before powder bed fails or shears
Stability & Variable Flow Rate	Basic Flowability Energy (BFE) (mJ), Specific Energy (SE) (mJ/g), Conditioned Bulk Density (CBD) (g/ml)	Conditioning refers to gentle displacement of the powder sample so as to loosen the powder, consisting of an upward and downward traverse of a rotating blade at a 5 degree angle. Each conditioning is followed by *testing*, that is, measuring energy consumption. This provides the measuremens: BFE (the total energy consumed in test 7, out of 7 tests), SE (the sum of the Up Energy of Cycle 6 and of Cycle 7 (energy measured on upward traverse of blade), divided by 2 times the Split Mass (powder mass measured after first conditioning)), CBD (the Split Mass divided by the Split Volume (25ml))
Tapped Consolidation	Consolidated Energy (50 taps)	Total energy consumed after 50 taps (controlled vibrations of the powder)
Wall Friction	Shear Stress (Kinematic) (kPa)	Wall friction head induces vertical and rotational stresses. Maximum shear stress occurs as the resistance is overcome

Table 2.1: List of tests that can be carried out with the powder rheometer, with the variables that are measured and how they are defined.

powder flow and *powder bed deposition*. The general focus is to make improvements in the quality of parts produced by the AM methodology *Laser Sintering*; as discussed above, these improvements are dependent on how well a powder flows through chutes in the machine, and how well the powder deposits onto the powder bed. It was also noted that two useful proxies for these components are *angle of repose* for powder flow and *tapped density* for powder bed deposition, which have been measured multiple times on seven different powder materials (each time using a different 'subsample' of the powder). The two proxies are treated as the response variables in this work, and their relationship with variables measured from a powder rheometer are estimated in the modelling. Identifying these relationships between tapped density and angle of repose and those variables extracted from the powder rheometer can provide an understanding of which aspects of the powder rheometer are relevant for describing powder flow and powder bed deposition, and what specific traits a powder should possess in order to flow optimally through the Laser Sintering machine and deposit optimally on the powder bed. This example, most importantly, demonstrates a statistically sound approach for optimising input variables to produce a desired response vector for real-world data that is transferable to several applications within AM.

2.3 Statistical methods in AM

This section reviews some prior applications of statistical methods in the additive manufacturing industry with the aim of improving AM technology. The main focus of this section is to highlight pieces of literature using statistical methods being applied in AM that are related to those considered in this work (i.e., linear regression and Gaussian processes).

The review paper Stavropoulos, Panagiotis & Foteinopoulos, Panagis (2018) is a review of literature on the modelling of existing AM processes, and provides a suitable starting point for identifying applications of statistical methods within AM research. The paper refers to all AM processes, and notes which 'approach' each referenced work considers, with the approaches being 'analytical', 'numerical', 'empirical', or some combination of the three. The empirical approaches that consider linear regression are discussed in this section. Given the application in this work is in laser sintering, those papers that consider this methodology are looked at firstly below, with other methodologies considered after. The limitations of the existing research, and how this work is able to overcome these, is discussed in Section 2.3.3.

2.3.1 Laser Sintering

Linear regression

A common theme in the literature looking into linear regression applications in laser sintering shows that linear regression is often applied in conjunction with an experimental design. The discussion of this begins here with the paper Raghunath & Pandey (2007). The paper aims to improve the accuracy of the final part in laser sintering by modelling 'shrinkage' against various processing parameters. The percentage difference in each of the three dimensions of the CAD model and a fabricated prototype (3D object) are used as response variables, and the processing parameters laser power, scan speed, hatch spacing, part bed temperature and scan length are taken to be the explanatory variables. A Taguchi method (a popular tool in engineering applications for building optimal experimental design; see Rosa et al. (2009) for more details) for designing an optimal experiment is utilised that suggests 16 observations at four levels given the five explanatory variables. Having carried out the experiments, the response variables are treated individually, with the shrinkage % in each dimension being modelled against the explanatory variables. For each of the 16 observations, the shrinkage % is measured and recorded three times. An average of the shrinkage % is provided, implying that measurement error is observed for each observation of the experiment. Signal-to-noise ratios are calculated by $-10\log\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right)$, where n is the total number of experiments, yet the ratio is calculated for each of the 16 observations (in each dimension). Then the signal-to-noise ratios are used to choose the optimal settings for each of the processing parameters, though the process for making the choice is unclear. An application of Analysis of Variance (ANOVA) is then carried out, in order to identify which processing parameters are influential on the shrinkage % in each dimension. The full model (with all five explanatory variables) is compared with five nested models, having removed each of the explanatory variables and fitted the model with the four remaining explanatory variables. Those variables whose removal results in a small p-value (the probability of repeating the experiment and the new data from the experiment providing a more extreme result, i.e., a smaller F-statistic of the F-test) are all removed from the model.

In a similar fashion, the paper Wegner & Witt (2012) also considers an experimental design framework to help identify the most influential input variables on the mechanical properties of final parts in laser sintering. In comparison, a central composite design is implemented, as opposed to the

Taguchi method in Raghunath & Pandey (2007), which accounts for higher-order and interactions terms in the design. A second-order response surface model is fitted (effectively a linear regression model with a quadratic and interactions terms given an optimal experimental design), and again, as above, utilises the ANOVA in order to identify which input variables are most important for predicting mechanical properties of final parts.

Furthermore, the paper Lappo et al. (2003) investigates the use of multiple materials simultaneously in laser sintering. The method they implemented was an experimental design in order to determine the possibility of an 'electrostatic powder removal', as opposed to vacuum removal, of the first powder. They investigate the influence of the three variables powder removal height (in mm), charge voltage (V), and the powder material, on the volume of powder removed through electrostatic removal. Each of the three input variables had two levels, and interaction terms were considered, leading to an experimental design with 8 possible combinations of the input variables, and each combination (trial) was carried out twice. The ANOVA was applied to the three-covariate model with interactions between each of the input variables (but no polynomial terms for each variable), and indicated that the interaction term between powder removal height and powder material was insignificant. The p-values of the F-tests were not reported, instead, the Probability of the F-test was given, with values 'close' to 1 indicating values of high confidence (indicating this gives 1 minus the corresponding p-value). The authors note that, with only three input variables considered, with little inclusion of processing parameters, potentially valuable information is not captured in their model. This is an appropriate conclusion, particularly given that the best-fitting model given 16 observations included 7 model coefficients (possible overfitting issues should be noted).

The work carried out in Jain et al. (2008) is similar to that of Raghunath & Pandey (2007), with the implementation of the Taguchi method for optimal experimental design to investigate the effects of the input parameters layer thickness, part bed temperature, refresh rate and hatch pattern on the tensile strength of the final part. Once more, the ANOVA is carried out to indicate which variables have a statistically significant effect. Again, a four-level design with 16 observations is carried out, and the resulting ANOVA indicates that all input parameters have a statistically significant influence on the final part tensile strength. They conclude that inclusions of laser power, beam speed and hatch spacing may lead to further improvements in the model.

The aim of the paper Majewski et al. (2008) is to improve the under-

standing of final-part mechanical properties from laser sintering. They note that laser sintering final parts, while typically having appropriate levels of tensile strength, often suffer from low levels of ductility. A possible reason for this was that the consistency of melting throughout the powder during the printing was low. Differential scanning calorimetry (DSC) tests are able to detect two separate melt peaks, corresponding to a fully-melted region of material and a region which has not fully melted. The term degree of particle melt then defines the variation in particle melt. The experiment in their work looks at investigating the influence of degree of particle melt (calculated using DSC tests) on the mechanical properties. Having carried out appropriate experiments in order to investigate this, simple linear regression models were applied to investigate the relationship between each output variable (elongation at break, tensile strength and E-Modulus (elastic modulus, or Young's modulus)) against core peak height (relating to degree of particle melt). In each case R^2 was used to determine how well the regression line represents the actual data, and it is then claimed that R^2 values above 0.9 indicate the data follow a significant trend.

Another example of the combination of experimental design and using the ANOVA for variable selection is provided in Wegner et al. (2015). The paper investigates the effects of the processing parameters in laser sintering on the mechanical properties when the material Polyamide 11 is used. This is notable since Polyamide 12 is definitively the predominant material applied in LS. The authors implement the response surface model which is also implemented in Wegner & Witt (2012), and a central composite experimental design is once again used.

Ha et al. (2018) aimed to develop a 'compensation' algorithm (previously developed in the AM and also subtractive manufacturing literature), for the purpose of better understanding the relationship between final-part dimensional inaccuracy in laser sintering and CAD design. The compensation algorithm is effectively a process for printing 3D objects given an initial CAD design, adjusting the initial CAD design based on inverting predictions of measured deformations from a quadratic regression model of surface data, and measuring the improvement in dimensional inaccuracy given the updated CAD design. The work is an extension of the previous paper Ha et al. (2015). The choice of implementing a quadratic regression model, while also based on the evidence provided in Ha et al. (2015), was on the basis of comparing R^2 values of the linear, quadratic, cubic (and even higher polynomial terms) regression models. Crucially, it is also noted that the estimated model coefficients from the quadratic regression model are used further in the compensation algorithm, with no consideration of the

uncertainty in those estimates.

A work that is closely related to this PhD is that of Vetterli (2019), a PhD thesis with the interest of optimising powders for use in laser sintering. One of the main interests of the work was to apply statistical regression methods combined with experimental work to better understand the influence of 'intrinsic' and 'extrinsic' particle properties on the flowing and packing behaviour of powders that could potentially be used in laser sintering. Therefore, the application of statistical regression models and the case study of powder flow and powder bed deposition (discussed further in Section 2.2) align to a large extent with the work in Vetterli (2019). As described in Section 2.2, this work considers (input) variables measured from a powder rheometer to investigate their influence on tapped density (proxy for powder bed deposition) and angle of repose (proxy for powder flow). Conversely, Vetterli (2019) considers different explanatory variables for different output variables. Firstly, the response variable 'avalanche angle' is modelled against elliptic smoothness (particle shape), Young's modulus (stiffness of material), compressibility, and inverse of specific density (including an interaction term between elliptic smoothness and Young's modulus). Secondly, the response variable 'surface fractal' (of an avalanching event) is modelled against the span of the number-based size distribution, the square of the median particle size value, Young's modulus and the relative bulk density. Finally, the response variable 'relative bulk packing density' is modelled against the square of the aspect ratio, particle solidity, Young's modulus, tensile strength, and the median particle size value. Further discussion of limitations of this work (in general terms) is given later.

A final example of an application of linear regression in laser sintering is in Baturynska (2018), where the focus is on improving dimensional accuracy by considering processing parameters. Interestingly, the data collection process involved taking mean values of repeated measurements of length, width and thickness, and using these mean values as data. This was carried out in order to reduce the impact of measurement error, as opposed to fitting an errors-in-variables model similarly to what is carried out in this work.

Gaussian processes

The more recent statistical methods applied to aspects of laser sintering, and other AM methodologies, are machine learning techniques, which have become prevalent in recent years, and are a fast-moving field like AM. In this work, Gaussian process regression (GPR or GP regression) is of interest. Applications of Gaussian processes within laser sintering are novel, with the

first paper considered here being Song et al. (2018). They note that experiments have shown possible improvements in thin film coating in laser sintering with laser-improved nano particle deposition (as opposed to nano particle deposition without laser), but further understanding is required for this to be reliably used. The main desire of the work is to apply controlled kinetic Monte Carlo simulation to simulate the laser-improved nano particle deposition, to improve the understanding of this process, with regards to possible materials to be used. The use of Gaussian processes here is limited to a validation process, to compare their simulation results from kinetic Monte Carlo with experimental results. The reason for using Gaussian processes to carry this out was that this method provides more flexibility compared with linear regression models, where new data can be incorporated into the GP regression model without amending the previously built model (this is not the case for linear regression). They fitted the difference between the simulation prediction and the experimental measurement, where both are functions of laser power, with a Gaussian process with constant mean and the squared exponential kernel as the covariance function. It is unclear what method was used to estimate the hyperparameters of the GP, and any uncertainty in these estimates is not reported.

The work of Czelusniak & Amorim (2020) aligns with the work in this thesis in multiple ways; their desire is to implement a Gaussian process regression to investigate the relationship between final-part properties (mechanical properties, dimensional inaccuracy) and laser sintering processing parameters, with the use of stochastic multi-objective optimisation to find ideal processing parameters. The key reason for implementing a Gaussian process was its modelling flexibility, where a predetermined relationship between the outputs and inputs need not be given. Notably, the work identifies several response variables relating to mechanical properties (such as tensile strength at yield, elastic modulus, etc.) and chooses to fit multiple single-output Gaussian process models with laser power, laser speed and scan line spacing as input variables. An experimental design is also carried out. They use leave-one-out cross-validation root mean squared error (or RMSE, discussed in Section 3.5.4) to evaluate the predictive performance of each GP regression, as well as the coefficient of determination R^2. They elect to consider only those GP models whose $R^2 > 0.9$ for optimisation. The optimisation process, using multi-objective stochastic optimisation, is notably different to what is performed in this work. Their multi-objective algorithm approach optimises the input variables for maximisation or minimisation of multiple objective functions simultaneously, and in this case a multi-output GP is not necessarily required. In comparison, the methods

in this thesis allow multiple response variables to be considered simultaneously in the modelling, and an inverse modelling approach is applied (in a Bayesian setting) for finding optimal input values for desired response values. The optimal input parameters in Czelusniak & Amorim (2020) were found using an evolutionary algorithm stochastic optimisation. The method introduces a population of individuals, where, for each individual in the population, a 'mutant' individual is created, and combined with the initial individual to give a 'trial' individual, who is evaluated using some 'fitness' function. The best individual between the initial and trial is chosen based on the fitness function, and it remains in the algorithm for future iterations. This algorithm is carried out until the chosen number of iterations have all been performed. The work carries out three two-objective optimisations: optimising manufacturing time and dimensional accuracy in Z-direction, optimising all mechanical properties and dimensional accuracy in all directions, and optimising all mechanical properties and manufacturing time. In each case, the remaining response variables that are not optimised were assigned 'hard lower (or upper) limits', where any optimal solutions with values of these variables smaller (larger) than these lower (upper) limits are discarded. The input values that maximise (or minimise where appropriate) the output variables are identified for each of the three optimisations. Given the trade-offs for each response variable, multiple solutions sets are identified.

The boom in application of Bayesian optimisation (BO) in AM is evidenced by multiple papers in recent years. The work of Batabyal et al. (2022) provides an example of BO being applied within laser sintering, for the purpose of finding optimal values for the parameters 'surface diffusivity' and 'interparticle distance' in order to produce desired values of 'size of neck region between two particles' during sintering. The work makes use of a phase-field model which replicates the evolution of microstructure and physical powder properties, providing the data for which the Gaussian process (GP) modelling is trained. An experimental design (as part of the BO) is applied for data simulation, notably a four-level full factorial design, where the phase-field model is run for 16 simulations at uniformly sampled values of the two input variables, with different endpoints depending on whether the particles were equal-sized or not. The GP hyperparameters are estimated with maximum likelihood estimation, and the model predictions were tested using RMSE cross-validation (training the GP on 75% of the data, then testing the GP on the remaining 25% of the data). Having trained and tested the GP, input parameter optimisation is carried out using BO, with the aim of choosing values for the two input variables in order to maximise the output variable. Two acquisition functions were considered for the pro-

cess of finding optimal input values: Probability of Improvement (choosing the input values $\boldsymbol{x}$ to maximise the probability that $\frac{Y(\boldsymbol{x})-y_{\max}}{y_{\max}}$, where $y_{\max}$ is the maximum observed value so far) and Expected Improvement (instead choosing $\boldsymbol{x}$ to maximise the expectation of $\frac{Y(\boldsymbol{x})-y_{\max}}{y_{\max}}$). The efficient global optimisation approach was then implemented for each acquisition function to maximise the output variable. Further details can be found in Batabyal et al. (2022).

Limitations

The limitations of the papers discussed above are reviewed here, and demonstrate some suboptimal aspects of current statistical research being applied in Laser Sintering. The use of R^2 as a model *comparison* tool is not advised. While it is true that larger R^2 values indicate that a larger amount of the variance in the response variable is explained by the model, it is tricky to use as a suitable measure of model performance. Values of R^2 will increase with the inclusion of polynomial (and interaction) terms, supposing the same explanatory variables are considered, and so R^2 can always be increased. In particular, there often appears to be a desire for $R^2 > 0.9$; while values this large are not necessarily bad, they could indicate that the model is overfitted. This is explained further in Section 3.4.1. Using R^2 for model selection is not recommended, for reasons noted above (adjusted versions of R^2 exist that do not necessarily increase as more explanatory variables are included in the model, making these versions more suitable for model comparison; further details are not provided here). Performing a F-test to investigate whether, for example, the quadratic coefficient provides an improvement in the understanding of the response variable, is preferable. Moreover, model evaluation of GPs using R^2 is not advised, since overfitting may occur.

It is also clear that errors-in-variables models are very rarely considered, despite the frequency with which they would seem appropriate; instead, averages of the data are calculated which are then used in the modelling as the data points. This eliminates the possibility of understanding how the variability in 'replicate' measurements can influence the statistical model.

A linear regression approach in a Bayesian framework is hardly explored in the AM literature. The likely reason for this is the boom of ML techniques in recent years, due to the ease of application and understanding. Therefore, a lot of the statistical research in AM in recent years considers ML techniques, with linear regression becoming less popular, hence Bayesian regression is hardly considered.

In multiple cases above, variable selection using the ANOVA possibly

eliminates variables too early, without checking the significance of the variables in the nested models. Moreover, the application of the ANOVA led to simultaneous elimination of multiple explanatory variables, as opposed to removing each explanatory variable in turn, refitting the model, and carrying out the ANOVA again.

Referring in particular to Vetterli (2019), the optimisation of powder properties considered a screening process for composite powders that satisfy already known 'optimal' powder properties. In comparison, the benefit of the methods in this work for optimising input variables is that a statistically sound inversion of the estimated multivariate regression model is implemented, given desired response values, and so a screening process is not essential.

Finally, there are several instances of multiple response variables being identified, followed by fitting univariate regression models, as opposed to multivariate regression models. Multivariate regression is beneficial since it is able to capture the relationship between the response variables.

2.3.2 Other AM methodologies

While the application of this work is for laser sintering, given the versatility of the statistical modelling, the possibilities for applying these methods within other AM methodologies are numerous. For this reason, some examples of statistical research from other methodologies are considered here.

Linear regression

The only paper which considers linear regression that is discussed in this section is Vigneshwaran & Venkateshwaran (2019). In their work, linear regression was used to predict mechanical properties in fused deposition modelling such as tensile strength, tensile modulus, and energy absorption rate with layer height, infill and three different patterns of samples prepared with biodegradable wood-PLA (polylactic acid). A standard multiple linear regression was fitted multiple times (each with different response variables, the mechanical properties mentioned above), as opposed to a multivariate linear regression to consider the potential relationships between the response variables. The estimates of the model coefficients are given without mention of confidence intervals of these estimates. The coefficient of determination R^2 was used to assess the fit of the model.

Gaussian processes

While a surge in GP modelling has occurred very recently, some relatively older examples are also discussed here. The paper Tapia et al. (2016) looks at porosity in metal-based AM, and tries to predict the final part porosity in selective laser melting by estimating its relationship with other selective laser melting parameters. Their method is to use Bayesian inference in order to estimate the hyperparameters in the Gaussian process to then make predictions of the porosity (note this is also carried out in this thesis, but generally it is uncommon in GP applications to see hyperparameter estimation with Bayesian methods). A notable aspect of this research is the use of a non-constant mean function, more specifically, a full quadratic linear predictor in two input variables. As discussed further in Section 3.5.1, a constant mean function is generally preferred, since the covariance kernel can capture all of the variation in the response variable. The assumption of a non-constant mean for a GP is somewhat restrictive, and takes away from one of the benefits of the GP, where it does not require the assumption of a parametric relationship between the output(s) and input(s).

The work of Li et al. (2018) considers thermal field prediction in fused deposition modelling (a method of material extrusion) dependent on processing parameters. They combine a physically-based model, which allows data to be collected that are modelled on thermal field evolution, and fit the data to a Gaussian process. With the output variable of layer-to-layer thermal field (temperature measurements) and input variables of layer thickness, printing speed, nozzle temperature (all relating to printer settings), layer index, printing pattern direction and time, a Gaussian process (again with non-constant mean, but seemingly unspecified) was fitted, estimating the hyperparameters using maximum likelihood. The model was trained on all but two data points, which were left out for model validation by comparing predictions of the model for the 'test' points with actual outcomes using RMSE. Two parameters, heat convection coefficient and latent heat of fusion, were fixed variables during the Gaussian process fitting, which are then calibrated using Bayesian calibration. The Bayesian calibration was carried out by considering the discrepancy between the model predictions and the experimental outcomes, using the equation

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}_0, \eta, t) + \boldsymbol{\delta}(\boldsymbol{x}_0, t) + \boldsymbol{\epsilon},$$

with $\boldsymbol{f}(\boldsymbol{x}_0, \eta, t)$ representing model predictions given input variables $\boldsymbol{x}_0$, the calibration parameters η and time t, and the terms $\boldsymbol{\delta}(\boldsymbol{x}_0, t)$ representing model discrepancy and $\boldsymbol{y}$ representing the observed experimental model

outcomes. An appropriate prior distribution is placed on η, which is updated using the model predictions $\boldsymbol{f}$ and experimental outcomes $\boldsymbol{y}$ to provide a posterior distribution for the calibration parameters. Due to the time dependency in the model, the assumption of independent and identically distributed error terms is violated, and so an online updating process is implemented using an autoregressive (AR) model to better understand the behaviour of the model error terms.

This Bayesian calibration process is noteworthy, as it represents a 'backward' modelling process, where the authors try to understand the behaviour of calibration parameters given the output variable. The backward modelling process in this thesis follows a similar ideological process, where the behaviour of the input variable(s) is conditioned on some desired value of the response variable(s). A more recent example of Bayesian-calibration based work within AM is Ye et al. (2022), which aims to improve the understanding of several aspects of metal-based powder bed fusion simultaneously with a melt pool model and a phase field model. Each model is calibrated in a similar way to that in Li et al. (2018), using a discrepancy term. Most notably, the GP hyperparameters in this work are estimated in a Bayesian setting using Hamiltonian Monte Carlo, which is discussed multiple times in this work (although not implemented), with notable discussion in Chapter 6.

Sharpe et al. (2018) look at optimising the design of mechanical 'metamaterials' (in a sense, custom-built materials than can exceed any naturally occurring materials in particular properties, like bulk stiffness). This provides another example of BO being applied in AM. Specifically, with the knowledge that the design of lattice structures at microscales of a material can improve bulk stiffness, they look at optimising these structures for maximising the bulk stiffness. Because of the microscale on which the lattice must be designed, it must be implicitly represented by an approximation, using finite element analysis. The work continues with the fitting of a Gaussian process regression to estimate the functional relationship relating the bulk stiffness and lattice structures (using a Matérn covariance kernel), followed by a Bayesian optimisation using expected improvement as the acquisition function. Moreover, the authors are working with a heavily constrained problem, and so apply nonlinear inequality constraints on the input space.

The examples of BO applications in AM continue with the paper Liu et al. (2022), who consider the AM methodology material extrusion (specifically considering fused filament fabrication) for optimising properties of polymer nanocomposites (graphene composition %) and printer processing parameters (extruder temperature, print speed, layer thickness) to achieve

the best surface roughness. For GP fitting, an exponential kernel function (using absolute values of distance within the kernel, as opposed to squaring the distance) is chosen. There is a suggestion that a non-constant mean was used, but this is unclear. The method for estimating the hyperparameters is also unclear. Interestingly, the authors compare the efforts of the GP predictions with predictions from other possible statistical modelling methods, such as linear regression, regression trees and support vector regression. These comparisons were carried out using R^2, RMSE and MAPE (mean absolute percentage error). As previously noted, it is difficult to compare statistical models using R^2 (with linear regressions, R^2 simply improves with more complicated linear predictors).

Of interest in Lu et al. (2022) is the relationship between final-part quality (in particular, relative density) and processing parameters in laser powder bed fusion. More specifically, they note that the understanding between these variables is good when considering particular alloy materials, such as Inconel 718, but not so good for Inconel 625. Their aim is to carry out 'knowledge transfer' by implementing Bayesian learning, to improve the understanding of this relationship when using Inconel 625 given the high level of understanding of the relationship when using Inconel 718. To do this, they consider two models,

$$\begin{aligned} Y_i &= f(\boldsymbol{x}) + \epsilon_i, \\ Y_j &= f(\boldsymbol{x}) + \delta(\boldsymbol{x}) + \epsilon_j, \end{aligned}$$

with the first model explaining the well-understood relationship while using Inconel 718, and the second model incorporating a discrepancy function $\delta(\boldsymbol{x})$, representing the difference between the output variables when considering the different alloys. Note that the variable $\boldsymbol{x}$ represents the various processing parameters, such as laser power, scanning speed, and so on. Naturally, the relationship f must still be estimated, which is done so with a GP, with constant mean and a squared exponential automatic relevance determination kernel. The discrepancy function $\delta(\boldsymbol{x})$ also assumes a GP with the same assumptions. The hyperparameters of both GPs are estimated simultaneously by posterior distributions found using MCMC.

In order to improve the decision-making process for final part polishing in metal AM, Jin et al. (2020) implement a Gaussian process regression to improve the time taken to carry out the polishing process, where time is often wasted with repeated inspections (checking how well polished the part is) during the polishing process. The GP informs a decision rule for when the polishing has been carried out sufficiently. The data collection

and processing is extensive, and so is not discussed here. The GP models surface height against the location of the measurement and the quantile of surface height. It is also noted that these surface heights are measured at numerous T time points, and for each time point, a GP is fitted. They use maximum likelihood to estimate their hyperparameters, which include some constant mean, and notably multiple distance-scaling parameters (one for the locations, one for the quantiles). The decision rule is then influenced by the distance-scale parameter for the locations (that for the quantiles is deemed to be unimportant), which notably change in relation to time—as the polishing process continues, the surface heights are being reduced, and hence the distance-scales become smaller, as they are relative not absolute. This is a notable takeaway from this work, as they note that the value of the estimate of the distance-scaling parameter is relative to both the input variables and the output variables. This is also discussed in this work (Section 3.5.1 and Section 4.3).

The work of Maculotti et al. (2022) aims to deal with the issues in inaccuracy of surface topography measurements that can occur in AM and can possibly lead to suboptimal process parameter choice based on these inaccurate measurements. They consider two case studies, with the second related to electron beam melting, a metal powder bed fusion method, which is able to build geometrically complex parts and in particular, they are concerned with non-measured points. Most notably, the work investigates the appropriateness of multiple covariance kernels and chooses the best one to represent their modelling situation, using the analysis of a variogram, and also factors in the minimisation of RMSE. This had not been considered in this work. They opt for the squared-exponential covariance kernel (interestingly having selected a Matérn kernel in their first case study, based on the variogram analysis), then use a form of cross-validation, training their model on a subset of the sampled data and testing on the remaining dataset to validate the model. Once validated, the model is used to correct the non-measured points, in other words, predictions of the surface topography are carried out based on the GP (which they suggest outperforms a spline regression in this case).

A complex 'hybrid hierarchical modelling' (HHM) approach, which implements Gaussian process modelling, is developed in Yang et al. (2022), in an attempt to identify the various possible sources in AM (their specific case study considers vat photopolymerisation) that lead to geometric final part accuracy (specifically wall thickness). These possible sources include the machines that produce the parts, the individual parts themselves, and position of the parts within the machines. It is firstly notable that they adopt a

Bayesian approach to the hierarchical modelling, which is notable since, of the work from the literature considered here, a fully Bayesian approach is very uncommon. The hierarchical setup includes the part-level modelling, and the feature-level modelling, with the former investigated with a two-level hierarchical Bayesian linear model (HBLM, with level 1 concerning part-related predictors of positioning of the object within the machine in absolute (x- and y-directions) and relative terms (distance from the centre), and level 2 concerning the machine index), and the latter being investigated using GPs. Thus their HBLM assumes the hierarchical structure

$$y_{i,j} = \beta_{0,i} + \beta_{x,i} s_{x,i,j} + \beta_{y,i} s_{y,i,j} + \beta_{\text{dist},i,j} + \epsilon_{i,j},$$

where there is assumed to be different model coefficients for each machine (printer) i, and within each group i, and the subscript j represents the part j printed by machine i. Some normal prior distribution is assumed, then the parameters are estimated by their posterior distributions, updating the priors using the data. The model is compared with other competing models from the literature using RMSE in three ways: looking at the entire data set, looking at just training data (in-sample predictions), then looking at test data (out-of-sample predictions). The feature level is then modelled using a GP with standard assumptions (constant mean), where the area of each part j is partitioned into grids, represented by the subscript k. The response variable of the GP is then the residuals of the grid averages of wall thickness and part averages of wall thickness. The introduction of the feature-level modelling using GPs into the HHM (so the hybrid model is some combination of the HBLM and the GP) provides an improvement over the HBLM on its own, evaluating model predictions using RMSE again.

Chen et al. (2022) have a specific focus in laser melting. The authors combine a 'multimodal' Gaussian process, which incorporates a mixture of various distributions (some potentially non-Gaussian), and a linear model of corregionalisation (this is discussed further in Section 3.5.5), a type of multi-output Gaussian process, which captures correlations between output variables. In doing so, they aimed to estimate relationships between design parameters in AM and final part characteristics specifically when building thin walls with 'laser powder bed fusion'. This appears to be the first instance of a MOGP (multi-output Gaussian process) being applied within AM.

The paper Tapia et al. (2018) considers the methodology Laser Powder-Bed Fusion, a metal-based process, with the desire to predict melt pool depth given parameters of the process–laser power, scan speed, and laser beam size

combination. Notably, their method for estimating the hyperparameters from the Gaussian process is carried out using Bayesian inference, which is uncommon.

2.3.3 Conclusions from the literature

It is clear from the literature that there is a gap in which the work in this thesis fits appropriately. In terms of applications of linear regression, it is very common in the literature for variable/model selection to be carried out suboptimally. In addition, it is noted that there is a clear lack of errors-in-variables approaches to the linear modelling, in spite of the prevalence of measurement error. The benefits of Bayesian regression, which are discussed further throughout this thesis, have been almost entirely unexplored in the AM literature. Finally, consideration of modelling multiple response variables simultaneously is also lacking in the literature. The work here addresses all of these possibilities noted above.

As far as GPs in AM, it has become clear that, with the increasing popularity of machine learning in general, that GPs are becoming more commonly applied in AM research, which is a notable positive. The majority of the examples noted above are very recent (within the past year). This work will contribute further to those examples, with novel applications of errors-in-variables GPs to account for measurement error on both output and input variables.

The ultimate aim of this project, to invert fitted models for optimising input variables given some desired output variables, is explored to some extent in the literature, in particular with the use of BO. While the benefits of BO are clear, this work provides an alternative approach to optimising input variables, which, in relative terms, is applied more simply by not having to deal choice of acquisition function and optimisation algorithm.

The following chapter deals with the statistical methods that are applied in this work.

Chapter 3

Statistical methods

This chapter details the statistical methods being applied in this thesis. The chapter is broken down into six main sections: Bayesian statistics, errors-in-variables problems, Bayesian errors-in-variables regression, evaluating and comparing model fits, nonparametric modelling, and inverse problems. Within each section, the existing methods in the literature are outlined, which is followed by the setting up of the preferred notation here.

Bayesian statistics is firstly discussed broadly as a contrasting interpretation of statistics and probability, with further detail written about Bayes' Theorem, Bayesian inference, prior distributions, hierarchical modelling, Markov Chain Monte Carlo simulation, and linear regression.

The 'errors-in-variables problems' section comments on directionality of assumed error, attenuation bias, measurement error on the response variable, and replicate measurements.

'Bayesian errors-in-variables regression' combines the first two sections to discuss one of the two central methods applied to the data in this thesis. The importance of this method is highlighted in further detail in Section 3.3, but in short, it is able to account for the complicated structure of the data set (i.e., the measurement error in both the response variable(s) and the explanatory variable(s)) with relative ease compared with the classical errors-in-variables model. Moreover, the ultimate goal of being able to invert the relationship between the response variable(s) and the explanatory variable(s), to find optimal values of the explanatory variable(s) to produce some desired values of the response variable(s), is simplified greatly when considered in a Bayesian setting. This is discussed further in Section 3.6.

'Evaluating and comparing model fits' considers appropriate methods for evaluating and comparing classical linear regressions and goes on to

discuss why these methods are not appropriate for the errors-in-variables Bayesian regression, and what methods of evaluating and comparing models are appropriate for this particular regression.

Nonparametric modelling discusses an alternative modelling method where the relationship between the response variable(s) and the explanatory variable(s) is not predetermined, which allows for a more flexible model. Among other advantages, one for this work is the reduction in the number of potential models which could be fitted, especially when considering multiple response variables and several explanatory variables. In this work the method of nonparametric modelling considered is Gaussian processes.

Finally, the inverse problems section relates to the ultimate goal of the work; to optimise explanatory variables in order to produce a desired value for the response variable(s). Generally speaking, the modelling process in this work is divided into two sections. Firstly, the 'forward' model is discussed from Sections 3.1.5 to 3.5.5, where methods are discussed for estimating the relationship between response variables and explanatory variables. Secondly, the 'backward' model is discussed, where the 'best-fitting' forward model is refitted with a fixed, 'desired' value for the response variable(s), with which optimal explanatory variable values can be found. The inverse problems section firstly looks at inverse problems in a general sense before concentrating on the model inversion for Bayesian errors-in-variables regression and errors-in-variables Gaussian processes.

3.1 Bayesian statistics

Bayesian statistics is known as being a contrasting way of interpreting probability compared with frequentist statistics, in that probability is *not* seen as an event's relative frequency after a large number of trials, that is, given a repeatable experiment, the probability of an event A occurring is given by

$$P(A) = \lim_{n_t \to \infty} \frac{n_s}{n_t},$$

where n_s and n_t are the number of 'successes' (i.e. the number of times the events occurs) and the number of trials of the event A, respectively. Instead, the probability of an event A occurring from a Bayesian perspective can be understood as one's subjective probability of the event occurring. In an example of a simple case, such as rolling a die, frequentist interpretation of probability may be suitable, but events which are not easily repeatable (if at all) will often lead to people giving different degrees of belief of the event occurring.

To construct a subjective probability of an event occurring, fair bets are often used. Moreover, a bet is fair if you have no preference on which outcome you decide to bet. Suppose you win £1 (arbitrary) if an event A occurs, and lose £x (some amount you choose to bet) if A does not occur. Then, suppose you choose to bet £y so that your expected profit is zero (meaning you gain no advantage depending on which side of the bet you take, i.e. you do not mind whether A occurs or not), and let the variable Y be the amount of money you have after the bet is settled. Then,

$$E(Y) = P(A) \times 1 + (1 - P(A)) \times (-y) = 0,$$

which rearranges to

$$P(A) = \frac{y}{1+y}.$$

Clearly, you believe the event is extremely likely to occur if you make your bet of £y large.

Another distinction between classical and Bayesian statistics is how a statistical model depends on its parameters; in a classical setting, the parameters are understood as unknown, but fixed quantities, whereas in the Bayesian setting these parameters are random variables, described in terms of a probability distribution.

3.1.1 Bayes' Theorem

The mindset of Bayesian statistics is built fundamentally on the understanding of Bayes' Theorem; the conditional probability of an event A given an event B is expressed as

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}. \tag{3.1.1.1}$$

In terms of Bayesian inference, this can be rewritten using more suitable notation for data and parameters. Consider the vector of data $\boldsymbol{Y}$ along with the parameter θ. Defining $p(\theta)$ and $p(\boldsymbol{Y})$ as the probability distributions of θ and $\boldsymbol{Y}$ respectively and applying Bayes' Theorem appropriately from Equation 3.1.1.1 gives

$$p(\theta|\boldsymbol{Y}) = \frac{p(\theta,\boldsymbol{Y})}{p(\boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\theta)p(\theta)}{p(\boldsymbol{Y})},$$

where $p(\theta,\boldsymbol{Y})$ is the joint probability distribution of θ and $\boldsymbol{Y}$, and $p(\theta|\boldsymbol{Y})$ is the conditional distribution of θ given $\boldsymbol{Y}$. In Bayesian inference, the conditional distribution $p(\theta|\boldsymbol{Y})$ is defined as the *posterior distribution* (sometimes

shortened to the *posterior*) of θ, and the marginal distribution $p(\theta)$ is defined as the *prior distribution* (the *prior*) of θ. Since $p(\boldsymbol{Y})$ does not depend on θ and can be considered as a constant, another form of the posterior distribution of θ known as the *unnormalised posterior density* is often considered:

$$p(\theta|\boldsymbol{Y}) \propto p(\theta)p(\boldsymbol{Y}|\theta).$$

As described in Gelman et al. (2013), the process of Bayesian inference comes in three steps. Firstly, a description of initial knowledge is specified for θ, known as the *prior*, which could be based on underlying knowledge of the problem, a professional's opinion, or failing those, an alternative which expresses weak prior knowledge. This initial knowledge is *updated* using the distribution (or likelihood) of the data $p(\boldsymbol{Y}|\theta)$, which gives the *posterior* of θ, $p(\theta|\boldsymbol{Y})$. With the combination of the data and the prior knowledge, a good understanding of the parameter should be found in the posterior. Finally, the fit of the model must be evaluated to determine the suitability of the posterior distribution $p(\theta|\boldsymbol{Y})$.

In very simple modelling scenarios, the method for finding the posterior distribution $p(\theta|\boldsymbol{Y})$ (using Bayes' Theorem) is straightforward, where it is assumed it is simple to calculate the likelihood, the prior distribution and the marginal distribution of $\boldsymbol{Y}$. In slightly more complicated modelling scenarios, where there are multiple observations $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ and θ is a parameter vector, the method becomes more complicated and requires algebraic manipulation and evaluating integrals. As the modelling continues to get more complicated (and, often, more appropriately corresponds to the scenario being modelled), a point is reached where some computing power and well-developed algorithms are required.

The decision to consider a Bayesian modelling approach in this PhD was based on a few reasons. The benefits of Bayesian methods over frequentist methods are well known; namely the opportunity to incorporate expert opinion into the modelling, and to naturally take into account error within certain aspects of the model. Another reason it is beneficial is apparent when considering the model inversion, which is discussed further in Section 3.6.

3.1.2 Prior distributions

As mentioned in Section 3.1.1, the combination of the *prior distribution* $p(\theta)$ and the data $p(\boldsymbol{Y}|\theta)$, leading to the posterior distribution, is utilised to gather information on the parameter θ, where any initial knowledge that is known is incorporated into the prior distribution. The distribution and the

parameters of the distribution that are chosen are dependent on the amount of initial knowledge, *conjugacy*, and personal choice. Given the posterior distribution is effectively a weighting between the prior distribution and the likelihood of the data, it is true that the posterior distribution is more influenced by the likelihood of the data as the amount of data is increased. This means that, with a sufficiently large amount of data, Bayesian analyses using different prior distributions will agree with each other.

With respect to initial knowledge, prior distributions can be split into a few categories:

- informed—a lot of initial knowledge has been incorporated into the prior;
- weakly informative—small amount of initial knowledge has been incorporated into the prior;
- uninformative—no initial knowledge has been incorporated in the prior.

As an example, suppose data is available to investigate the relationship between the tensile strength of a final part printed by a laser sintering printer, given by Y_i, and the bulk density of the powder used, given by x_i, where (Y_i, x_i) is a paired observation. Also suppose that a simple linear regression model is fitted to find this relationship, that is, the linear model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \tag{3.1.2.1}$$

is fitted. Note that this model and any other linear regression model can also be written in matrix form, given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3.1.2.2}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, $\boldsymbol{X}$ is the design matrix with dimension $n \times p$ (for n observations and p parameters), the parameter vector $\boldsymbol{\beta}$ has length p and the model error is given by $\boldsymbol{\epsilon}$. In the case of the simple linear model, the design matrix has $p = 2$ columns, with the first column being a column of 1's, and the second column being $(x_1, \ldots, x_n)^T$, and the parameter vector has length $p = 2$, with $\boldsymbol{\beta} = (\beta_0, \beta_1)$.

As a non-expert in this field without any initial knowledge on the relationship, I would choose an uninformative prior for the intercept β_0 and gradient of the straight line β_1 which defines the relationship between tensile strength and bulk density, such that the prior distribution could take negative and positive values. Then, suppose I contact an expert for more

information, and they tell me that there is without a doubt a positive relationship between these two variables, meaning that, as bulk density is increased, so too is tensile strength. In this case, I would elect to change the parameters on my initial choice of prior distribution, so that it could only take positive values—in this case, an informed prior is being used, which is derived based on the information elicited from the expert. This is an example of *elicitation*.

An important aspect regarding the choice of the prior distribution is *conjugacy*. Informally, a prior is a *conjugate* prior if the parametric form of the prior remains unchanged in its posterior distribution. For example, if a normal distribution is used as a prior for some parameter of interest, and a normal distribution is then produced for the posterior of this parameter, then the normal distribution is a conjugate prior distribution for the parameter. Whether this happens depends on two things—the parameter in question (specifically, the purpose of the parameter in the modelling scenario) and the distribution assumed for the data, $p(\boldsymbol{Y}|\theta)$.

In most situations, a conjugate prior distribution is preferred, as the complexity of the process of updating the prior distribution is reduced; it becomes a case of updating the parameters in the prior.

In the case that a model is complicated enough to require computational methods in order to calculate the posterior distribution, the need for a conjugate prior to be used is reduced. The deciding factor for which prior distribution to use comes down to personal choice, or at least carrying out the Bayesian modelling multiple times with different prior distributions and comparing the quality of the inference to make a decision.

3.1.3 Hierarchical modelling

Hierarchical modelling is a form of modelling required in many real-life scenarios, where the scenario being investigated is structured in a way that means multiple parameters in the modelling are related and are therefore dependent on one another through a hierarchical or multi-level structure. An example of hierarchical modelling could be an investigation of the flowability of some powder material, where the powder material has been manufactured multiple times by multiple companies. The material composition manufactured by company j has a 'general flowability metric' of θ_j, which is to be estimated. It would be fair to assume that there is a relationship between the flowability metrics and their estimates, and so each prior distribution on θ_j will be dependent on some shared parameter ϕ. Consequently, a hierarchical modelling structure is introduced, where the parameter ϕ is unknown

and requires its own prior distribution.

As a result, the terms *hyperparameter* and *hyperprior* distribution are introduced. That is, to investigate the parameter θ_j using Bayesian analysis requires some prior distribution $p(\theta_j)$, and if the prior distribution $p(\theta_j)$ itself is defined by some parameter ϕ, then the parameter ϕ is known as a *hyperparameter*. If the hyperparameter ϕ is unknown and requires a prior distribution, the prior distribution is defined as a *hyperprior* distribution.

The notion of 'exchangeability' is also key in a lot of Bayesian hierarchical modelling analyses. This notion effectively states that, with the same prior distribution for each θ_j, the ordering (or, in a sense, the labelling) of each θ_j is arbitrary. In other words, the joint distribution of (θ_1, θ_2) is equivalent to that of (θ_2, θ_1).

Moreover, many hierarchical models will assume conditional independence between observations and parameters that correspond to different groups. It is often fair to assume that observations from group j are conditionally independent from $\theta_{j'}$, for $j \neq j'$. In other words, learning about $\theta_{j'}$ does not provide any information about the observations for group j.

3.1.4 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) algorithms have been developed to carry out parameter estimation in situations where the posterior distribution is tricky to derive analytically. With these algorithms, the posterior distribution $p(\theta|\boldsymbol{Y})$ is constructed through sampling. With the definition of a Markov chain whose state space is equal to the parameter space of the model, the chain is built in such a way that the realisations of the chain are easy to sample and the stationary distribution of the chain is equal to the joint posterior distribution of the parameters. Formally, a Markov chain is defined as a sequence of random variables $\theta_1, \theta_2, \ldots$, such that the conditional distribution $p(\theta_t|\theta_1, \ldots, \theta_{t-1})$ can be rewritten as

$$p(\theta_t|\theta_1, \ldots, \theta_{t-1}) = p(\theta_t|\theta_{t-1})$$

for any given t. Informally, the conditional distribution of θ_t given all previous values in the sequence depends only on the value directly previous, θ_{t-1}.

For each iteration t, a 'proposal' or 'candidate' point θ^* is sampled from a proposal (or 'jumping') distribution (denoted as $J(\theta^*|\theta_{t-1})$ in Gelman et al. (2013), for example), and θ^* undergoes an acceptance/rejection rule step,

given by

$$\theta_t = \begin{cases} \theta^* & \text{with probability } \min(r, 1), \\ \theta_{t-1} & \text{otherwise.} \end{cases} \tag{3.1.4.1}$$

That is, the draw of θ at iteration t is equal to either the proposal θ^* or the previous draw θ_{t-1}. The equation for calculating the quantity r is dependent on the algorithm, but involves the ratio of the target distribution at θ^* to the target distribution at θ_{t-1}, where the target distribution at draw s, θ_s, is given by

$$p(\theta_s|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\theta_s)p(\theta_s)}{p(\boldsymbol{Y})} \tag{3.1.4.2}$$

using Bayes' theorem. For the Metropolis-Hastings algorithm, the general form of the acceptance probability $P(\theta^*|\theta_{t-1})$ is given by

$$P(\theta^*|\theta_{t-1}) = \min\left\{1, \mathrm{HR}(\theta^*|\theta_{t-1})\right\},$$

where HR is the Hastings ratio, defined to be

$$\mathrm{HR}(\theta^*|\theta_{t-1}) = \frac{J(\theta_{t-1}|\theta^*)p(\theta^*|\boldsymbol{Y})}{J(\theta^*|\theta_{t-1})p(\theta_{t-1}|\boldsymbol{Y})}.$$

The Hastings ratio can be simplified using Equation 3.1.4.2 and because of the symmetry of the proposal distribution.

A multivariate extension of the random walk Metropolis-Hastings algorithm is the Gibbs sampler. This algorithm is relevant for this work due to its ability to cope with high-dimensional hierarchical models. The sampler follows the Metropolis-Hasting algorithm roughly sketched out above, but considers a multidimensional parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. At each iteration of the sampler, each element $\boldsymbol{\theta}$ is drawn from and updated in turn, conditionally on the previous elements which have been drawn and updated in that iteration. That is, at each iteration t and for each subvector of $\boldsymbol{\theta}$, an ordering of the m subvectors is chosen, and the j^{th} element of the parameter vector at iteration t, $\theta_{t,j}$, is sampled from the conditional distribution given all the other components of $\boldsymbol{\theta}$. Defining the following vector (taken from Gelman et al. (2013))

$$\theta_{t-1,-j} = (\theta_{t,1}, \ldots, \theta_{t,j-1}, \theta_{t-1,j+1}, \ldots, \theta_{t-1,m}),$$

the conditional distribution of $\theta_{t,j}$ given all other components of θ is given by

$$p(\theta_{t,j}|\theta_{t-1,-j}, \boldsymbol{Y}).$$

The MCMC simulation is run for long enough, so that the Markov chain will reach its stationary distribution (i.e. the posterior distribution), and the subsequent values of the sequence will be samples from the stationary distribution. Then, if a large number of samples are drawn, these can be used to evaluate the posterior distribution.

In order to check that the output from the MCMC simulation is from the posterior distribution, the convergence to the posterior distribution can be checked using the potential scale reduction factor. This approach was proposed in Gelman & Rubin (1992) and developed further for the multivariate case (assessing convergence of the joint posterior distribution) in Brooks & Gelman (1998). The approach requires at least two parallel chains of MCMC output, as it requires a measure of between-chain variance. Taken from Gelman et al. (2013), each scalar estimand ψ has its simulation samples labelled as $\psi_{i,j}$, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$, with n being the number of samples in each chain, and m being the number of chains. Each chain is also split in half, so that non-stationarity in each chain can be checked simultaneously (poor mixing between two halves of a chain may still lead to appropriate levels of convergence). So, the value m is technically double the number of parallel chains, meaning the value n is technically half of the number of stored posterior samples for each chain. With the simulation samples $\psi_{i,j}$, the between- and within-sequence variances are calculated, given by

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\psi}_{\cdot j} - \bar{\psi}_{\cdot\cdot})^2, \tag{3.1.4.3}$$

where $\bar{\psi}_{\cdot j} = \frac{1}{n} \sum_{i=1}^{n} \psi_{i,j}$, and $\bar{\psi}_{\cdot\cdot} = \frac{1}{m} \sum_{j=1}^{m} \bar{\psi}_{\cdot j}$, and

$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2, \tag{3.1.4.4}$$

where $s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\psi_{i,j} - \bar{\psi}_{\cdot j})^2$. An estimate of the marginal posterior variance of the estimand ψ is calculated with a weighted average of W and B, given by

$$\widehat{\text{var}}(\psi|\boldsymbol{Y}) = \frac{n-1}{n} W + \frac{1}{n} B \tag{3.1.4.5}$$

The potential scale reduction factor is then estimated by

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}(\psi|\boldsymbol{Y})}{W}}, \tag{3.1.4.6}$$

which tends towards 1 as $n \to \infty$. This is because the expectation of W tends towards $\text{var}(\psi|\boldsymbol{Y})$ as $n \to \infty$. The interpretation of the potential scale reduction factor is that, in the limit $n \to \infty$, the estimate represents the possible scale reduction of the current distribution of ψ if the simulations were continued. A large value of the potential scale reduction factor would suggest that collecting more simulations from the MCMC may lead to a better approximation of the posterior distribution.

Moreover, an issue which also needs to be checked with the posterior distribution is the severity of the autocorrelation between the posterior samples. It is possible that the MCMC algorithm becomes stuck in a particular region of the posterior distribution, and large chunks of consecutive posterior samples can be very similar values, suggesting the posterior distribution has not been explored exhaustively and effectively. Some level of autocorrelation naturally occurs since MCMC is predicated on Markov chains, where there is dependence on the directly previous random variable. A measure of the severity of the autocorrelation is obtained by comparing the number of iterations with the effective sample size, which is defined in the following paragraphs. In order to increase effective sample size, an increase in the number of posterior samples drawn from the MCMC algorithm could be considered. Additionally, in order to avoid storing too many samples from the MCMC output, a thinning parameter, say k_{thin}, is used, which indicates that only every $k_{\text{thin}}^{\text{th}}$ sample is stored from the post-burn-in posterior samples. This can lead to overly long running times for the MCMC, in which case, an adjustment to the proposal distribution, or even changing the actual MCMC algorithm, could be considered.

The effective sample size as described in Gelman et al. (2013) is also described here. In very simple terms, the effective sample size of a sequence is an adjustment of the actual sample size of the sequence based on the autocorrelation between the samples of the sequence. Considering how effective the simulation average for the estimand ψ, $\bar{\psi}_{\cdot\cdot}$, is as an estimate of the posterior mean, formally defined as $\text{E}(\psi|\boldsymbol{Y})$, is a possible way of defining effective sample size. What follows is the asymptotic formula for the variance of this simulation average, given by

$$\lim_{n\to\infty} mn\text{var}(\bar{\psi}_{\cdot\cdot}) = \left(1 + 2\sum_{t=1}^{\infty}\rho_t\right)\text{var}(\psi|\boldsymbol{Y}), \tag{3.1.4.7}$$

where the term ρ_t is defined to be the autocorrelation of the sequence of posterior samples for ψ at lag t (the distance between two terms of this sequence is given by t). If zero autocorrelation in the sequence existed (each

value in the sequence was independent), then

$$\begin{aligned}\text{var}(\bar{\psi}_{..}) &= \text{var}\left(\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}\psi_{i,j}\right) = \frac{1}{(mn)^2}\sum_{j=1}^{m}\sum_{i=1}^{n}\text{var}(\psi_{i,j}) \\ &= \frac{1}{mn}\text{var}(\psi|\boldsymbol{Y}),\end{aligned} \tag{3.1.4.8}$$

each autocorrelation ρ_t at lag t would be 0, and the sample size would simply be mn. Given the existence of correlation, the effective sample size is defined to be

$$n_{\text{eff}} = \frac{mn}{1 + 2\sum_{t=1}^{\infty}\rho_t}, \tag{3.1.4.9}$$

that is, the ratio (from Equation 3.1.4.7) between the scale of the variance of the average of the simulations and the scale of the true variance of the estimand ψ.

In order to estimate the effective sample size, the correlations ρ_t observed between the simulations must be estimated. Firstly, the variogram V_t at each lag t is calculated, defined by

$$V_t = \frac{1}{m(n-t)}\sum_{j=1}^{m}\sum_{i=t+1}^{n}(\psi_{i,j} - \psi_{i-t,j})^2, \tag{3.1.4.10}$$

which determines the variance between the sequence of samples given some distance t between the samples. With this, the correlations ρ_t can be estimated by

$$\hat{\rho}_t = 1 - \frac{V_t}{2\widehat{\text{var}}(\psi|\boldsymbol{Y})}, \tag{3.1.4.11}$$

with the estimate of the variance of the the estimand ψ taken from Equation 3.1.4.5.

Summing these estimates of the correlations in order to estimate n_{eff} using Equation 3.1.4.9 cannot be done as for large values of t, the correlations become effectively random, so instead, a partial sum is calculated, from $t = 0$ up to some value T, the first odd value for which the sum $\hat{\rho}_{T+1} + \hat{\rho}_{T+2}$ is negative. Therefore, the estimate of the effective sample size is given by

$$\hat{n}_{\text{eff}} = \frac{mn}{1 + 2\sum_{t=1}^{T}\hat{\rho}_t}. \tag{3.1.4.12}$$

Both the effective sample size and the potential scale reduction factor are used frequently in the assessments of the MCMC output to ensure that

inference can subsequently be carried out on the output. If poor values are observed for the potential scale reduction factor (values greater than 1.1 for the potential scale reduction factor are typically suggested (Gelman et al. 2013) to be too large), then, as suggested above, more samples are drawn from the MCMC algorithm, and a longer burn-in period could also be recommended. For the estimate of effective sample size, the recommendation for appropriate values is $5m$ (Gelman et al. 2013), where m is twice the number of parallel chains. If this is not satisfied, it is again recommended to run the algorithm for longer and store more posterior samples.

Some examples of computing environments that are commonly utilised to construct posterior distributions are BUGS (Bayesian inference using Gibbs sampling, Lunn et al. (2009)), JAGS (just another Gibbs sampler, Depaoli et al. (2016)), Stan (Gelman et al. 2015), mcsim (Bois & Maszle 1997), PyMC (Salvatier et al. 2016) and HBC (Daumé III 2008). The work carried out in this thesis was performed using JAGS, which provides the tools for fitting hierachical Bayesian models.

3.1.5 Bayesian linear regression

In order to understand a Bayesian linear regression, an understanding of a classical linear regression is useful, and so an example is given here to demonstrate.

As in Section 3.1.2, suppose there is data available to investigate the relationship between the tensile strength of a final part printed by a laser sintering printer, and the bulk density of the powder used. In this experiment, for each sample of powder used, the bulk density of the powder is measured, then the sample of powder is used in the laser sintering machine to print some arbitrary 3D-printed object. Once the object is printed, its tensile strength is measured. The data are therefore paired observations (Y_i, x_i), where Y_i denotes the tensile strength of the final part for powder i, and x_i denotes the bulk density of powder i, for $i = 1, \ldots, n$ (the vectors $\boldsymbol{Y}$ and $\boldsymbol{x}$ correspond to $(Y_1, \ldots, Y_n)$ and $(x_1, \ldots, x_n)$ respectively). The simple linear regression model from Equation 3.1.2.1 is to be fitted to the data, that is, the model where ϵ_i is a random error term, or random noise, which takes the distribution $\epsilon_i \sim \mathrm{N}(0, \tau_\epsilon)$. For this example, it is assumed that τ_ϵ is known. Furthermore, the subscript ϵ is included to help distinguish between precision parameters more easily, as several are introduced. Note that, in this instance, the normal distribution is parameterised by the mean and the precision, where the precision is defined to be inverse variance, that is, for some variance σ^2 of a random variable, the precision τ of the random

variable is $\tau = \frac{1}{\sigma^2}$. For the majority of this thesis, the (univariate) normal distribution is parameterised in this way, but it will be explicitly clear otherwise. The method of fitting a simple linear model from Equation 3.1.2.1 using a classical (non-Bayesian) approach is to estimate values of β_0 and β_1 such that the vector of random errors, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$, is minimised. This is usually carried out with either the method of least squares, or maximum likelihood estimation (MLE) (details of these methods are omitted here), which both lead to the estimates

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x},$$

where $S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})$, $S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, and $\bar{Y}$ and $\bar{x}$ represent the sample means of $\boldsymbol{Y}$ and $\boldsymbol{x}$ respectively. As noted in Section 3.1, the classical approach to this model assumes that the parameters β_0 and β_1 are fixed, unknown constants that are estimated by $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively.

Transferring from a classical approach to a Bayesian approach, the parameters β_0 and β_1 are considered to be random variables instead of fixed constants, and instead of estimating the fixed constants using MLE or the method of least squares, the estimation of the random variables is carried out using prior distributions and, in the simplest sense, applying Bayes' Theorem, to create the posterior distributions of β_0 and β_1.

For this simple linear model, there are two parameters of interest (assuming τ_ϵ is known): β_0 and β_1. The prior distribution for these parameters would need some consideration, but this is not of direct interest here. Typically, the conjugate priors would be used, so for example

$$\beta_0 \sim \mathrm{N}(0, \tau_{\beta_0}),$$

$$\beta_1 \sim \mathrm{N}(0, \tau_{\beta_1}),$$

could be used, that is, assuming normal distributions for both parameters, with means 0 and some precisions τ_{β_0} and τ_{β_1} respectively. Defining $\boldsymbol{\beta} = (\beta_0, \beta_1)$, its joint prior distribution is given by

$$\boldsymbol{\beta} \sim \mathrm{N}_2(0, T_\beta),$$

withe T_β being the diagonal (precision) matrix with τ_{β_0} and τ_{β_1} as the diagonal elements. Again, for simplifying this example, it is assumed these

precisions are known. So, in order to estimate the posterior distribution, Bayes' Theorem is applied:

$$p(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{x},\tau_\epsilon,T_\beta) = \frac{p(\boldsymbol{Y}|\boldsymbol{x},\boldsymbol{\beta},\tau_\epsilon)p(\boldsymbol{\beta}|T_\beta)}{p(\boldsymbol{Y}|\boldsymbol{x})}. \tag{3.1.5.1}$$

The posterior distribution can be found by considering

$$p(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{x},\tau_\epsilon,T_\beta) \propto p(\boldsymbol{Y}|\boldsymbol{x},\boldsymbol{\beta},\tau_\epsilon)p(\boldsymbol{\beta}|T_\beta), \tag{3.1.5.2}$$

where

$$p(\boldsymbol{Y}|\boldsymbol{x},\boldsymbol{\beta},\tau_\epsilon) \propto \exp\left\{-\frac{\tau_\epsilon}{2}\left(\boldsymbol{Y}-X\boldsymbol{\beta}\right)^T\left(\boldsymbol{Y}-X\boldsymbol{\beta}\right)\right\},$$

and

$$p(\boldsymbol{\beta}|T_\beta) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^T T_\beta \boldsymbol{\beta}\right\}.$$

The two above distributions are only given up to proportionality (as well as the posterior distribution) since the interest is in factors which are not constant—any constant elements from the distributions, that do not depend on $\boldsymbol{\beta}$, would be found in the normalising constant, the distribution $p(\boldsymbol{Y}|\boldsymbol{x})$. Note that X is simply the design matrix of the linear regression, which, in the case of the simple linear model, is given by

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

It follows that Equation 3.1.5.2 is proportional to

$$p(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{x},\tau_\epsilon,T_\beta) \propto \exp\left\{-\frac{Q}{2}\right\},$$

where

$$Q = \boldsymbol{\beta}^T T_\beta \boldsymbol{\beta} + \tau_\epsilon\left(\boldsymbol{Y}-X\boldsymbol{\beta}\right)^T\left(\boldsymbol{Y}-X\boldsymbol{\beta}\right).$$

Hence the parameters of the posterior distribution of $\boldsymbol{\beta}$, its mean $\boldsymbol{m}$ and precision matrix P (knowing its posterior is normally distributed since the conjugate prior was used), can be found by rearranging Q into the form

$$Q = \left(\boldsymbol{\beta}-\boldsymbol{m}\right)^T P\left(\boldsymbol{\beta}-\boldsymbol{m}\right),$$

since this provides the general form of a multivariate normal distribution with mean $\boldsymbol{m}$ and precision matrix P. A useful equation for carrying this out is

$$\left(\boldsymbol{Y}-X\boldsymbol{\beta}\right)^T\left(\boldsymbol{Y}-X\boldsymbol{\beta}\right) = \left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^T\left(X^TX\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right) + S,$$

where

$$\hat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T \boldsymbol{Y}$$

and

$$S = \left(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{Y} - X\hat{\boldsymbol{\beta}}\right).$$

Thus, Q can be written as

$$\begin{aligned}
Q &= \boldsymbol{\beta}^T T_\beta \boldsymbol{\beta} + \tau_\epsilon \left(\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \left(X^T X\right) \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + S \right) \\
&= \boldsymbol{\beta}^T T_\beta \boldsymbol{\beta} + \tau_\epsilon \left(\left(\boldsymbol{\beta}^T X^T X - \hat{\boldsymbol{\beta}}^T X^T X\right) \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + S \right) \\
&= \boldsymbol{\beta}^T T_\beta \boldsymbol{\beta} + \tau_\epsilon \left(\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} - \boldsymbol{\beta}^T X^T X \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T X^T X \boldsymbol{\beta} + \hat{\boldsymbol{\beta}}^T X^T X \hat{\boldsymbol{\beta}} + S \right) \\
&= \boldsymbol{\beta}^T T_\beta \boldsymbol{\beta} + \tau_\epsilon \left(\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} - \boldsymbol{\beta}^T X^T \boldsymbol{Y} - \boldsymbol{Y}^T X \boldsymbol{\beta} \right) + K \\
&= \boldsymbol{\beta}^T T_\beta \boldsymbol{\beta} + \tau_\epsilon \boldsymbol{\beta}^T X^T X \boldsymbol{\beta} - \tau_\epsilon \boldsymbol{\beta}^T X^T \boldsymbol{Y} - \tau_\epsilon \left(\boldsymbol{\beta}^T X^T \boldsymbol{Y} \right)^T + K,
\end{aligned}$$

where the matrix K is just constant. Recognising this as a matrix quadratic form, this leads to completing the square in $\boldsymbol{\beta}$, giving

$$Q = \left(\boldsymbol{\beta} - \boldsymbol{m}\right)^T P \left(\boldsymbol{\beta} - \boldsymbol{m}\right),$$

with

$$P = T_\beta + \tau_\epsilon X^T X$$

and

$$\boldsymbol{m} = P^{-1} X^T \boldsymbol{Y},$$

that is, $\boldsymbol{\beta} | \boldsymbol{Y}, \boldsymbol{x}, \tau_\epsilon, T_\beta \sim \mathrm{N}_2(\boldsymbol{m}, P)$.

If a particular value needs to be chosen as an estimate for these parameters, the mode of the posterior distribution (often called the posterior mode) is suggested. In the case of using noninformative priors for β_0 and β_1 in a simple linear model, the posterior mode for β_0 and β_1 is equal to the estimates from the method of least squares and MLE.

3.2 Errors-in-variables problem

3.2.1 Simple linear model with errors-in-variables

Consider the simple linear model from Equation 3.1.2.1, where the model error ϵ_i assumes a normal distribution with mean 0 and precision τ_ϵ. The only error assumed in the model is in the y-direction; that is, the explanatory variable is assumed to be measured exactly, and ϵ_i measures the vertical

distance (error) between the observed data point Y_i and some straight line $\beta_0 + \beta_1 x$, for any real x. The assumption that the explanatory variable x_i is measured exactly greatly helps in simplifying the estimation of the parameters.

In many scenarios with real data, the explanatory variable is not measured exactly. For example, say the bulk density of different powder materials is to be measured. The experiment carried out to obtain the bulk density is conducted in the same way each time. The machine that measures the density (mass) of the powder does not measure it *exactly* (there could be several reasons for this: incorrectly calibrated/faulty, rounding error, user error), and so the measurements of density that are recorded have measurement error. The 'true' bulk density of the powder does exist, but it cannot be measured exactly (with the machine); the true bulk density is unobservable.

Taking the example above, let the observed data be X_i for powder material i. The observed data X_i are of the form

$$X_i = \tilde{X}_i + \delta_i,$$

where the variable $\tilde{X}_i$ is thought to be the 'true', unobservable bulk density for powder material i, and the random variable δ_i is measurement error, usually assumed to be normally distributed with mean 0 and some precision τ_δ. The change from the lower case, noise-free explanatory variable x_i to the upper case, noisy explanatory variable is because of the understanding of the two variables; the observed x_i are exact observations and are not a function of anything, whereas the observed X_i is a random variable (since it is the sum of the true explanatory variable and some measurement error random variable δ_i), and random variables are commonly written with capital letters to represent them.

If a simple linear model is applied to this example, using the observed data X_i as the explanatory variable, i.e.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

'attenuation' bias is observed in the estimate of the parameter β_1. Examining the situation in a classical setting, the estimate of β_1 using either the method of least squares or MLE is given by

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}},$$

where typically $S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{\boldsymbol{x}})^2$. Replacing the noise-free data x_i with the noisy data X_i, the expression S_{xx} becomes

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i + \delta_i - \bar{X})^2.$$

With some algebraic manipulation, it can be shown that the estimate of the parameter β_1 in the linear model is altered by a factor of

$$\frac{\frac{1}{\tau_\delta}}{\frac{1}{\tau_\delta} + \frac{1}{\tau}} = \frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma^2},$$

which has to lie between 0 and 1, since each term must be positive, which makes the denominator larger than the numerator, and so the estimate of β_1 is biased towards 0. This is discussed further in Fuller (1987).

Due to this attenuation bias, it is clear that the typical estimate of β_1 from the simple linear model is not adequate here, and so an alternative model is required. This leads to the *errors-in-variables* model—taking the example from Section 3.2.1, where the observed data are of the form

$$X_i = \tilde{X}_i + \delta_i,$$

the errors-in-variables model investigates the relationship between the response variable Y_i and the true, unobservable variable $\tilde{X}_i$, leading to the model

$$Y_i = \beta_0 + \beta_1\tilde{X}_i + \epsilon_i.$$

3.2.2 Measurement error on the response variable

Measurement error can also occur on the response variable. Although error is already accounted for in the y-direction with the model error term ϵ_i, it is possible that the observations on the response variable are known to have been measured with error. If some true, unobservable variable $\tilde{Y}_i$ is introduced in a similar fashion to the true value $\tilde{X}_i$ for the explanatory variable, this makes a clear distinction between the error in the model, ϵ_i, and the measurement error in the response variable, say η_i, where

$$Y_i = \tilde{Y}_i + \eta_i,$$

and $\eta_i \sim \mathrm{N}(0, \tau_\eta)$. The variable $\tilde{Y}_i$, like $\tilde{X}_i$, is thought of as some true value of the response variable for group i which cannot be observed. If a simple linear model is then applied to investigate the linear relationship between $\tilde{Y}_i$ and $\tilde{X}_i$, then the true relationship between the variables is being investigated.

3.2.3 Replicate measurements

In many cases, the data on both the response variable and the explanatory variable are collected with replicate measurements. Extending the example given in Section 3.1.5, suppose that the measurements of the bulk density of each powder material are carried out on several subsamples of the powder material. It is noted that the measurements of the bulk density for each subsample are not all the same, and therefore this case corresponds to some measurement error in the explanatory variable. For this situation, the notation is adjusted based on the belief that the powder material is designed to be consistent throughout, leading to the conclusion that there is a true, unobservable value of bulk density for the powder material, and the bulk density measurement for each subsample of powder is some noisy observation of this true value (as opposed to each subsample of powder having its own true, unobservable value). This leads to the notation for the observed data for the explanatory variable

$$X_{i,k} = \tilde{X}_i + \delta_{i,k},$$

where the subscript i refers to the powder material, and the subscript k refers to the subsample of powder. For example, the observed data point $X_{3,4}$ is the bulk density measured on the fourth subsample of the third powder material. The random error term is given by $\delta_{i,k} \sim \mathrm{N}(0, \tau_\delta)$, so it is assumed that the measurement error comes from the same distribution for each subsample of each powder material.

As in Section 3.2.2, the same scenario can be applied to the response variable, leading to the notation for the observed data for the response variable

$$Y_{i,j} = \tilde{Y}_i + \eta_{i,j},$$

where the subscript i again corresponds to the powder material, and the subscript j is used here for the j^{th} subsample of powder. It could be the case that the same subsamples of powder that are used to measure bulk density (the explanatory variable) and the subsamples of powder that are used to create the 3D-printed object (on which the tensile strength is measured, from the example in Section 3.1.5) is the same, in which case the subscript k could be used for the observed data on both variables. The notation given above covers any possible scenario where the subsamples for each variable are two different sets of subsamples.

This leads to the simple linear model with measurement error on both the response variable and the explanatory variable

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_i + \epsilon_i, \tag{3.2.3.1}$$

where $\epsilon_i \sim \mathrm{N}(0, \tau_\epsilon)$, and the observed data are of the form

$$Y_{i,j} = \tilde{Y}_i + \eta_{i,j}, \;\; X_{i,k} = \tilde{X}_i + \delta_{i,k}, \tag{3.2.3.2}$$

where $\eta_{i,j} \sim \mathrm{N}(0, \tau_\eta)$ and $\delta_{i,k} \sim \mathrm{N}(0, \tau_\delta)$.

3.3 Bayesian errors-in-variables regression

While the errors-in-variables model is completely justified for this situation, it is difficult to apply with a classical approach, with additional strong assumptions required to be able to estimate the model parameters—assuming known precisions, or assuming the ratio between the precisions τ_ϵ and τ_δ. In order to apply the errors-in-variables model with more ease, the approach can be transferred from a classical framework to a Bayesian framework. There are many examples in the literature of this being carried out theoretically, including Fang et al. (2017), Dellaportas & Stephens (1995), Muff et al. (2015), Rozliman et al. (2017) and de Castro et al. (2013). Some practical examples are given in Kelly (2007) (astronomy) and Denham et al. (2011) (ecology).

The model error term ϵ_i from Equation 3.2.3.1 is discussed briefly here. It is assumed that the $\epsilon_i \sim \mathrm{N}(0, \tau_\epsilon)$ are independent and identically distributed. In this instance of estimating the relationship between the true values of the response variable and the true values of the explanatory variable(s), it could be argued that it is too strong an assumption that these model error terms, or model discrepancy, are independent of one another.

One possible alteration to account for this would be to assume a Gaussian process for the model error terms, which is a natural way of accounting for possible dependencies between the model discrepancy (in particular, see Kennedy & O'Hagan (2001) for a suitable approach to this). This formulation fits very naturally into the more general Gaussian process modelling that is used later in this thesis in Section 3.5.2, and so is not considered further just now.

The application of errors-in-variables in a Bayesian setting is natural, as error and noise are inherently accounted for in a Bayesian setting, with the model parameters treated as random variables. The method for estimating the parameters β_0 and β_1 is similar to what it would be for a Bayesian simple linear model without errors-in-variables, like in Section 3.1.5, using an appropriate MCMC algorithm. As in the example in Section 3.1.5, the parameters β_0 and β_1 take normal prior distributions, and, without any prior knowledge, they assume a normal distribution with mean 0 and precisions

τ_{β_0} and τ_{β_1}. These precision parameters are hyperparameters which take uninformative gamma hyperprior distributions.

It is noted here that there are some additional parameters that need some attention—the true, unobservable value $\tilde{X}_i$, and the measurement error precisions τ_η and τ_δ. The primary decision to be made is whether these parameters should be fixed at a value or given a prior distribution and allowed to vary.

The first parameter to consider is $\tilde{X}_i$. Typically, an errors-in-variables model in a classical setting can take two forms—a structural model or a functional model. In the Bayesian setting, those options are both possible; the value $\tilde{X}_i$ could be fixed at a particular value (as in a functional model; see Fuller (1987)). Without knowledge of what this particular value should be, a structural model seems to make more sense, and given the Bayesian setting and the ability to account for error naturally, treating the true value as a random variable is justified. Of course, as a response variable in a regression model, the true value $\tilde{Y}_i$ is already treated as a random variable.

The value $\tilde{X}_i$ is therefore assumed to be a random variable and must take a prior distribution in this Bayesian approach. The prior used here is a normal distribution, with some mean $\mu_{\tilde{X}}$ and some precision $\tau_{\tilde{X}}$. The mean $\mu_{\tilde{X}}$ can then be thought of as the explanatory variable mean over all groups, and the precision $\tau_{\tilde{X}}$ is some measure of variability between groups. That means, within each group i, there are the observed data $X_{i,k}$, of the form

$$X_{i,k} = \tilde{X}_i + \delta_{i,k},$$

where the observed data are then also thought of as random variables. The measurement error $\delta_{i,k}$ takes mean 0 and precision τ_δ, and, conditioned on $\tilde{X}_i$, the observed data $X_{i,k}$ take a normal distribution with mean $\tilde{X}_i$ and precision τ_δ. So, the precision τ_δ corresponds to measurement error.

What remains to be considered are the precision parameters τ_δ, τ_η and $\tau_{\tilde{X}}$, which can again either take a fixed value or a prior distribution. If the precision is known, then a fixed value should be used. Otherwise, there are a multiple possibilities for the prior distribution. Typically, gamma distributions are used as the prior distribution for precision parameters, as the gamma distribution is bounded from below by 0, which ensures that any sampled value from the distribution is a valid precision. Another reason is that the gamma distribution is the conjugate prior for precision parameters, when considering normally distributed likelihoods for the data. An alternative is to utilise a uniform prior distribution on a corresponding standard deviation or variance, where each sampled value can be rearranged into the

appropriate precision parameter. These hyperprior distributions may be uninformative, or, with elicitation or some other appropriate prior knowledge, may be informed.

Firstly define the parameter vector $\boldsymbol{\theta}$ to be

$$\boldsymbol{\theta} = (\beta_0, \beta_1, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}, \tau_{\tilde{X}}, \tau_\epsilon, \tau_\delta, \tau_\eta, \tau_{\beta_0}, \tau_{\beta_1}), \tag{3.3.0.1}$$

where the true, unobservable values for the response variable and explanatory variable are written as $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_{n_g})$, and $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{n_g})$ respectively, where n_g is the number of groups (powders/materials in the example). Secondly, define the observed data Y as

$$Y = \begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n_r} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n_r} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n_g,1} & Y_{n_g,2} & \cdots & Y_{n_g,n_r} \end{pmatrix}, \tag{3.3.0.2}$$

and the observed data X as

$$X = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,n_c} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,n_c} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_g,1} & X_{n_g,2} & \cdots & X_{n_g,n_c} \end{pmatrix}, \tag{3.3.0.3}$$

where n_r and n_c are the number of replicate observations on each group for the response variable and for the explanatory variable respectively. Then, the conditional distribution of interest is

$$p(\boldsymbol{\theta}|Y, X), \tag{3.3.0.4}$$

the joint posterior distribution of $\boldsymbol{\theta}$. This contains all the information that is known about the model parameters given the data used in the model, with which the model can then be assessed and possibly compared with other fitted models. Given the multidimensional parameter vector and the hierarchical model structure, the Gibbs sampling algorithm is required in order to find the posterior distribution $p(\boldsymbol{\theta}|Y, X)$.

3.3.1 Extending the simple linear model for EIV Bayesian regression

The simple linear model in Equation 3.2.3.1 is further developed by adding other terms into the relationship between the response variable and the

explanatory variable. An example of how this can be carried out is with a polynomial function of the sole explanatory variable, e.g., the quadratic model

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_i + \beta_{11} \tilde{X}_i^2 + \epsilon_i. \tag{3.3.1.1}$$

Whereas the quadratic model in classical EIV regression requires some thought as to how the quadratic in $X_{i,k}$ affects the parameter estimates for β_0, β_1 and β_{11}, the quadratic model in EIV Bayesian regression is straightforward to fit, with the only further prior distributions required for β_{11} and the hyper-prior on its precision parameter. Since there is already a prior distribution on $\tilde{X}_i$, there is no prior needed for $\tilde{X}_i^2$ (any sample of $\tilde{X}_i$ during the MCMC is simply squared for $\tilde{X}_i^2$, so no further sampling is required).

It is also noted that more explanatory variables can be included in the modelling. Extending the example from Section 3.2.3, suppose the relationship between the tensile strength of a final part printed by a laser sintering machine and two explanatory variables is of interest, with the first explanatory variable being bulk density of a powder, and the second being the aeratability of a powder (how easy it is to pass air through a powder; details of how this is measured are not included in this example). Suppose again that the method for carrying out the experiments leads to the same data structure as before, with replicate measurements on multiple groups. Some attention is needed to describe the relationship between the measurement error on multiple explanatory variables.

In general, for two explanatory variables, we can write that

$$X_{1,i,k_1} = \tilde{X}_{1,i} + \delta_{1,i,k_1} \tag{3.3.1.2}$$

and

$$X_{2,i,k_2} = \tilde{X}_{2,i} + \delta_{2,i,k_2}, \tag{3.3.1.3}$$

where:

- the value X_{1,i,k_1} represents replicate measurement k_1 on material i for the first explanatory variable, where $k_1 = 1, \ldots, n_{c_1}$,
- the value X_{2,i,k_2} represents replicate measurement k_2 on material i for the second explanatory variable, where $k_2 = 1, \ldots, n_{c_2}$,
- the values $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$ represent the true, unobservable values for material i for the first and second explanantory variables respectively,
- the variable δ_{1,i,k_1} represents the measurement error for replicate measurement k_1 on material i for the first explanatory variable,

- the variable δ_{2,i,k_2} represents the measurement error for replicate measurement k_2 on material i for the second explanatory variable.

The term n_{c_1} is written as such as the 'c' represents covariate, and so n_{c_1} represents the number of replicate observations on the first covariate.

This problem is further defined depending on the understanding of the replicate measurements. Suppose that two unique sets of subsamples were utilised to carry out the replicate measurements on bulk density and on aeratability (i.e., one set was used for bulk density, and the other set for aeratability). Given that different subsamples are used to measure X_{1,i,k_1} and X_{2,i,k_2}, even for $k_1 = k_2$, there are no grounds to assume some relationship between the measurement error variables δ_{1,i,k_1} and δ_{2,i,k_2}, and so these variables take separate (marginal) distributions, leading to

$$\delta_{1,i,k_1} \sim \mathrm{N}(0, \tau_{\delta_1}) \tag{3.3.1.4}$$

and

$$\delta_{2,i,k_2} \sim \mathrm{N}(0, \tau_{\delta_2}), \tag{3.3.1.5}$$

where the parameters τ_{δ_1} and τ_{δ_2} are measurement error precision for the first and second explanatory variables respectively.

If the exact same set of subsamples of powder is used to find the replicate measurements for both explanatory variables, then it is fair to expect there could be some relationship in the measurement error for both explanatory variables. For example, knowing that a particular subsample of powder measures higher than expected for bulk density could inform the measurement of aeratability for the particular subsample of powder, hence a relationship is assumed between the measurement errors. If the set of subsamples of powder is the same for both explanatory variables (and for each material i), then the subscripts k_1 and k_2 can be replaced by the subscript k, the number of replicate observations for both explanatory variables can be written as n_c, and the observed data for the explanatory variables can be written as

$$\begin{pmatrix} X_{1,i,k} \\ X_{2,i,k} \end{pmatrix} = \begin{pmatrix} \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \end{pmatrix} + \begin{pmatrix} \delta_{1,i,k} \\ \delta_{2,i,k} \end{pmatrix}, \tag{3.3.1.6}$$

Moreover, a joint normal distribution is assumed for $\delta_{1,i,k}$ and $\delta_{2,i,k}$, that is,

$$\begin{pmatrix} \delta_{1,i,k} \\ \delta_{2,i,k} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, T_\delta \right), \tag{3.3.1.7}$$

where the matrix T_δ is the precision matrix, which describes the information about the variability within groups for each explanatory variable, and how

the replicate measurements carried out on the same subsample vary with respect to each other.

Some more unique scenarios may occur. One example could be that the first n_{c_1} replicate measurements on both explanatory variables (for each material i) are carried out on the same subsample of powder, but there are a further $n_{c_2} - n_{c_1}$ (with $n_{c_2} > n_{c_1}$) replicate measurements on both explanatory variables for which different subsamples of powder were used. In this case, for $k_1, k_2 = 1, \ldots, n_{c_1}$ with $k_1 = k_2$, it would be assumed that

$$\begin{pmatrix} \delta_{1,i,k_1} \\ \delta_{2,i,k_2} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, T_\delta \right),$$

and for $k_1, k_2 = n_{c_1} + 1, \ldots, n_{c_2}$, with $k_1 = k_2$, it would be assumed that

$$\delta_{1,i,k_1} \sim \mathrm{N}(0, \tau_{\delta_1})$$

and

$$\delta_{2,i,k_2} \sim \mathrm{N}(0, \tau_{\delta_2}),$$

with τ_{δ_1} and τ_{δ_2} being the diagonal elements of the matrix T_δ. Other unique scenarios could also be explained here, but the details all lead to the same instruction; if the same subsample of powder is used to measure more than one explanatory variable observation, then the measurement error on these observations should be related by a joint distribution, and otherwise, the measurement errors can take marginal distributions.

With the aim of fitting an EIV Bayesian regression model to seek the relationship between the response variable and the two explanatory variables, firstly considering just an additive linear predictor, the model of interest is written as

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_{1,i} + \beta_2 \tilde{X}_{2,i} + \epsilon_i. \tag{3.3.1.8}$$

Again, it is noted here that there are additional parameters to consider—the regression coefficient β_2, the true, unobservable values $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$ and the precision matrix T_δ for the joint measurement error vector $(\delta_{1,i,k}, \delta_{2,i,k})'$ (in the case that the measurement errors δ_{1,i,k_1} and δ_{2,i,k_2} are not related, consider the details in Section 3.3). It is clear that β_2 is treated in the same fashion as the other regression coefficients β_0 and β_1, with a normal prior distribution and some hyperprior on its precision hyperparameter. In the case of one explanatory variable, a normal prior distribution is placed on the true, unobservable value $\tilde{X}_i$, which suggests possibly placing a normal prior distribution on both $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$, leading to the question of whether to place a joint prior on these true values, or separate priors on each true

value. Given that in any modelling case, there is expected to be some relationship between two explanatory variables (if they could both relate to the response, they could in some way then relate to each other), it makes sense to place a joint prior distribution on the true values. It follows that

$$\begin{pmatrix} \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \end{pmatrix} \sim \mathrm{N}(\boldsymbol{\mu}_{\tilde{X}}, T_{\tilde{X}}), \tag{3.3.1.9}$$

where the vector $\boldsymbol{\mu}_{\tilde{X}} = (\mu_{\tilde{\boldsymbol{X}}_1}, \mu_{\tilde{\boldsymbol{X}}_2})$ is the mean vector, comprising the mean for the true values for the first explanatory variable $\tilde{\boldsymbol{X}}_1$ and the mean for the true values for the second explanatory variable $\tilde{\boldsymbol{X}}_2$, and the matrix $T_{\tilde{X}}$ is the precision matrix, which describes the information about the variability between groups for each explanatory variable, and how the two explanatory variables vary with respect to each other.

The final two parameters that need to be considered are the precision matrices T_δ and $T_{\tilde{X}}$, which represent some measure of the within-groups and between-groups variability respectively. In a similar way to how the precisions τ_η, τ_δ and $\tau_{\tilde{X}}$ can be dealt with multiple possible prior distributions, there are a multiple prior distribution options for these precision matrices. Typically, the Wishart distribution is used as the prior distribution for a precision matrix and can be thought of as the multidimensional equivalent to the gamma distribution in this sense, as it naturally produces valid precision matrices, and is the conjugate prior distribution for precision matrices, when considering normally distributed likelihoods for the data. The Wishart distribution is defined by some scale matrix and some degrees of freedom, that is, some random variable ∇ takes a Wishart distribution means

$$\nabla \sim \mathrm{Wishart}(\nabla_0, \nu).$$

The scale matrix ∇_0 itself must be a valid precision matrix, and any sample from this distribution has the same dimensionality as ∇_0. If ∇_0 is a $p \times p$ matrix, then the degrees of freedom ν must simply be greater than or equal to p.

Alternatively, a precision matrix could be created by inverting a covariance matrix built through standard deviation parameters and some correlation coefficient. This is discussed below, with further details in Browne et al. (2006) and again in Gelman et al. (2006).

As an example, take the 2×2 *covariance* matrix S for two variables X_1 and X_2, written as

$$S = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{pmatrix}. \tag{3.3.1.10}$$

By definition, the covariance of two variables X_1 and X_2 is

$$\mathrm{Cov}(X_1, X_2) = \rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2},$$

where σ_{X_1} and σ_{X_2} are the respective standard deviations of the variables X_1 and X_2, and ρ_{X_1,X_2} is the correlation between the variables X_1 and X_2. This gives an alternative way to write S,

$$S = \begin{pmatrix} \sigma^2_{X_1} & \rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} \\ \rho_{X_1,X_2}\sigma_{X_2}\sigma_{X_1} & \sigma^2_{X_2} \end{pmatrix}. \tag{3.3.1.11}$$

With this formulation of S, the precision matrix S^{-1} can be created by placing prior distributions on the parameters σ_{X_1}, σ_{X_2} and ρ_{X_1,X_2}, then forming the matrix S and inverting it. It is often the case that uniform prior distributions are used for the standard deviations and correlation coefficient (again, see Browne et al. (2006) and Gelman et al. (2006) for theory and discussion on this).

Extending the EIV Bayesian regression with two explanatory variables further, an 'interaction' term between the two explanatory variables can be included in the relationship between the response variable and the explanatory variables. This would produce the model

$$\tilde{Y}_i = \beta_0 + \beta_1\tilde{X}_{1,i} + \beta_2\tilde{X}_{2,i} + \beta_{12}\tilde{X}_{1,i}\tilde{X}_{2,i} + \epsilon_i \tag{3.3.1.12}$$

The only new term in this model is the coefficient β_{12}, which captures a level of 'interaction' between the two explanatory variables. Referring to the example with bulk density and the aeratability of a powder as the explanatory variables, this model takes into account the possibility that the effect of bulk density on tensile strength changes depending on the value of the aeratability (or vice versa; the effect of aeratability on tensile strength changes depending on the value of bulk density). Similarly to fitting the quadratic model from Equation 3.3.1.1, all that is required in order to fit this model is to place a prior distribution on β_{12} and a hyperprior on its precision parameter.

The model can then be extended to some finite number of explanatory variables n_p and all their potential powers and interactions terms, where the observed data are of the form

$$X_{m,i,k_m} = \tilde{X}_{m,i} + \delta_{m,i,k_m},$$

where X_{m,i,k_m} is the k_m^{th} replicate measurement on the i^{th} group for the m^{th} explanatory variable, and the true, unobservable value for each group

within each explanatory variable is given by $\tilde{X}_{m,i}$. The distribution(s) for δ_{m,i,k_m} is dependent on how the data are collected (see discussion earlier in this section). The subscript m for the explanatory variables varies from $m = 1, \ldots, n_p$, the subscript i for the number of groups varies from $i = 1, \ldots, n_g$, and the subscripts k_m for the number of replicate measurements on explanatory variable m vary from $k_m = 1, \ldots, n_{c_m}$.

3.3.2 Multivariate response for EIV Bayesian regression

Again, extending the example scenario in Section 3.1.5, suppose additionally that there is interest in the relationship between the Young's modulus of a final part printed by a laser sintering printer and the bulk density of the powder used in the printer. In this case, there are two potential response variables: tensile strength and Young's modulus. One possible method for investigating these relationships is to fit two EIV Bayesian regression models, and, under the assumption of a linear relationship between the response variables and the explanatory variable, the simple linear model from Equation 3.2.3.1 would be fitted. That is, the model

$$\tilde{Y}_{1,i} = \beta_{0,1} + \beta_{1,1}\tilde{X}_i + \epsilon_{1,i}, \tag{3.3.2.1}$$

where $\tilde{Y}_{1,i}$ is the true value of tensile strength for powder i, is fitted, as well as the model

$$\tilde{Y}_{2,i} = \beta_{0,2} + \beta_{1,2}\tilde{X}_i + \epsilon_{2,i}, \tag{3.3.2.2}$$

where $\tilde{Y}_{2,i}$ is the true value of Young's modulus for powder i. This would be a legitimate method for investigating these relationships, and posterior distributions for $\beta_{0,1}$ and $\beta_{1,1}$ should give a good understanding as to how tensile strength relates to bulk density, and posterior distributions for $\beta_{0,2}$ and $\beta_{1,2}$ should give a good understanding as to how Young's modulus relates to bulk density.

On the other hand, what if there is some relationship between tensile strength and Young's modulus? What if, knowledge of potentially good values for $\beta_{0,1}$ and $\beta_{1,1}$ informs potentially good values for $\beta_{0,2}$ and $\beta_{1,2}$? If that knowledge is available, then a bivariate response model should definitely be considered. That is, the model

$$\begin{pmatrix} \tilde{Y}_{1,i} \\ \tilde{Y}_{2,i} \end{pmatrix} = \begin{pmatrix} \beta_{01} + \beta_{11}\tilde{X}_i \\ \beta_{02} + \beta_{12}\tilde{X}_i \end{pmatrix} + \begin{pmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{pmatrix} \tag{3.3.2.3}$$

should be fitted. For the bivariate response model, it is assumed that there is some relationship between the model errors $\epsilon_{1,i}$ and $\epsilon_{2,i}$, whereas in the

previous case with fitting two univariate response models, there was no relationship assumed between $\epsilon_{1,i}$ and $\epsilon_{2,i}$. In other words, the bivariate response model error is distributed by a bivariate normal distribution, with mean 0 and some precision matrix T_ϵ, i.e.

$$\begin{pmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, T_\epsilon \right) \tag{3.3.2.4}$$

Moreover, the bivariate response model has many benefits over two univariate response models when considering the backward model, which is explained further in Section 3.6. In short, if the behaviour of the explanatory variable(s) given the response variables is considered, then knowing the relationship between the response variables has an impact on the behaviour of the explanatory variable(s) (in the same sense, if the relationship between one response variable and two explanatory variables is of interest, then it would make sense to fit a model including both explanatory variables at once, instead of separate models).

In the Bayesian setting, the multivariate response model is handled with relative ease. Prior distributions are placed on parameters of interest in an identical fashion to the univariate response models. The key difference, the relationship between the model error terms $\epsilon_{1,i}$ and $\epsilon_{2,i}$, is dealt with by assuming the bivariate normal distribution for the model error, where the prior on the precision matrix T_ϵ can be chosen from the possibilities mentioned in Section 3.3.1.

The final thing to consider is the observed data on the two response variables, which is the equivalent scenario to the observed data for the two explanatory variables that is discussed in Section 3.3.1. For two response variables, the replicate measurements are of the general form

$$\begin{pmatrix} Y_{1,i,j_1} \\ Y_{2,i,j_2} \end{pmatrix} = \begin{pmatrix} \tilde{Y}_{1,i} \\ \tilde{Y}_{2,i} \end{pmatrix} + \begin{pmatrix} \eta_{1,i,j_1} \\ \eta_{2,i,j_2} \end{pmatrix}, \tag{3.3.2.5}$$

where the subscripts $j_1 = 1, \ldots, n_{r_1}$ and $j_2 = 1, \ldots, n_{r_2}$ refer to the replicate measurements on the first and second response variables respectively. The terms η_{1,i,j_1} and η_{2,i,j_2} account for the measurement error and are therefore random variables. There are several possibilities for the data collection process which then influences the relationship between these measurement error random variables. Again, as in Section 3.3.1, this boils down to whether or not the measurements for the replicates Y_{1,i,j_1} and Y_{2,i,j_2} were found using the same subsample of powder. If the same subsample of powder was used

for both, then a joint distribution can be assumed such that

$$\begin{pmatrix} \eta_{1,i,j_1} \\ \eta_{2,i,j_2} \end{pmatrix} \sim \mathrm{N}(0, T_\eta). \tag{3.3.2.6}$$

If a different subsample of powder was used to measure the replicates Y_{1,i,j_1} and Y_{2,i,j_2}, then marginal prior distributions are assumed such that

$$\eta_{1,i,j_1} \sim \mathrm{N}(0, \tau_{\eta_1}) \tag{3.3.2.7}$$

and

$$\eta_{1,i,j_2} \sim \mathrm{N}(0, \tau_{\eta_2}). \tag{3.3.2.8}$$

Appropriate prior distributions for both scenarios are dealt with in an analogous fashion to the replicate measurement cases for the two explanatory variables in Section 3.3.1.

3.4 Evaluating and comparing model fits

It is clear from Sections 3.3.1 and 3.3.2 that there is a plethora of EIV Bayesian regression models that can be fitted to a data set, with varying numbers of explanatory variables, polynomial terms and interaction terms. Given that one of the crucial aims of this work is to identify which combination of explanatory variables gives the best understanding of the response variable(s), there needs to be some way of deciding on the best-fitting model, which means each model has to be evaluated in some way, to then be able to compare the models with one another. It needs to also be clear what is meant by 'best-fitting' model—is this the model which fits the data set the best, or is it the model which is best at predicting the response variable(s) for some new value(s) of the explanatory variable(s)?

3.4.1 Classical setting

Typically, within a classical regression, there are numerous metrics used to evaluate the fit of a linear regression model. The R^2 coefficient of determination gives a percentage of the information in the data that is explained by the model, calculated by

$$R^2 = \frac{SS_{total} - SS_r}{SS_{total}}, \tag{3.4.1.1}$$

where SS_{total} calculates some measure of the entire variance in the response data, and SS_r calculates some measure of the distance between the 'line

of best fit' and the response data points. The smaller the total squared distance between the line of best fit and the data points, the smaller the value of SS_r, and therefore the larger the value of R^2, suggesting that the model explains more of the variability within the data. Ideally a large value of R^2 is desired, suggesting that the model fits the data set well. A value of $R^2 = 1$ suggests that the model explains the entire variability in the data set, which guarantees that the model is overfitted (no statistical model can completely encapsulate the variability of a real-life scenario). The R^2 coefficient is a useful tool for evaluating the fit of a model to a data set, but it is not an ideal model comparison tool—the value of R^2 can only increase with more explanatory variables, polynomial terms and interaction terms included in the model, so using this metric to measure which models fits best, when comparing models with different explanatory variables, is not adequate and should not be carried out. It is noted that R^2 can be replaced with an adjusted R^2 coefficient which takes into account the number of explanatory variables within the model, which essentially punishes models for having too many explanatory variables.

In terms of comparing the fit of multiple models in a classical setting, there are again several options. An F-test can be used for comparing nested models, where a model is nested within another if a constraint being imposed on one of the model coefficients leads to the models being identical. For example, the simple linear model, $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, is nested within the quadratic model, $Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \epsilon_i$, as imposing $\beta_{11} = 0$ leads to the models being identical. The F-test is a hypothesis test which answers the question 'is it worthwhile adding (or removing) this term to (or from) the model?', or more specifically, 'is there enough of a decrease in the residual sum of squares SS_r to justify adding (or removing) this term?'. While this has some use, it is limited to comparing nested models, and carrying out F-tests using analysis of variance (ANOVA) can lead to different conclusions depending on the ordering of the hypothesis tests.

Alternatively, the log-likelihood is often used for comparing statistical models. In this sense, the data can be treated as information, and including more data (information) provides more evidence for the estimated parameters. This leads to its use in information criteria. Akaike's information criterion (AIC) is a suitable method for comparing model fits, which does not require the models to be nested, and is built on the idea of penalised likelihood. This method takes $(-2\times)$ the log-likelihood of the data at hand evaluated at the maximum-likelihood estimates of the parameters, and adds a penalty for models which take into account more explanatory variables.

That is,

$$\text{AIC} = -2\log(p(Y|\widehat{\boldsymbol{\theta}}_{\text{mle}})) + 2k, \tag{3.4.1.2}$$

where $\boldsymbol{\theta}$ is the vector of parameters to be estimated in the model (e.g., in the simple linear model assuming known variance/precision, $\boldsymbol{\theta} = (\beta_0, \beta_1)$), $\widehat{\boldsymbol{\theta}}_{\text{mle}}$ is the maximum-likelihood estimate of these parameters, and $\log(p(Y|\widehat{\boldsymbol{\theta}}_{\text{mle}}))$ then gives an indication of how likely the response data Y would come from a model with the estimated parameters $\widehat{\boldsymbol{\theta}}_{\text{mle}}$, where a larger value suggests a better model fit.

The smaller the value of AIC, the better the fit of the model (this is true of all 'information criterion' methods). This tool is effective for comparing models, but it should not be the sole method for evaluating the fit of the model (given that there is no limit to AIC, it is hard to say what value equates to a fit being 'good'). Another benefit of using AIC (and information criterion generally) in classical linear regression is the ability to compare models that are not 'nested'. A model M_1 is nested within a model M_2 if and only if setting one (or multiple) model coefficients equal to 0 in M_2 leads to a model which is equivalent to M_1. This is discussed further in Section 3.4.2.

There are visual methods for evaluating model fits, such as index plots, Q-Q plots, plots of residuals against fitted values, all of which check the assumptions made by the model are suitable for the data set. While these are informal checks, they can quickly give an indication as to whether the model that has been fitted is appropriate. If it is clear that an assumption is violated, the data may be transformed in some way to account for this.

3.4.2 Evaluating and comparing model fits in a Bayesian setting

While the measures discussed in Section 3.4.1 are effective for classical linear regression, they do not transfer directly to Bayesian regression. This can be explained by the key difference between classical linear regression and Bayesian linear regression—the parameters in the classical approach are unknown constants, and in the Bayesian approach they are random variables. The estimates for the parameters in the classical approach are point estimates, and in the Bayesian approach, the estimates for the parameters are the posterior distributions. While R^2 and AIC cannot directly be used to compare fits of Bayesian regression models, both can be adjusted to account for uncertainty in the posterior distributions of the parameters. The reasons for AIC not being directly applied to Bayesian models are discussed below, and alternative information criteria methods are provided. In this work,

Bayesian R^2 is not considered. It is also noted that the visual methods for checking model fits are still appropriate in the Bayesian setting.

There are multiple options for an alternative to AIC which is more appropriate for comparing the fits of Bayesian models. An example of this is deviance information criteria (DIC), which is very similar to AIC. The MLE estimate of the parameters, $\widehat{\theta}_{\text{mle}}$, is replaced with the posterior mean estimate of the parameters, $\widehat{\theta}_{\text{Bayes}}$, as written in Gelman et al. (2013), with some adjustments to the penalty k from Equation 3.4.1.2. Although this alternative can be applied with ease and is more appropriate than AIC, it is noted that this method is not fully 'Bayesian', as DIC is evaluated using only the posterior mean estimates of the parameters, so a method which can evaluate the model fit while taking into account the uncertainty in the posterior is preferred.

This leads to two further established metrics for comparing fits of Bayesian regression models: Watanbe-Akaike (or widely applicable) information criteria (WAIC) and leave-one-out cross-validation (LOO-CV, or, as often stated in this work, LOO-CV-IC, leave-one-out cross-validation information criterion). Both methods follow the same principles as AIC and DIC, in that they depend on some form of log-likelihood, but WAIC and LOO-CV also take into account the uncertainty in the posterior distribution for the parameters.

The method of choice in this work is to use LOO-CV-IC. A brief overview of LOO-CV-IC is provided here, with a more complete description found in Appendix A. As a general concept, the idea of cross-validation is to investigate the effectiveness of a statistical analysis of an entire data set by repeating the analysis on a subset of the data set (commonly referred to as the 'training' data), then seeing how well the remaining data points in the data set (the 'test' data) can be predicted. In essence it evaluates the out-of-sample predictive performance of the statistical model. A commonly used cross-validation technique is leave-one-out cross-validation, where 'leave-one-out' refers to training on $n - 1$ data points, leaving out one data point to be tested on, which is an exhaustive method in that the cross-validation is repeated n times, with each iteration using a unique data point as the test data.

The choice of whether to use WAIC or LOO-CV as a method to compare the fits of Bayesian models is somewhat unimportant, as, in the limit $n \to \infty$, the methods are asymptotically equal. In reality n is of course finite, so, with reference to Vehtari et al. (2017), LOO-CV using Pareto smoothed importance sampling (PSIS-LOO) is preferred due to robustness in this case with weak prior information or influential observations (where PSIS-LOO cross validation is an approximation of exact LOO-CV). The method of

PSIS-LOO is described in detail in Appendix A.2.

As noted in Appendix A.1, without knowing the distribution of the new pair of true values $(\tilde{Y}_{\text{new}}, \tilde{X}_{\text{new}})$ (considering just one response and one explanatory variable), the best way to estimate out-of-sample predictive performance of a fitted model is the exact LOO-CV-IC, with the adjustment to account for measurement error on both the response and the explanatory variable. This is given by

$$\begin{aligned} \text{lppd}_{\text{LOO-CV-EIV}} &= \sum_{i=1}^{n_g} \log(p_{(\text{post},-i)}(\tilde{Y}_i)) \\ &= \sum_{i=1}^{n_g} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(\tilde{Y}_{i,s}|\boldsymbol{\theta}_{-i,s}, \tilde{X}_{i,s})\right), \end{aligned} \tag{3.4.2.1}$$

where $\boldsymbol{\theta}_{-i,s}$ is joint sample s from the parameter vector posterior, having 'trained' the model without group i. Note that the subscript 'post' refers to the posterior density, i.e., the density of $\tilde{Y}_i$ given the posterior distribution.

The additional issue of still having to estimate the pair of true values $(\tilde{Y}_i, \tilde{X}_i)$ with group i as the test point provides further reason for using the approximate version of PSIS-LOO-CV-IC (Pareto-smoothed importance sampling LOO-CV-IC). As derived in Appendix A.2, this is given by

$$\widehat{\text{elpd}}_{\text{PSIS-LOO-CV,EIV}} = \sum_{i=1}^{n_g} \log\left(\frac{\sum_{s=1}^{S} w_{i,s} p(\tilde{Y}_{i,s}|\boldsymbol{\phi}_{i,s})}{\sum_{s=1}^{S} w_{i,s}}\right), \tag{3.4.2.2}$$

where each w_i represents a Pareto-smoothed weighted importance ratio (again, see Appendix A.2 for further details). Note that $\boldsymbol{\phi}_{i,s} = (\beta_{0,s}, \beta_{1,s}, \tilde{X}_{i,s}, \tau_{\epsilon,s})$ is introduced as a way of avoiding the condition $\tilde{Y}_{i,s}|\tilde{Y}_{i,s}$ which occurs when conditioning on the parameter vector θ.

Both statistics given in Equations 3.4.2.1 and 3.4.2.2 are multiplied by -2 so that the values are placed on the deviance scale.

The application of PSIS-LOO-CV-IC method in R uses the package `loo` (Vehtari et al. 2023), and specifically the function `loo` from the same package. The input for the function is a matrix of posterior samples of the log-likelihoods from the model, i.e., $\log(p(\tilde{Y}_{i,s}|\boldsymbol{\phi}_{i,s}))$. In the case of the EIV Bayesian regression model defined in Equation 3.2.3.1, this log-likelihood is the log of the probability density function of the normal distribution $\text{N}(\widetilde{\mu}_{i,s}, \tau_{\epsilon,s})$ evaluated at $\tilde{Y}_{i,s}$.

What remains to be detailed is how to choose which models should be compared using the LOO-CV-IC. In the case of having, say, seven explana-

tory variables, there are 128 possible combinations of the explanatory variables that lead to a linear regression model with an additive linear predictor (including the model with just an intercept term). Using well-established statistical software this can generally be carried out with relative ease, that is, to fit these models and compare them to choose which is the best-fitting model. In the case of this work, where bespoke code is written in order to fit these models, it is infeasible to be able to fit all 128 models and compare their model fits. An alternative to speed up the process and still carry out a legitimate process is to consider a form of stepwise regression. This is primarily used in classical linear regression with nested models (see Section 3.4.1), but is implemented here for the purpose of not having to fit 100+ models using bespoke coding. The stepwise regression considered here is backwards variable selection—this requires fitting the full model including all explanatory variables, then comparing this model fit with each of the model fits having removed one of the explanatory variables. Whichever model fit provides the best improvement in the model (according to LOO-CV-IC) is chosen to be the new 'full' model, and the process is repeated until the nested models provide no improvement on the full model. A backwards variable selection is preferred over the forwards variable selection (which starts with the null model, includes each explanatory variable in turn, chooses the model which provides the greatest improvement in model fit, and continues until no explanatory variables can be included in the model to improve the fit), so as to capture any possible combinations of variables that might not occur in forwards variable selection.

In the case of visual checks of errors-in-variables Bayesian regression models, there are multiple 'fitted' values plots that could be considered. Generally speaking, fitted plots give a visual understanding of how well the model 'fits' to the data. In a classical sense, fitted values are defined to be the values of the response variable that are produced by putting some explanatory variable values into the fitted model (that is, the model with the parameter estimates). In the case of a classical simple linear model, the fitted values are defined to be

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

where the values $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates of β_0 and β_1. These fitted values are plotted against the observed data for the response; if the model is a perfect fit for the data, all the points in the plot lie on $y = x$ (or fitted = observed).

In the case of a Bayesian errors-in-variables regression model, there are three 'fitted' plots to be considered. The first two plots are not typical 'fit-

ted' values plots—the posterior distributions for $\tilde{Y}_i$ and $\tilde{X}_i$ can be checked by considering how the posterior distribution of $\tilde{Y}_i$ compares with the observed $Y_{i,j}$, and how the posterior distribution of $\tilde{X}_i$ compares with the observed $X_{i,k}$. To take into account the uncertainty in the posterior, 95% credible intervals are provided for $\tilde{Y}_i$ and $\tilde{X}_i$. The fitted values (as in the typical understanding) for the Bayesian errors-in-variables regression model are redefined to be

$$\widehat{Y}_{i,s} = \beta_{0,s} + \beta_{1,s}\tilde{X}_{i,\text{post},s},$$

where the subscript s denotes posterior sample s, with $s = 1, \ldots, S$. Producing a fitted value for each sample (or at least a large subsample) from the posterior distributions provides the uncertainty that is captured in the posterior distribution. Then, taking the posterior samples of $\tilde{Y}_{i,s}$, a joint 95% credible ellipse over $\widehat{Y}_{i,s}$ and $\tilde{Y}_{i,s}$ can be plotted for each group, and these credible ellipses can be compared with the line $\widehat{Y} = \tilde{Y}$, on which these credible ellipses would roughly lie if the model is appropriate for the data. A plot demonstrating this is provided in Figure 3.1, based on the simulation example discussed in Section 4.2.3 (identical plot provided in Figure 4.12).

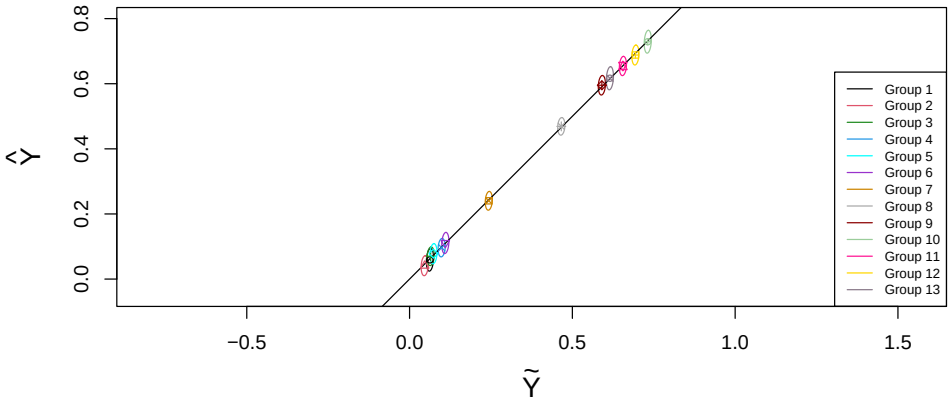

Figure 3.1: An example plot illustrating 95% credible ellipses over posterior samples $\widehat{Y}_{i,s}$ and $\tilde{Y}_{i,s}$ from a simulated data example.

It must be noted more specifically that there is a difference between the fitted values $\widehat{Y}_{i,s}$ and the posterior samples $\tilde{Y}_{i,s}$—the posterior of $\tilde{Y}_i$ can be estimated well given enough observations on the response variable for group i, and the same can be said of the posterior of $\tilde{X}_i$, but if these distributions have no relationship, the fitted values of the model are poor. The reason for

also considering the plots of the posterior of $\tilde{Y}_i$ against the observed data $Y_{i,j}$ and of the posterior of $\tilde{X}_i$ against the observed data $X_{i,k}$ is that these posteriors may be very poor, and at the same time the plot of the 95% credible ellipses of $\widehat{Y}_i$ and $\tilde{Y}_i$ (the fitted values vs. the true response values) may be very good. To get an idea whether the model fit is good, all three plots must be considered and must be 'good', i.e., close to the line $y = x$.

Finally, the plot of the 'fitted' model is also of interest. In a standard classical regression, with fitting a simple linear model, the fitted model is given by plotting the straight line estimated by the model, i.e. $y = \hat{\beta}_0 + \hat{\beta}_1 x$. This is typically plotted over the observed data (Y_i, x_i). In the case of an EIV model, specifically with observed data that do not have a one-to-one correspondence (the data point x_i does not 'produce' the data point Y_i), then the observed data are not plotted, and instead the true values for the response variable and explanatory variable, for each group i, are given by 95% credible ellipses estimated from the posterior distribution. The fitted model in a Bayesian regression is given by the posterior mean of the linear predictor, given by

$$\mathrm{E}_\theta(\beta_0 + \beta_1 x),$$

for some $x \in \mathbb{R}$. The uncertainty in the model fit is provided by 95% prediction intervals over the posterior distribution; that is, for some explanatory variable value $x \in \mathbb{R}$, the predictions of the response variable $\widehat{Y}_s$ are given by

$$\widehat{Y}_s = \beta_{0,s} + \beta_{1,s} x + \epsilon_s,$$

where the subscript s refers to the s^{th} joint posterior sample, and $\epsilon_s \sim \mathrm{N}(0, \tau_{\epsilon,s})$. This provides a distribution of predictions at x, whose mean is equal to the expectation $\mathrm{E}_\theta(\beta_0 + \beta_1 x)$ from above, and taking the 2.5% and 97.5% quantiles of $\widehat{Y}_s$ produces a centred 95% prediction interval for the response variable. This is carried out for a large range of explanatory variable values, whose posterior mean and 95% prediction intervals provide the fitted model.

The fitted model for the case of a model with one explanatory variable and one response variable is straightforward to visualise, with difficulties arising as models with more inputs and outputs are considered. With one response variable and two explanatory variables, contour plots and heatmaps can be utilised as 2D representations of the model—in these cases, displaying uncertainty in the model fit is difficult. Alternatively, a 3D plot can also be produced. When dealing with multiple response variable and multiple explanatory variables, the preference with this work is to produce predictions of the response variables jointly, then plot these distributions marginally

(i.e., plotting each response variable in turn). For example, in the case of two input variables, predictions of the response variables are produced for a straight line that intersects the 2D input space, which can be represented in a 2D plot with each response variable against one of the input variables.

3.5 Nonparametric modelling

The regression models found in Equations 3.1.2.1 and 3.2.3.1 are examples of linear modelling. In this case, the relationship between the response variable and explanatory variable is built through a linear combination of model coefficients $\beta_0, \beta_1, \ldots$, where the model is still said to be linear if nonlinear terms of the explanatory variable are involved in the relationship, that is,

$$y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \epsilon_i \tag{3.5.0.1}$$

is still a linear regression model. If the term β_1 were replaced by β_1^2, this would be a nonlinear regression model.

The simple linear model and the quadratic model are both examples of parametric regression—the relationship between the response variable and the explanatory variable in a linear regression is predetermined and clearly defined. For the simple linear model, this relationship is a straight line, and for the regression model in Equation 3.5.0.1, this relationship is a quadratic curve. There is no possibility of finding a quadratic relationship between the response variable and the explanatory variable when fitting a simple linear model. This can be seen as both an advantage and a disadvantage; the linear regression model is being restricted to a subset of potential relationships, and in many cases, this is appropriate, but if there is no initial understanding of the relationship between the response variable and the explanatory variable, a relationship between them may be forced when restricting to some parametric modelling method.

As an alternative, nonparametric regression modelling can be utilised. The advantage of nonparametric modelling is that there are almost no restrictions as to what relationship can be found between the response variable and the explanatory variable, so in the case where there is no knowledge of what to expect for the relationship, a nonparametric regression method is appropriate.

There are several methods of nonparametric regression—the method that has been focused on in this work is Gaussian processes.

3.5.1 Gaussian processes

A Gaussian process is a collection of random variables $Y_i = Y(\boldsymbol{x}_i)$ with $i = 1, 2, \ldots$, such that any finite subset of random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_n) = (Y(\boldsymbol{x}_1), \ldots, Y(\boldsymbol{x}_n))$ (for $\boldsymbol{x}_i \in \mathbb{R}^d$) has a multivariate normal distribution. Gaussian processes are discussed extensively in Williams & Rasmussen (2006); in this section, the case with one response variable and one explanatory variable is considered first, with $i = 1, \ldots, n$ defining the observations for both variables, and $\boldsymbol{x} = (x_1, \ldots, x_n)$.

It is often assumed that observations are made subject to noise, i.e. $Y_i = f(x_i) + \epsilon_i$ with $\epsilon_i \sim \mathrm{N}(0, \sigma_\epsilon^2)$. In this case, the Gaussian process for any finite set $\boldsymbol{Y}$ is given by

$$\boldsymbol{Y} \sim \mathrm{N}_n(\boldsymbol{m}(\boldsymbol{x}), V(\boldsymbol{x}, \boldsymbol{x})), \tag{3.5.1.1}$$

for some mean vector

$$\boldsymbol{m}(\boldsymbol{x}) = \begin{pmatrix} m(x_1) \\ m(x_2) \\ \vdots \\ m(x_n) \end{pmatrix}$$

and covariance matrix $V(\boldsymbol{x}, \boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}) + \sigma_\epsilon^2 I_n$, where

$$K(\boldsymbol{x}, \boldsymbol{x}) = \begin{pmatrix} k_{\mathrm{cov}}(x_1, x_1) & \cdots & k_{\mathrm{cov}}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k_{\mathrm{cov}}(x_n, x_1) & \cdots & k_{\mathrm{cov}}(x_n, x_n) \end{pmatrix}.$$

The function $k_{\mathrm{cov}}(\cdot, \cdot)$ is known as the covariance function, kernel function, or covariance kernel. Any function can be chosen to be k_{cov} as long as it produces a matrix K which is symmetric and positive semi-definite. An example which is often used is the squared exponential covariance function, which can be written in many forms, often as

$$k_1(x_i, x_{i'}) = \exp(-\frac{1}{2l^2}|x_i - x_{i'}|^2). \tag{3.5.1.2}$$

An alternative form of the squared exponential covariance function is

$$k_2(x_i, x_{i'}) = \exp(-l|x_i - x_{i'}|^2). \tag{3.5.1.3}$$

In both cases, l represents the 'length-scale' or 'distance-scaling' parameter, and is very important in GP modelling. This is discussed further below, under the heading 'Distance-scaling parameter'.

The covariance function is the main driving force in finding the relationship between $\boldsymbol{Y}$ and $\boldsymbol{x}$; its purpose is to suggest that values for the response variable are similar if the corresponding values of the explanatory variables are similar. Given that k_1 and k_2 have upper bounds at 1, the functions $k_3 = \sigma_k^2 k_1$ and $k_4 = \sigma_k^2 k_2$ can be used in an attempt to capture larger covariance (rather than just correlation).

Moreover, the matrix $\sigma_\epsilon^2 I_n$ represents some 'noise' variance at each data point x_i, where I_n is the $n \times n$ identity matrix. As in the linear regression models, and any non-deterministic model, we assume that the model cannot predict the response variable perfectly at each data point, and so σ_ϵ^2 induces some error in the GP at each x_i.

The mean function m is discussed less than the covariance kernel k_{cov} in the Gaussian process literature. This is generally because, if the mean function is assumed to be a constant, the covariance kernel dictates the behaviour of the Gaussian process completely, which makes managing the process easier. The mean function is also typically assumed to be 0, as the notation in the GP posterior is simplified (see Williams & Rasmussen (2006)). Furthermore, the choice of a non-constant mean function should be based on some assumption of the relationship between the response variable and the explanatory variable, which should be well informed. It could be argued that, in a Bayesian sense, the constant mean function corresponds to a incorporating no prior information into the mean function except a single overall mean value. Given that the GP can be well fitted without the need to consider a non-constant mean function, it is often more convenient to focus on the covariance kernel. In this work, the mean function is assumed to be some value α for any x, i.e. $m(x) = \alpha$. In the case of the mean vector $\boldsymbol{m}$, each element is set equal to α, and so $\boldsymbol{m}(\boldsymbol{x}) = \boldsymbol{\alpha} = (\alpha, \ldots, \alpha) \in \mathbb{R}^n$.

Like in a parametric regression, the method for estimating the unknown parameters in the Gaussian process can be carried out in a classical setting, using a method such as MLE. Given the desires in this work for working in a Bayesian environment, this is the preference here. This means placing prior distributions on the parameters of interest, namely σ_ϵ (standard deviation on the model error), α (constant value of mean function), l and, when k_3 or k_4 is considered as the covariance kernel, σ_k.

Distance-scaling parameter

The variable l is the most important variable of the Gaussian process when using the squared exponential kernel. It is the distance-scaling parameter of the function, which, if ignored, can lead to inappropriate values of the

covariance kernel, which leads to poor fits for the Gaussian process. For example, take the following set of explanatory variable data points:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 50 \\ 53 \\ 72 \\ 89 \\ 91 \end{pmatrix}.$$

The idea of the covariance kernel is to suggest that similar values for the explanatory variable corresponds to similar values of the response. Larger values of k_1 (note it is bounded above by 1) suggest that the two data points are very close, and that it is expected that the response values should be similarly close. Choosing $l = 1$ for the covariance kernel k_1 produces the covariance matrix

$$K = \begin{pmatrix} k_1(x_1, x_1) & k_1(x_1, x_2) & \cdots & k_1(x_1, x_5) \\ k_1(x_2, x_1) & k_1(x_2, x_2) & \cdots & k_1(x_2, x_5) \\ \vdots & \vdots & \ddots & \vdots \\ k_1(x_5, x_1) & k_1(x_5, x_2) & \cdots & k_1(x_5, x_5) \end{pmatrix}$$
$$= \begin{pmatrix} 1.00 & 0.01 & 0.00 & 0.00 & 0.00 \\ & 1.00 & 0.00 & 0.00 & 0.00 \\ & & 1.00 & 0.00 & 0.00 \\ & & & 1.00 & 0.14 \\ & & & & 1.00 \end{pmatrix},$$

to 2 decimal places. This matrix suggests that the data points x_1 and x_2 are hardly correlated, as with x_3 and x_5, despite them being relatively close together, given the (small) data set. Furthermore, the data point x_4 is closer to the pair of data points x_1 and x_2 than either of x_3 and x_5, yet the correlation between either x_1 or x_2 and x_4 is the same as between either x_1 or x_2 and either x_3 or x_5, i.e., there is no correlation between x_1 and x_4.

Choosing $l = 200$ for the covariance kernel k_1 produces the covariance matrix

$$K = \begin{pmatrix} 1.0000 & 0.9999 & 0.9940 & 0.9812 & 0.9792 \\ & 1.0000 & 0.9955 & 0.9839 & 0.9821 \\ & & 1.0000 & 0.9964 & 0.9955 \\ & & & 1.0000 & 1.0000 \\ & & & & 1.0000 \end{pmatrix},$$

to 4 decimal places. While there could be an argument that this is an improvement, with some distinction between the correlations, there is little difference between these correlations, like with the first example.

Finally, choosing $l = 20$ for the covariance kernel k_1 produces the covariance matrix

$$K = \begin{pmatrix} 1.000 & 0.989 & 0.546 & 0.149 & 0.122 \\ & 1.000 & 0.637 & 0.198 & 0.165 \\ & & 1.000 & 0.697 & 0.637 \\ & & & 1.000 & 0.995 \\ & & & & 1.000 \end{pmatrix}, \tag{3.5.1.4}$$

to 3 decimal places. Before considering any other information other than the five explanatory variable data points, this matrix seems like a good starting point for variability between the corresponding response data points. The value of $k_1(x_1, x_2) = 0.9888$ suggests that the corresponding response variable values for these two data points should be very similar; the value of $k_1(x_1, x_3) = 0.1223$ suggests the corresponding response variable values for these two data points should not be similar; the value of $k_1(x_1, x_4) = 0.5461$ suggests the corresponding response variable values for these two data points could be similar.

It it clear that the parameter l has an important role in determining how the response variable behaves. For the example above, it was clear that $l = 20$ was a better value for producing an appropriate covariance matrix K. Had the data been (500,530,910,720,890), then $l = 200$ would produce the same covariance matrix found in Equation 3.5.1.4. This leads to a clear relationship between the scale of the data and l, where the scale also helps determine the distance between the maximum and minimum data points in the data set. As far as the prior distribution is concerned for l, there are a couple of options, but in each case, one must be aware of the scale of the data. The uniform distribution is an option, where its lower bound must be greater than 0. To avoid having to explicitly specify the range of the uniform prior, a gamma distribution is an alternative, which always produces nonnegative values and can be skewed towards certain values, which is preferable here.

Other GP parameters

The GP parameters σ_ϵ, α and σ_k are considered here, specifically, their prior distibutions. Referring back to Section 3.1.2, conjugate priors for normally distributed data could be considered, while requiring placing a prior on some transformation of σ_ϵ; an inverse-gamma distribution is the conjugate prior for the variance, and the gamma distribution is the conjugate prior for the precision. The standard deviation parameter σ_ϵ on the model error can be dealt with in multiple ways, with, generally speaking, a uniform prior being

a common prior distribution. In some work, such as Lalchand & Rasmussen (2020), a lognormal distribution is utilised as the prior distribution for all non-noise hyperparameters, each with mean 0 and variance 3, considered to be vague hyperpriors. In this case (given the data set considered in Lalchand & Rasmussen (2020)), the hyperparameters were weakly identified, and so a gamma distribution was utilised for noise variance (i.e., σ_ϵ^2), although it is not clear how the gamma distribution is parameterised. The work in Flaxman et al. (2015) considers a lognormal prior also for the noise variance σ_ϵ^2, as well as for the signal variance parameter σ_k^2, both with mean 0 and variance 1.

The prior distribution on α is down to personal choice. As described above, it is common to fix this at 0. In the papers Lalchand & Rasmussen (2020) and Flaxman et al. (2015), there is no prior placed on the mean function. In this work, in the case of a weakly informed prior, a normal distribution is used, with mean 0.5 and variance 100. The reason for nonzero mean is due to the scaling of the data, and so the mean of the response variable a priori is the midpoint of the range [0,1]. In the case of fitting the fully Bayesian GP with informed priors, the mean of α is adjusted accordingly.

In this work, gamma prior distributions are considered for each of σ_ϵ, l and σ_k, with small shape values. This is due to familiarity with the distribution, and the scaling of the data, which causes relatively small values for each of these parameters.

GP posterior

Having chosen the prior distributions, the posterior distribution of the hyperparameters is estimated using MCMC. The priors are updated using the data, where the multivariate normal distribution is assumed for the response data points $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$.

Once the estimation of the unknown parameters has been performed, predictions of the response variable at certain values of the input variable can be carried out using the conditional Gaussian distribution, in other words, the Gaussian process posterior distribution. Analogously to the start Section 3.5.1, the GP posterior here is described for the case of a single input variable and single response variable.

Firstly, note that, for any vectors $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ with joint Gaussian distribution

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

the conditional distribution of $\boldsymbol{Y}_2|\boldsymbol{Y}_1$ is given by

$$\boldsymbol{Y}_2|\boldsymbol{Y}_1 \sim \mathrm{N}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1\right), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right). \tag{3.5.1.5}$$

Taking x_{new} to be a new observation for the explanatory variable and Y_{new} to be its corresponding value for the response variable to be predicted, the joint distribution for the GP, given known values of the hyperparameters, is given by

$$\begin{pmatrix} \boldsymbol{Y} \\ Y_{\text{new}} \end{pmatrix} \mid \boldsymbol{\phi} \sim \mathrm{N}_{n+1}\left(\begin{pmatrix} \boldsymbol{m}(\boldsymbol{x}) \\ m(x_{\text{new}}) \end{pmatrix}, \begin{pmatrix} V(\boldsymbol{x},\boldsymbol{x}) & \boldsymbol{k}(\boldsymbol{x}, x_{\text{new}}) \\ \boldsymbol{k}(x_{\text{new}},\boldsymbol{x}) & v(x_{\text{new}}, x_{\text{new}}) \end{pmatrix}\right)$$

where

$$\boldsymbol{k}(\boldsymbol{x}, x_{\text{new}}) = \begin{pmatrix} k_{\text{cov}}(x_1, x_{\text{new}}) \\ \vdots \\ k_{\text{cov}}(x_n, x_{\text{new}}) \end{pmatrix} = (\boldsymbol{k}(x_{\text{new}}, \boldsymbol{x}))^T, \tag{3.5.1.6}$$

$$v(x_{\text{new}}, x_{\text{new}}) = k_{\text{cov}}(x_{\text{new}}, x_{\text{new}}) + \sigma_\epsilon^2, \tag{3.5.1.7}$$

and

$$\boldsymbol{\phi} = (\alpha, l, \sigma_\epsilon, \sigma_k).$$

It is noted that the notation v is used for a scalar variance here (i.e., marginal, or one-dimensional variance parameter), as opposed to the more natural s. The reason for this choice is to avoid awkwardness when introducing posterior sample s of the marginal variance, thus $v(x_{\text{new}}, x_{\text{new}})$ is preferred to $s(x_{\text{new}}, x_{\text{new}})$. Similarly, the reason for the notation V to define the covariance matrices in the EIV GP (those which include model error variance along the diagonal elements) is because the more natural S is used to denote the number of posterior samples.

Then, the conditional distribution of $Y_{\text{new}}|\boldsymbol{Y}$, given known values of the hyperparameters and with the help of Equation 3.5.1.5, is given by

$$Y_{\text{new}}|\boldsymbol{Y}, \boldsymbol{\phi} \sim \mathrm{N}\left(m^*, v^*\right), \tag{3.5.1.8}$$

with

$$m^* = m(x_{\text{new}}) + \boldsymbol{k}(x_{\text{new}}, \boldsymbol{x})V(\boldsymbol{x},\boldsymbol{x})^{-1}(\boldsymbol{Y} - \boldsymbol{m}(\boldsymbol{x})), \tag{3.5.1.9}$$

and

$$v^* = v(x_{\text{new}}, x_{\text{new}}) - \boldsymbol{k}(x_{\text{new}}, \boldsymbol{x})V(\boldsymbol{x},\boldsymbol{x})^{-1}\boldsymbol{k}(\boldsymbol{x}, x_{\text{new}}). \tag{3.5.1.10}$$

Note again, the use of the notation v^*, as opposed to the more natural s^*, for the conditional variance term.

Since the model parameters are summarised by thousands of samples from the joint posterior distribution, it is most appropriate to use a subset of these posterior samples to produce estimates of the predictions of Y_{new}, since this takes into account the uncertainty in the posterior distribution. This is carried out as opposed to using just a point estimate from the posterior distributions, such as the mode, to give a single prediction of Y_{new}. In addition, predictions of the response can be carried out at multiple points for the explanatory variable, which can be generalised by treating both x_{new} and Y_{new} as vectors, with the conditional normal distribution working in the same fashion as above. An example of this is given in Section 3.5.3, also considering measurement error on the response variable and explanatory variables.

The Gaussian process defined above assumes that there is no error observed on the explanatory variable x_i. Some adjustments need to be made so that the Gaussian process can account for this. There are many examples in the literature of fully Bayesian Gaussian processes being implemented that discuss in further detail the choice of prior distribution for the parameters and hyperparameters in the model, including Lalchand & Rasmussen (2020), Flaxman et al. (2015), Frigola et al. (2013) and Oyebamiji et al. (2017).

3.5.2 Errors-in-variables Gaussian processes

Consider again the example given in Section 3.2.3, where the observed data that has been collected for the response variable is of the form

$$Y_{i,j} = \tilde{Y}_i + \eta_{i,j}$$

and for the explanatory variables is of the form

$$X_{i,k} = \tilde{X}_i + \delta_{i,k},$$

where $\eta_{i,j} \sim \text{N}(0, \tau_\eta)$ and $\delta_{i,k} \sim \text{N}(0, \tau_\delta)$. That is, it is known that the data for both the response variable and the explanatory variable has been observed with measurement error and collected on replicate measurements, with potentially different subsamples used for the response variable and for the explanatory variable (i.e., $Y_{1,1}$ was measured on a different subsample of powder to the subsample of powder used to measure $X_{1,1}$). The variables $\tilde{Y}_i$ and $\tilde{X}_i$ are true, 'unobservable' values of the response variable and the explanatory variable respectively, and the relationship between these values is to be investigated.

Coping with this measurement-error setup in Bayesian linear regression is relatively straightforward, with the natural treatment of the true values as random variables, as well as their respective observed data. The papers Cervone & Pillai (2015) and Zhou et al. (2019) provide methodology for dealing with noisy input variables in a Gaussian process regression, with some focus in Cervone & Pillai (2015) on a fully Bayesian methodology, and the work in Zhou et al. (2019) providing further theoretical justification for the fully Bayesian approach. The work in McHutchon & Rasmussen (2011) also defines a noisy-input Gaussian process, abbreviated to NIGP, which deals with the problem via an approximation and likelihood estimation. The paper Cervone & Pillai (2015) also references the work on developing the non-fully-Bayesian 'Kriging adjusting for location error', or KALE, shown in Cressie & Kornak (2003). Cervone & Pillai (2015) notes that the Gaussian process with location error in the input variable is not available in closed form, hence MCMC simulation is required. There appears to be no examples in the literature of also including measurement error on the output variable while fitting a Gaussian process regression. This work presents a fully Bayesian errors-in-variables Gaussian process regression with measurement error for both the output and input variable. This method is then applied in Chapter 4.

As defined in Section 3.3, the true, 'unobservable' values of the response variable and the explanatory variable are introduced as $\tilde{Y}_i$ and $\tilde{X}_i$ respectively, for $i = 1, \ldots, n_g$, with the observed data for each variable corresponding to noisy observations of the true values, that is,

$$Y_{i,j} = \tilde{Y}_i + \eta_{i,j},$$

and

$$X_{i,k} = \tilde{X}_i + \delta_{i,k},$$

where $\eta_{i,j} \sim \mathrm{N}(0, \tau_\epsilon)$, $\delta_{i,k} \sim \mathrm{N}(0, \tau_\delta)$, and τ_ϵ and τ_δ representing measurement error precisions. It is assumed that the collection of random variables $\tilde{Y}_i = f(\tilde{X}_i) + \epsilon_i$ with $i = 1, 2, \ldots$ is such that any finite subset of random variables $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_{n_g})$ has a multivariate normal distribution. The Gaussian process in this case is given by

$$\tilde{\boldsymbol{Y}} \sim \mathrm{N}_{n_g}(\boldsymbol{m}(\tilde{\boldsymbol{X}}), V(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}})),$$

for some mean vector

$$\boldsymbol{m}(\tilde{\boldsymbol{X}}) = \begin{pmatrix} m(\tilde{X}_1) \\ \vdots \\ m(\tilde{X}_{n_g}) \end{pmatrix}$$

and covariance matrix $V(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}) = K(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}) + \sigma_\epsilon^2 I_{n_g}$, where

$$K(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}) = \begin{pmatrix} k_{\text{cov}}(\tilde{X}_1, \tilde{X}_1) & \cdots & k_{\text{cov}}(\tilde{X}_1, \tilde{X}_{n_g}) \\ \vdots & \ddots & \vdots \\ k_{\text{cov}}(\tilde{X}_{n_g}, \tilde{X}_1) & \cdots & k_{\text{cov}}(\tilde{X}_{n_g}, \tilde{X}_{n_g}) \end{pmatrix}.$$

The model hyperparameters σ_ϵ, l, σ_k and the true values $\tilde{Y}_i$, $\tilde{X}_i$ for $i = 1, \ldots, n_g$ are estimated jointly, with the respective prior distrbutions for each parameter updated by the data to give the joint posterior distribution of the hyperparameters, whose posterior samples are used to predict the output variable given the GP posterior, analogously to that defined in Section 3.5.1. That is, the Gaussian process is trained on the true, unobservable values. The following section on EIV GPs with multiple explanatory variables has an example of prediction of the true value of the response at some vector of true values for two input variables subject to error.

3.5.3 Errors-in-variables Gaussian processes with multiple explanatory variables

The EIV GP developed in Section 3.5.2 can be extended to incorporate multiple explanatory variables. Considering again the example from Section 3.3.1, with two explanatory variables being the bulk density of a powder and the aeratability of a powder, where replicate measurements are recorded on multiple powders, leading to measurement error on the explanatory variables. This leads to the formulation of Equation 3.3.1.6. Suppose the response variable, represented by the true values $\tilde{\boldsymbol{Y}}$, is related to the two explanatory variables, represented by the true values $\tilde{\boldsymbol{X}}_1$ and $\tilde{\boldsymbol{X}}_2$, by a Gaussian process, that is,

$$\tilde{\boldsymbol{Y}} \sim \text{N}_{n_g}(\boldsymbol{m}(\tilde{X}), V(\tilde{X}, \tilde{X})), \tag{3.5.3.1}$$

where the matrix $\tilde{X}$ is defined as the matrix of true values for the explanatory variables, which, for this example with two explanatory variables, is given by

$$\tilde{X} = \begin{pmatrix} \tilde{\boldsymbol{X}}_1 & \tilde{\boldsymbol{X}}_2 \end{pmatrix} = \begin{pmatrix} \tilde{X}_{1,1} & \tilde{X}_{2,1} \\ \tilde{X}_{1,2} & \tilde{X}_{2,2} \\ \vdots & \vdots \\ \tilde{X}_{1,n_g} & \tilde{X}_{2,n_g} \end{pmatrix}. \tag{3.5.3.2}$$

Then, the mean vector $\boldsymbol{m}(\tilde{X})$ and covariance matrix $V(\tilde{X}, \tilde{X})$ are defined as

$$\boldsymbol{m}(\tilde{X}) = \begin{pmatrix} m(\tilde{\boldsymbol{X}}_{,1}) \\ \vdots \\ m(\tilde{\boldsymbol{X}}_{,n_g}) \end{pmatrix} = \begin{pmatrix} \alpha \\ \vdots \\ \alpha \end{pmatrix} \in \mathbb{R}^{n_g} \tag{3.5.3.3}$$

and $V(\tilde{X}, \tilde{X}) = K(\tilde{X}, \tilde{X}) + \sigma_\epsilon^2 I_{n_g}$, where

$$K(\tilde{X}, \tilde{X}) = \begin{pmatrix} k_{\text{cov}}(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_{\text{cov}}(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,n_g}) \\ \vdots & \ddots & \vdots \\ k_{\text{cov}}(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_{\text{cov}}(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,n_g}) \end{pmatrix}, \tag{3.5.3.4}$$

where the notation $\tilde{\boldsymbol{X}}_{,i}$ refers to the vector of true values for the two input variables for material i. Note that the mean function remains as a constant, so that the covariance matrix controls the entire variability in the GP.

The covariance functions for k_{cov} described in Section 3.5.1 need slight modification when considering multiple input variables, since the function must evaluate a pair of vectors instead of a pair of scalar points. For example, the squared exponential kernel k_1, which evaluates the two points x_i and $x_{i'}$ can be vectorised to give

$$k_{\mathbf{1}}(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) = \exp\left(-\frac{1}{2l^2}||\boldsymbol{x}_i - \boldsymbol{x}_{i'}||_2^2\right), \tag{3.5.3.5}$$

for two vectors $\boldsymbol{x}_i, \boldsymbol{x}_{i'} \in \mathbb{R}^2$, where $||\boldsymbol{x}||_2$ is defined to be the Euclidean norm, given by $||\boldsymbol{x}||_2 = \sqrt{x_1^2 + \cdots + x_p^2}$, for some $\boldsymbol{x} \in \mathbb{R}^p$, with $p = 2$ here. Therefore the function $k_{\mathbf{1}}$ can be rewritten as

$$k_{\mathbf{1}}(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) = \exp\left(-\frac{1}{2l^2}\sum_{d=1}^{2}(x_{i,d} - x_{i',d})^2\right). \tag{3.5.3.6}$$

In this form, $k_{\mathbf{1}}$ is an appropriate covariance kernel for a Gaussian process with two input variables, in that it will produce a positive definite covariance matrix. However, if the variables are not standardised appropriately, it is unlikely the estimate for the distance-scaling parameter l is suitable for both input variables. It would then be more appropriate to consider the squared exponential automatic relevance determination (SE-ARD) kernel, which vectorises the scalar parameter l, so that each input variable has its own distance-scaling parameter. So, the SE-ARD covariance kernel is

defined as

$$k_{\mathrm{ARD}}(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) = \exp\left(-\frac{1}{2}\sum_{d=1}^{2}\frac{(x_{i,d} - x_{i',d})^2}{l_d^2}\right), \tag{3.5.3.7}$$

for $\boldsymbol{l} = (l_1, l_2) \in \mathbb{R}^2$. As for the covariance kernel for one input variable, the covariance term σ_k^2 can be included as a scaling parameter in front of the exponential term, giving for example $k_{\mathbf{3}} = \sigma_k^2 k_{\mathbf{1}}$. For the example below, the function $k_{\mathbf{cov}}$ is the analogous covariance kernel function to some generic covariance kernel k_{cov} for multiple input variables.

Then, for some new true values of the explanatory variables given by $\tilde{\boldsymbol{X}}_{\mathrm{new}} = (\tilde{X}_{\mathrm{new},1}, \tilde{X}_{\mathrm{new},2})$, a prediction of its corresponding true value for the response variable $\tilde{Y}_{\mathrm{new}}$ using the GP posterior requires formulating the normal distribution

$$\begin{pmatrix}\tilde{\boldsymbol{Y}}\\ \tilde{Y}_{\mathrm{new}}\end{pmatrix} \sim \mathrm{N}_{n_g+1}\left(\begin{pmatrix}\boldsymbol{m}(\tilde{X})\\ m(\tilde{\boldsymbol{X}}_{\mathrm{new}})\end{pmatrix}, \begin{pmatrix}V(\tilde{X},\tilde{X}) & \boldsymbol{k}(\tilde{X},\tilde{\boldsymbol{X}}_{\mathrm{new}})\\ \boldsymbol{k}(\tilde{\boldsymbol{X}}_{\mathrm{new}},\tilde{X}) & v(\tilde{\boldsymbol{X}}_{\mathrm{new}},\tilde{\boldsymbol{X}}_{\mathrm{new}})\end{pmatrix}\right), \tag{3.5.3.8}$$

where

$$\boldsymbol{k}(\tilde{X}, \tilde{\boldsymbol{X}}_{\mathrm{new}}) = \begin{pmatrix}k_{\mathbf{cov}}(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{\mathrm{new}})\\ \vdots \\ k_{\mathbf{cov}}(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{\mathrm{new}})\end{pmatrix}, \tag{3.5.3.9}$$

$\boldsymbol{k}(\tilde{\boldsymbol{X}}_{\mathrm{new}}, \tilde{X}) = (\boldsymbol{k}(\tilde{X}, \tilde{\boldsymbol{X}}_{\mathrm{new}}))^T$, and $v(\tilde{\boldsymbol{X}}_{\mathrm{new}}, \tilde{\boldsymbol{X}}_{\mathrm{new}}) = k_{\mathbf{cov}}(\tilde{\boldsymbol{X}}_{\mathrm{new}}, \tilde{\boldsymbol{X}}_{\mathrm{new}}) + \sigma_\epsilon^2$. Then, the conditional distribution of $\tilde{Y}_{\mathrm{new}}|\tilde{\boldsymbol{Y}}, \tilde{X}, \boldsymbol{\phi}, \tilde{\boldsymbol{X}}_{\mathrm{new}}$, with the help of Equation 3.5.1.5, is given by

$$\tilde{Y}_{\mathrm{new}}|\tilde{\boldsymbol{Y}}, \tilde{X}, \boldsymbol{\phi}, \tilde{\boldsymbol{X}}_{\mathrm{new}} \sim \mathrm{N}(m^*, v^*), \tag{3.5.3.10}$$

where

$$m^* = m(\tilde{\boldsymbol{X}}_{\mathrm{new}}) + \boldsymbol{k}(\tilde{\boldsymbol{X}}_{\mathrm{new}}, \tilde{X})V(\tilde{X}, \tilde{X})^{-1}(\tilde{\boldsymbol{Y}} - \boldsymbol{m}(\tilde{X})) \tag{3.5.3.11}$$

and

$$v^* = v(\tilde{\boldsymbol{X}}_{\mathrm{new}}, \tilde{\boldsymbol{X}}_{\mathrm{new}}) - \boldsymbol{k}(\tilde{\boldsymbol{X}}_{\mathrm{new}}, \tilde{X})V(\tilde{X}, \tilde{X})^{-1}\boldsymbol{k}(\tilde{X}, \tilde{\boldsymbol{X}}_{\mathrm{new}}). \tag{3.5.3.12}$$

Furthermore, the vector $\boldsymbol{\phi}$ represents the joint posterior distribution of the model hyperparameters (excluding the true values $\tilde{\boldsymbol{Y}}$ and $\tilde{X}$), given by

$$\boldsymbol{\phi} = (\alpha, \sigma_\epsilon, \sigma_k, \boldsymbol{l}). \tag{3.5.3.13}$$

Note here the distinction between EIV GP posterior, which defines the conditional distribution of $\tilde{Y}_{\text{new}}$ given $\tilde{\boldsymbol{Y}}, \tilde{X}, \phi$ (i.e., the posterior predictive distribution), and the hyperparameter posterior distribution, which defines the posterior of the hyperparameter vector $\boldsymbol{\phi}$.

In order to take into account the uncertainty in the model when predicting the true value of the response, the values m^* and v^* are calculated for a large S subset of the posterior samples. This provides the distributions $\tilde{Y}_{\text{new},s}|\tilde{\boldsymbol{Y}}_s, \tilde{X}_s, \boldsymbol{\phi}_s, \tilde{\boldsymbol{X}}_{\text{new}}$, for $s = 1, \ldots, S$, as estimates of the distribution $\tilde{Y}_{\text{new}}|\tilde{\boldsymbol{Y}}, \tilde{X}, \boldsymbol{\phi}, \tilde{\boldsymbol{X}}_{\text{new}}$. Finally, combining a single sample from each distribution $\tilde{Y}_{\text{new},s}|\tilde{\boldsymbol{Y}}_s, \tilde{X}_s, \boldsymbol{\phi}_s, \tilde{\boldsymbol{X}}_{\text{new}}$ provides the complete distribution of $\tilde{Y}_{\text{new}}|\tilde{\boldsymbol{Y}}, \tilde{X}, \boldsymbol{\phi}, \tilde{\boldsymbol{X}}_{\text{new}}$.

This example with two explanatory variables demonstrates the EIV GP with multiple explanatory variables, and any case with more than two explanatory variables (and one response variable) follows this example in a similar fashion.

3.5.4 Evaluating and compare Gaussian process regression fits

As discussed in Section 3.4, the aim of model assessment is clear; to choose the best model, i.e., the model that best predicts the response of a future data point.

In Gaussian process regression, deciding which model is the best is often carried out by comparing some loss function, with the model that minimises the loss function being the best model. In general, some mean squared error function is used as the loss function, that is the function $\lambda(x)$ given by

$$\lambda(x) = (x - \hat{x})^2, \tag{3.5.4.1}$$

for some fixed $\hat{x} \in \mathbb{R}$. A common method for assessing Gaussian process regression models is to look at the root mean squared error with some form of cross-validation. There are numerous examples of this in the literature, including Oyebamiji et al. (2017), Xie et al. (2021), Amin et al. (2021) and Mukesh Kumar & Kavitha (2021).

Considering one of the desires of this work is to compare the parametric and nonparametric methods of EIV Bayesian regression and EIV Gaussian processes for the powder flow data, an approximate leave-one-out cross-validation information criterion (LOO-CV-IC) like what was considered in Section 3.4.2 is also considered here. Vehtari et al. (2016) discuss the implementation of a few approximate methods for applying LOO-CV-IC to

Gaussian latent variables models, with Gaussian process models as a specific example. Their approximations include importance sampling LOO, quadrature LOO, truncated quadrature LOO, Laplace approximation LOO and expectation propagation LOO. For this work, a mixed LOO-CV is applied; roughly speaking, the model hyperparameters of the EIV GP are estimated given all n_g groups, then the EIV GP posterior is fitted n_g times, each time training on $n_g - 1$ groups and testing on the remaining group. This is preferred here as it navigates any possible issues that can occur with PSIS-LOO-CV-IC, such as large estimates for the smoothing parameter of the generalised Pareto distribution.

As with the EIV BR, the full details of exact LOO-CV-IC and the approximate 'mixed LOO-CV-IC' for the EIV GP are found in Appendix A. The statistic used to evaluate EIV GPs using mixed LOO-CV-IC here is given by

$$\text{lppd}_{\text{mixed-LOO-CV-EIV}} = \sum_{i=1}^{n_g} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(\tilde{Y}_{i,s} | \boldsymbol{\phi}_s, \tilde{\boldsymbol{Y}}_{-i,s}, \tilde{\boldsymbol{X}}_s) \right), \quad (3.5.4.2)$$

with $\boldsymbol{\phi}_s$ being the s^{th} posterior sample of the EIV GP hyperparameters having been estimated given all n_g groups. The value in Equation 3.5.4.2 is multiplied by -2 so that the deviance scale is used.

3.5.5 Gaussian process with multivariate response

As in Section 3.3.2, where the example problem had been extended to consider two possible response variables and the multivariate EIV Bayesian regression had been considered, a multi-output extension of the Gaussian process is introduced here.

The extension of the Gaussian process to consider multiple response variables is achieved by applying a multi-output Gaussian process (MOGP). In the same vein as in Section 3.3.2, the two reasons for considering a multi-output EIV GP, as opposed to fitting multiple single-output EIV GPs, are that the EIV multi-output GP captures correlations between the response variables, which could increase model performance (Liu et al. 2018), and that considering the response variables in one model instead of multiple models is beneficial for the inverse problem, where the corresponding covariates can be optimised to produce the multivariate desired response (as opposed to being optimised multiple times for each element of the desired response vector).

The references to works in the literature regarding multi-output GPs all consider cases without measurement error on the response and explanatory variables. Similar cases of treating the input variables as random

variables are examined in the literature, typically considering uncertainty-quantification techniques in order to deal with any randonmness in the input variables. These techniques, such as stochastic collocation (see Kaintura et al. (2017)) and the spectral finite element method (see Bilionis & Zabaras (2012)), are not considered here, as these methods preemptively cope with uncertainty in input variables (as opposed to errors-in-variables methods, which account for *observed* data measured with error). The EIV MOGP, relating the true 'unobservable' values of multiple output variables to those of multiple input variables given observed data, is not covered in the literature (to the best of the author's knowledge). The errors-in-variables methodology here carries over from the single-output case with relative ease, and the description of relating the true values using a MOGP is covered below. Consequently, the descriptions are in terms of the true values of the output and input variables.

The multi-output Gaussian process can be split into two categories (which can be further split into subcategories)—symmetric and asymmetric multi-output Gaussian processes (see Liu et al. (2018)). The asymmetric multi-output GP is not of interest in this work. The symmetric multi-output GP corresponds to treating the first and second response variable equally. The subcategories of symmetric multi-output GPs are *separable models*, *process convolution* and *simple transformation models*, where the separable models are of interest here.

Before separable models are discussed here, the observed data and true values for the model are considered again, in the case of two response variable and two explanatory variables. That is, consider the observed data Y_{1,i,j_1}, Y_{2,i,j_2}, X_{1,i,k_1} and X_{2,i,k_2}, where

$$\begin{aligned} Y_{1,i,j_1} &= \tilde{Y}_{1,i} + \eta_{1,i,j_1}, \\ Y_{2,i,j_2} &= \tilde{Y}_{2,i} + \eta_{2,i,j_2}, \\ X_{1,i,k_1} &= \tilde{X}_{1,i} + \delta_{1,i,k_1}, \\ X_{2,i,k_2} &= \tilde{X}_{2,i} + \delta_{2,i,k_2}, \end{aligned} \tag{3.5.5.1}$$

where the observed data are considered to be noisy observations of the true values $\tilde{Y}_{1,i}$, $\tilde{Y}_{2,i}$, $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$ respectively, with η_{1,i,j_1}, η_{2,i,j_2}, δ_{1,i,k_1} and δ_{2,i,k_2} as the corresponding measurement errors. Note again that the subscript i refers to the group/material, the subscript j_1 refers to the subsample of group/material i for the first response, j_2 refers to the subsample of group/material i for the second response, k_1 refers to the subsample of group/material i for the first covariate, and k_2 refers to the subsample of group/material i for the second covariate.

There are multiple methods for defining a separable model for the MOGP; the first covered here is an intrinsic coregionalisation model (ICM), defined in Liu et al. (2018), which assumes that (for two output variables), that the first and second outputs are each a linear transformation of some Gaussian process $u^1\left(\tilde{\boldsymbol{X}}_i\right) \sim \mathrm{N}(0, k_1(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{X}}_{i'}))$, that is,

$$f_1(\tilde{\boldsymbol{X}}_i) = a_1^1 u^1\left(\tilde{\boldsymbol{X}}_i\right) \tag{3.5.5.2}$$

and

$$f_2(\tilde{\boldsymbol{X}}_i) = a_2^1 u^1\left(\tilde{\boldsymbol{X}}_i\right). \tag{3.5.5.3}$$

By writing the two outputs as a vector-valued function

$$\boldsymbol{f}(\tilde{\boldsymbol{X}}_i) = (f_1(\tilde{\boldsymbol{X}}_i), f_2(\tilde{\boldsymbol{X}}_i))$$

and writing the scalars as the vector $\boldsymbol{a} = \begin{pmatrix} a_1^1 & a_2^1 \end{pmatrix}^T$, the covariance of $\boldsymbol{f}(\tilde{\boldsymbol{X}}_i)$ and $\boldsymbol{f}(\tilde{\boldsymbol{X}}_{i'})$ is given by

$$\mathrm{cov}(\boldsymbol{f}(\tilde{\boldsymbol{X}}_i), \boldsymbol{f}(\tilde{\boldsymbol{X}}_{i'})) = \boldsymbol{a}\boldsymbol{a}^T k_1(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{X}}_{i'}). \tag{3.5.5.4}$$

This is derived using the definition of the covariance in terms of the difference between the expectation of the products and the products of the expectations.

Assuming that, for $i = 1, \ldots, n_g$,

$$\tilde{\boldsymbol{Y}}_{1,i} = f_1(\tilde{\boldsymbol{X}}_i) + \epsilon_{1,i}$$

and

$$\tilde{\boldsymbol{Y}}_{2,i} = f_2(\tilde{\boldsymbol{X}}_i) + \epsilon_{2,i},$$

where $\epsilon_{1,i} \sim \mathrm{N}(0, \sigma^2_{\epsilon_1})$ and $\epsilon_{2,i} \sim \mathrm{N}(0, \sigma^2_{\epsilon_2})$, then the true values $(\tilde{\boldsymbol{Y}}_1, \tilde{\boldsymbol{Y}}_2)'$ are multivariate normally distributed, that is

$$\begin{pmatrix} \tilde{\boldsymbol{Y}}_1 \\ \tilde{\boldsymbol{Y}}_2 \end{pmatrix} \sim N_{2n_g}\left(\begin{pmatrix} m(\tilde{X}) \\ m(\tilde{X}) \end{pmatrix}, \begin{pmatrix} (a_1^1)^2 K_1(\tilde{X}, \tilde{X}) + \sigma^2_{\epsilon_1} I_{n_g} & a_1^1 a_2^1 K_1(\tilde{X}, \tilde{X}) \\ a_2^1 a_1^1 K_1(\tilde{X}, \tilde{X}) & (a_2^1)^2 K_1(\tilde{X}, \tilde{X}) + \sigma^2_{\epsilon_2} I_{n_g} \end{pmatrix}\right), \tag{3.5.5.5}$$

where $\tilde{X} = (\tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2)$, and so $m(\tilde{X}) = \alpha \in \mathbb{R}^{n_g}$. In some cases, it is appropriate to use unique mean functions for the response variables, i.e., assuming that $\tilde{\boldsymbol{Y}}_1$ has mean $m_1(\tilde{X}) = \alpha_1 \in \mathbb{R}^{n_g}$ and $\tilde{\boldsymbol{Y}}_2$ has mean $m_2(\tilde{X}) = \alpha_2 \in \mathbb{R}^{n_g}$. Moreover, the covariance matrix for the Gaussian process has been divided into four block matrices denoted by $(a_1^1)^2 K_1(\tilde{X}, \tilde{X}) + \sigma^2_{\epsilon_1} I_{n_g}$, $a_1^1 a_2^1 K_1(\tilde{X}, \tilde{X})$, $a_2^1 a_1^1 K_1(\tilde{X}, \tilde{X})$ and $(a_2^1)^2 K_1(\tilde{X}, \tilde{X}) + \sigma^2_{\epsilon_2} I_{n_g}$, which are all $n_g \times n_g$ matrices.

This follows from the covariance between the two vectors $\boldsymbol{f}(\tilde{\boldsymbol{X}}_i)$ and $\boldsymbol{f}(\tilde{\boldsymbol{X}}_{i'})$ in Equation 3.5.5.4. As before, the matrix $K_1(\tilde{X}, \tilde{X})$ is the covariance kernel matrix given by

$$K_1(\tilde{X}, \tilde{X}) = \begin{pmatrix} k_1(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_1(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,n_g}) \\ \vdots & \ddots & \vdots \\ k_1(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_1(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,n_g}) \end{pmatrix} \tag{3.5.5.6}$$

The multi-output EIV GP can be defined more succinctly using the Kronecker product. For two matrices $A \in \mathbb{R}^{a_1 \times a_2}$ and $B \in \mathbb{R}^{b_1 \times b_2}$, where the matrix A is given by

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,a_2} \\ \vdots & \ddots & \vdots \\ a_{a_1,1} & \cdots & a_{a_1,a_2} \end{pmatrix}, \tag{3.5.5.7}$$

the Kronecker product $A \otimes B \in \mathbb{R}^{a_1 b_1 \times a_2 b_2}$ is defined as

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \cdots & a_{1,a_2}B \\ \vdots & \ddots & \vdots \\ a_{a_1,1}B & \cdots & a_{a_1,a_2}B \end{pmatrix}. \tag{3.5.5.8}$$

This allows for the multi-output EIV GP to be defined as

$$\begin{pmatrix} \tilde{\boldsymbol{Y}}_1 \\ \tilde{\boldsymbol{Y}}_2 \end{pmatrix} \sim N_{2n_g}\left(\begin{pmatrix} m(\tilde{X}) \\ m(\tilde{X}) \end{pmatrix}, \boldsymbol{a}\boldsymbol{a}^T \otimes K_1(\tilde{X}, \tilde{X}) + \begin{pmatrix} \sigma^2_{\epsilon_1} I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma^2_{\epsilon_2} I_{n_g} \end{pmatrix} \right), \tag{3.5.5.9}$$

with the matrix $0_{n_g \times n_g}$ being a matrix of zeroes with n_g rows and n_g columns. The ICM can be further extended to taking multiple samples from the Gaussian process $u^1\left(\tilde{\boldsymbol{X}}_i\right)$, and writing each functional relationship as a unique linear combination of the samples. The matrix $\boldsymbol{a}\boldsymbol{a}^T$ is redefined here to be

$$V_K = \begin{pmatrix} \sigma^2_{k,1} & \sigma_{k,1}\sigma_{k,2}\rho_{V_K} \\ \sigma_{k,1}\sigma_{k,2}\rho_{V_K} & \sigma^2_{k,2} \end{pmatrix}, \tag{3.5.5.10}$$

which effectively represents the covariance between the output variables.

The ICM provides the simplest case of the separable MOGP. The model assumes that the same covariance function captures the functional relationship between the first response variable and the explanatory variables, and between the second response variable and the explanatory variables. This

could be seen as advantageous in the sense of reducing the number of hyperparameters that need to be estimated, but is typically inappropriate if the relationships between each response variable and the input variables are sufficiently different. For example, in the case that the first input variable has a large influence on the first output variable (that is, the first response variable is very sensitive to small changes in the first input variable), then the corresponding distance-scaling parameter l_1 will be very small. However, if it is also the case that the first input variable has a small influence on the second output variable (that is, the second response variable is **not** very sensitive to small changes in the first input variable), then l_1 will be relatively much larger, making it infeasible to find values of l_1 that are suitable for the relationship between both the first input and the first output, and the first input and second output.

This can be dealt with by introducing more covariance kernels into the MOGP model, which extends the ICM to a semiparametric latent factor model (Teh et al. (2005), abbreviated to SPLF), which introduces a second Gaussian process $u^2\left(\tilde{\boldsymbol{X}}_i\right) \sim \mathrm{N}(0, k_2(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{X}}_{i'}))$, and defines the functions $f_1(\tilde{\boldsymbol{X}}_i)$ and $f_2(\tilde{\boldsymbol{X}}_i)$ as a linear combination of the two Gaussian processes $u^1\left(\tilde{\boldsymbol{X}}_i\right)$ and $u^2\left(\tilde{\boldsymbol{X}}_i\right)$, that is,

$$f_1(\tilde{\boldsymbol{X}}_i) = a_{1,1}u^1\left(\tilde{\boldsymbol{X}}_i\right) + a_{2,1}u^2\left(\tilde{\boldsymbol{X}}_i\right) \tag{3.5.5.11}$$

and

$$f_2(\tilde{\boldsymbol{X}}_i) = a_{1,2}u^1\left(\tilde{\boldsymbol{X}}_i\right) + a_{2,2}u^2\left(\tilde{\boldsymbol{X}}_i\right). \tag{3.5.5.12}$$

By writing the scalars as the vectors $\boldsymbol{a}_1 = \begin{pmatrix} a_{1,1} & a_{1,2} \end{pmatrix}^T$ and $\boldsymbol{a}_2 = \begin{pmatrix} a_{1,1} & a_{1,2} \end{pmatrix}^T$, the covariance of $\boldsymbol{f}(\tilde{\boldsymbol{X}}_i)$ and $\boldsymbol{f}(\tilde{\boldsymbol{X}}_{i'})$ is given by

$$\mathrm{cov}(\boldsymbol{f}(\tilde{\boldsymbol{X}}_i), \boldsymbol{f}(\tilde{\boldsymbol{X}}_{i'})) = \boldsymbol{a}_1(\boldsymbol{a}_1)^T k_1(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{X}}_{i'}) + \boldsymbol{a}_2(\boldsymbol{a}_2)^T k_2(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{X}}_{i'}). \tag{3.5.5.13}$$

Similarly to Equation 3.5.5.4, this is derived by defining to the covariance as the expectation of the products minus the products of the expectations (see Appendix). Then, the semiparametric latent factor model for the EIV MOGP is defined analogously to Equation 3.5.5.9, replacing

$$\boldsymbol{a}\boldsymbol{a}^T \otimes K_1(\tilde{X}, \tilde{X}) + \begin{pmatrix} \sigma_{\epsilon_1}^2 I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma_{\epsilon_2}^2 I_{n_g} \end{pmatrix}$$

with

$$\boldsymbol{a}_1(\boldsymbol{a}_1)^T \otimes K_1(\tilde{X}, \tilde{X}) + \boldsymbol{a}_2(\boldsymbol{a}_2)^T \otimes K_2(\tilde{X}, \tilde{X}) + \begin{pmatrix} \sigma_{\epsilon_1}^2 I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma_{\epsilon_2}^2 I_{n_g} \end{pmatrix}.$$

The matrices $K_1(\tilde{X}, \tilde{X}), K_2(\tilde{X}, \tilde{X})$ are the covariance kernel matrices, measuring spatial correlation between the input variables. Each element of the matrix is an evaluation of the squared exponential automatic relevance determination kernel for the true values of the inputs for each pair of groups $i, i' = 1, \ldots, n_g$, e.g., for $K_1(\tilde{X}, \tilde{X})$,

$$\begin{pmatrix} k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,n_g}) \\ \vdots & \ddots & \vdots \\ k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,n_g}) \end{pmatrix},$$

where

$$k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,i}, \tilde{\boldsymbol{X}}_{,i'}) = \exp\left\{ -\frac{1}{2} \sum_{d=1}^{2} \frac{(\tilde{X}_{d,i} - \tilde{X}_{d,i'})^2}{l_{1,d}^2} \right\}.$$

The covariance kernel matrix $K_2(\tilde{X}, \tilde{X})$ follows the same form, with $l_{1,d}$ replaced with $l_{2,d}$.

Similarly to the ICM, the matrices $\boldsymbol{a}_1(\boldsymbol{a}_1)^T$ and $\boldsymbol{a}_2(\boldsymbol{a}_2)^T$ are defined to be the covariance matrices $V_{K,1}$ and $V_{K,2}$. The matrix $V_{K,1}$ is defined to be

$$V_{K,1} = \begin{pmatrix} \sigma_{k,1}^2 & \sigma_{k,1}\sigma_{k,2}\rho_{V_K} \\ \sigma_{k,1}\sigma_{k,2}\rho_{V_K} & \sigma_{k,2}^2 \end{pmatrix} \tag{3.5.5.14}$$

The matrix $V_{K,2}$ is defined as

$$V_{K,2} = \begin{pmatrix} \lambda_1^2\sigma_{k,1}^2 & \lambda_1\sigma_{k,1}\lambda_2\sigma_{k,2}\rho_{V_K} \\ \lambda_1\sigma_{k,1}\lambda_2\sigma_{k,2}\rho_{V_K} & \lambda_2^2\sigma_{k,2}^2 \end{pmatrix}, \tag{3.5.5.15}$$

that is, the marginal variances in $V_{K,2}$ are scaled marginal variances from $V_{K,1}$, and the same correlation is assumed for both matrices.

Finally, the most notable case of a separable MOGP is the linear model of corregionalisation (LMC), which is discussed in Conti & O'Hagan (2010), Alvarez et al. (2011), Bilionis et al. (2013) and Liu et al. (2018), among others. The LMC can be thought of as a combination of the ICM and SPLF, where each functional relationship (the relationship between each output variable and the input variables) is a linear combination of multiple samples from multiple Gaussian processes. In Liu et al. (2018), the LMC is defined in terms of the function $f(\boldsymbol{x})$, where $y = f(\boldsymbol{x}) + \epsilon$, for some normally distributed model error $\epsilon \sim \text{N}(0, \sigma_\epsilon^2)$. For the EIV GP, define $f_1(\tilde{\boldsymbol{X}}_{,i})$ as the function evaluated at $\tilde{\boldsymbol{X}}_{,i}$ (i.e., the true values of the input variables for material i) for the first response variable. Then, the LMC expresses $f_1(\tilde{\boldsymbol{X}}_{,i})$

as a linear combination of Q latent functions, i.e.,

$$f_1(\tilde{\boldsymbol{X}}_{,i}) = \sum_{q=1}^{Q} a_{1,q} u_q(\tilde{\boldsymbol{X}}_{,i}), \tag{3.5.5.16}$$

where the function $u_q(\tilde{\boldsymbol{X}}_{,i})$ is a Gaussian process with mean 0 and covariance $\text{cov}(u_q(\tilde{\boldsymbol{X}}_{,i}), u_q(\tilde{\boldsymbol{X}}_{,i'})) = \boldsymbol{k}_q(\tilde{\boldsymbol{X}}_{,i}, \tilde{\boldsymbol{X}}_{,i'})$. Define now $\boldsymbol{f}(\tilde{\boldsymbol{X}}_{,i})$ as the multivariate (specifically bivariate in this case) function evaluated at $\tilde{\boldsymbol{X}}_{,i}$ for the first and second response variables. Then, in matrix form, the LMC expresses $\boldsymbol{f}(\tilde{\boldsymbol{X}}_{,i})$ as

$$\boldsymbol{f}(\tilde{\boldsymbol{X}}_{,i}) = B\boldsymbol{u}(\tilde{\boldsymbol{X}}_{,i}), \tag{3.5.5.17}$$

where $\boldsymbol{u}(\tilde{\boldsymbol{X}}_{,i}) = (u_1(\tilde{\boldsymbol{X}}_{,i}), \ldots, u_Q(\tilde{\boldsymbol{X}}_{,i}))$. The scalar value $a_{1,q}$ is the coefficient for the function $u_q(\tilde{\boldsymbol{X}}_{,i})$. Furthermore, the assumption is made that the latent functions $u_q(\tilde{\boldsymbol{X}}_{,i})$ and $u_{q'}(\tilde{\boldsymbol{X}}_{,i'})$ are orthogonal. So, for the second response variable, where

$$f_2(\tilde{\boldsymbol{X}}_{,i}) = \sum_{q'=1}^{Q} a_{2,q'} u_{q'}(\tilde{\boldsymbol{X}}_{,i}), \tag{3.5.5.18}$$

the covariance between the two response variables is given by

$$\begin{aligned} \boldsymbol{k}_{Y_1,Y_2}(\tilde{\boldsymbol{X}}_{,i}, \tilde{\boldsymbol{X}}_{,i'}) &= \sum_{q=1}^{Q} \sum_{q'=1}^{Q} a_{1,q} a_{2,q'} \text{cov}(u_q(\tilde{\boldsymbol{X}}_{,i}), u_{q'}(\tilde{\boldsymbol{X}}_{,i'})) \\ &= \sum_{q=1}^{Q} a_{1,q} a_{2,q} k_{Y_1}(\tilde{\boldsymbol{X}}_{,i}, \tilde{\boldsymbol{X}}_{,i'}) \end{aligned} \tag{3.5.5.19}$$

Further details of the LMC can be found in Alvarez et al. (2011), Liu et al. (2018). For this work, the SPLF model is preferred to the IMC and LMC—the advantage of having a linear combination of multiple Gaussian processes (as opposed to just one Gaussian process) to estimate multiple functional relationships which may differ in behaviour means the SPLF model is preferred over the ICM. Because of the limited amount of data considered in the work, the additional parameters required (which would need to be estimated) to fit the LMC could cause issues for the modelling. Moreover, the testing of these models on simulated data suggests that the ICM is insufficient in terms of providing accurate predictions of the response variables, whereas the SPLF model performed well in comparison.

Finally, for clarity, the EIV MOGP model to be investigated in this work is provided succinctly here. Defining the matrix of true values for the output variables as $\tilde{Y} \in \mathbb{R}^{n_g \times 2}$, the vectorisation of these output variables stacks the true values for the first output variable on top of the true values for the second output variable, giving the vector $\text{vec}(\tilde{Y}) = (\tilde{Y}_{1,1}, \ldots, \tilde{Y}_{1,n_g}, \tilde{Y}_{2,1}, \ldots, \tilde{Y}_{2,n_g})'$. This vector of true values is then assumed to be multivariate normal, with mean $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'$, and covariance matrix

$$V_{\text{MOGP},\tilde{X}} = V_{K,1} \otimes K_1(\tilde{X}, \tilde{X}) + V_{K,2} \otimes K_2(\tilde{X}, \tilde{X}) + \begin{pmatrix} \sigma^2_{\epsilon_1} I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma^2_{\epsilon_2} I_{n_g} \end{pmatrix},$$

that is,

$$\text{vec}(\tilde{Y}) \sim \text{N}_{2n_g}\left(\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}, V_{\text{MOGP},\tilde{X}} \right). \tag{3.5.5.20}$$

The EIV MOGP posterior distribution can be considered for the purpose of jointly estimating the output variables for a given input vector of true values $\tilde{\boldsymbol{X}}_{\text{new}}$. That is, the prediction of the response variables at the input vector $\tilde{\boldsymbol{X}}_{\text{new}}$ is given by the distribution $p(\tilde{\boldsymbol{Y}}_{\text{new}}|\boldsymbol{\theta}, \tilde{\boldsymbol{X}}_{\text{new}}, \tilde{Y}, \tilde{X})$, which is bivariate normally distributed with mean $\boldsymbol{m}^*$ and covariance matrix V^*, i.e.,

$$\tilde{\boldsymbol{Y}}_{\text{new}}|\boldsymbol{\theta}, \tilde{\boldsymbol{X}}_{\text{new}}, \tilde{Y}, \tilde{X} \sim \text{N}_2(\boldsymbol{m}^*, V^*) \tag{3.5.5.21}$$

where

$$\boldsymbol{m}^* = (\alpha_1, \alpha_2)' + \boldsymbol{K}(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}) V_{\text{MOGP},\tilde{X}}(\text{vec}(\tilde{Y}) - (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'), \tag{3.5.5.22}$$

and

$$V^* = V_{\text{MOGP},\tilde{\boldsymbol{X}}_{\text{new}}} - \boldsymbol{K}(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}) V_{\text{MOGP},\tilde{X}} \boldsymbol{K}(\tilde{X}, \tilde{\boldsymbol{X}}_{\text{new}}). \tag{3.5.5.23}$$

The covariance matrix $V_{\text{MOGP},\tilde{\boldsymbol{X}}_{\text{new}}}$ is analogous to that of $V_{\text{MOGP},\tilde{X}}$, replacing the matrix of existing true values of the input variables with the new true vector $\tilde{\boldsymbol{X}}_{\text{new}}$. The matrix $\boldsymbol{K}(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X})$ is defined to be

$$\boldsymbol{K}(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}) = V_{K,1} \otimes K_1(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}) + V_{K,2} \otimes K_2(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}),$$

and finally, $\boldsymbol{K}(\tilde{X}, \tilde{\boldsymbol{X}}_{\text{new}}) = \boldsymbol{K}(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X})^T$.

This concludes the discussion and definition of the statistical methods used in this work to estimate functional relationships between the response variable(s) and explanatory variable(s). The remainder of this chapter now considers appropriate statistical methods for investigating the explanatory variable(s) given the estimates of these functional relationships and given desired values of the response variable(s).

3.6 Inverse problems

The topic of inverse problems covers a wide range of applications. In this work, we are only interested in solving inverse problems within a statistical modelling framework. In this section, the ideas of 'forward' modelling and 'backward' modelling are more formally defined. Then, Bayesian solutions to the inverse problem in parametric and nonparametric regression are investigated, followed by details of the process for solving inverse problems in this work.

3.6.1 Defining the 'forward' and 'backward' models

Consider the paired observations of the response and explanatory variables given by (Y_i, x_i), for $i = 1, \ldots, n$, for which the relationship between the variables is of interest. Given that there are multiple possible relationships between these two variables (linear relationship, quadratic, etc.), a decision needs to be made as to which relationship is best. As discussed in Section 3.4, there are multiple potential methods to decide which parameters to include, where the preference here is to use some approximation of leave-one-out cross-validation information criterion (LOO-CV-IC). As an example, two models M_1 and M_2 can be fitted, with M_1 representing the simple linear model, and M_2 representing the quadratic model. For both models, the model parameters are estimated by the posterior distribution. Having fitted the models, the LOO-CV-IC can be evaluated for both models, and, along with other sensible considerations (fitted values plots, suitable model checking), the decision is made as to which model fits the data best.

This process becomes more extensive as more explanatory variables (and response variables) are available to be modelled, and this leads to more possible models to consider. Define the list of possible models to consider as $\mathcal{M} = \{M_1, M_2, \ldots\}$, leading to the same process described above, with the aim of finding the best-fitting model. This process of deciding which model is best in terms of describing the relationship between the response variable(s) and the explanatory variables(s) is known as the 'forward' modelling (or the forward modelling process), with each $M_1, M_2, \ldots$, defined to be a 'forward' model.

The methods described in Sections 3.3 and 3.5.1 have been described in terms of a *forward* model, with the aim of finding the relationship between the response variable(s) and explanatory variable(s), then deciding which model is the best-fitting with the methods described in Sections 3.4. The ultimate goal of this PhD is to find optimal values of the explana-

tory variable to produce some desired response value, given the estimates of the model parameters. The process of *inverting* the best-fitting model, to optimise the input variable(s) in order to produce some desired response value(s), is referred to as the 'backward' modelling. In other words, the forward modelling considers the behaviour of the response variable(s) given the explanatory variable(s), with the purpose of finding the best linear predictor in the explanatory variable(s) to predict the behaviour of the response variable(s). The 'backward' modelling assumes the fitted linear predictor and considers the behaviour of the explanatory variable(s) given the response variable(s), with the purpose of optimising the explanatory variable(s) in order to produce desired values of the response variable(s).

A more formal backward modelling definition is provided here. As an example, consider the case of a single output variable and a single input variable. If the desired response variable is defined as Y^*, and the corresponding explanatory variable as X^*, suppose the best-fitting model is identified as M_{best}. Then, the backward modelling process considers the backward model

$$Y^* = \hat{f}_{\text{best}}(X^*) + \epsilon^*, \tag{3.6.1.1}$$

where $\epsilon^* \sim \text{N}(0, \tau_\epsilon)$, and the function $\hat{f}_{\text{best}}$ represents the estimated relationship between the response variable and the explanatory variable from the best-fitting model. Then, the desired response variable Y^* is fixed at some value $Y^* = y^*$, and the backward model process is defined to be the optimisation the corresponding explanatory variable X^*, given $\hat{f}_{\text{best}}$ and $Y^* = y^*$. In the following sections, the ideal backward modelling process for this work is identified, with some more general background on the backward modelling process from the literature.

3.6.2 Linear regression inverse problems in a Bayesian setting

Given the issues identified in Chapter 1 regarding solving inverse problems in a classical setting, a Bayesian alternative is sought here.

The Bayesian approach to calibration was first developed by Hoadley (1970), where he explains how the 'classical estimator'

$$\hat{X}^*_{\text{C}} = \bar{\boldsymbol{x}} + \frac{S_{xx}}{S_{xy}} \left(Y^* - \bar{\boldsymbol{Y}} \right), \tag{3.6.2.1}$$

has some undesirable properties (namely the classical estimator has infinite mean square error) which he attempted to alleviate by considering the

problem from a Bayesian perspective. This Bayesian approach detailed by Hoadley entails assuming a joint prior distribution $p(\beta_0, \beta_1, \sigma_\epsilon^2, X^*)$ given by

$$p(\beta_0, \beta_1, \sigma_\epsilon^2, X^*) \propto p(\beta_0, \beta_1, \sigma_\epsilon^2) p(X^*),$$

with the noninformative prior distribution $p(\beta_0, \beta_1, \sigma_\epsilon^2) \propto \sigma_\epsilon^{-2}$ (with independence between X^* and $(\beta_0, \beta_1, \sigma_\epsilon^2)$). Hoadley assumed normally distributed errors, and derived the posterior density

$$p(X^*|Y^*, \boldsymbol{Y}, \boldsymbol{X}) \propto p(X^*) L(X^*),$$

where the likelihood $L(X^*)$ corresponds to the predictive density of Y^*, $p(Y^*|X^*)$. The vectors $\boldsymbol{Y}$ and $\boldsymbol{X}$ represent the observed data. To give a proper posterior density of X^*, the prior density $p(X^*)$ must also be proper. In particular, he assumed a non-central Student density centred at $\bar{x}$ as the prior density $p(X^*)$ and showed that the posterior distribution of X^* had mean

$$\hat{X}_{\mathrm{I}}^* = \bar{\boldsymbol{X}} + \frac{S_{xy}}{S_{yy}} \left(Y^* - \bar{\boldsymbol{y}}\right), \tag{3.6.2.2}$$

where $S_{yy} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{\boldsymbol{y}})^2$ (this provides the 'inverse estimator' solution from a classical setting, i.e., when regressing X on y). Further information of the univariate calibration in a Bayesian setting can be read in Hoadley (1970), Hunter & Lamboy (1981*b*) (which also considers cases of known σ_ϵ^2), and Hunter & Lamboy (1981*a*) (a response to the paper Hunter & Lamboy (1981*b*)).

The Bayesian methodology for calibration is preferable for this work for multiple reasons. It maintains consistency with EIV BR and EIV GPs used for the forward modelling, where the uncertainty in the forward modelling can be utilised in the backward modelling, so that in both the forward modelling and backward modelling, uncertainty of the estimates are taken into account. The concept of model inversion also naturally follows general Bayesian principles; having considered the behaviour of the response variable(s) conditioned on the explanatory variable(s) in the forward modelling, Bayes' theorem provides straightforward logic for investigating the behaviour of the explanatory variable(s) conditioned on the response variable(s). If required, elicitation can be carried out to provide further information through expert opinion for the backward modelling, for example for the prior distribution on X^*, which would provide more accurate estimates. In the non-Bayesian case, where there are two possible estimators to consider (the classical and the inverse), the Bayesian calibration (in this case) only requires a decision to be made on the prior distribution for X^*.

Once the posterior distribution of X^* is estimated, the posterior mode is the logical choice for the point-estimate solution to the inverse problem; the mode maximises the posterior density of X^*, i.e. where it provides the point where the probability density peaks. Other summary statistics such as the mean and median may not be appropriate, as the posterior of X^* may not be normal and could be multimodal.

The calibration in the case of a Bayesian multivariate parametric regression is a somewhat straightforward extension of the univariate case, as described in Brown (1982). Defining $p(\boldsymbol{X}^*)$ as the prior distribution for some multidimensional corresponding covariate $\boldsymbol{X}^*$, details are given to show the derivation of

$$p(\boldsymbol{X}^*|\boldsymbol{Y}^*, Y, X, \boldsymbol{\theta}) \propto p(\boldsymbol{X}^*)L(\boldsymbol{X}^*, \boldsymbol{\theta}),$$

where the likelihood $L(\boldsymbol{X}^*, \boldsymbol{\theta})$ corresponds to the predictive density of $\boldsymbol{Y}^*$, $p(\boldsymbol{Y}^*|\boldsymbol{X}^*, \boldsymbol{\theta})$. The matrices Y and X represent the observed data for the response variables and explanatory variables respectively. The demonstration then considers an invariant Jeffreys prior distribution for the model parameters, and some proper prior for $\boldsymbol{X}^*$.

The implementation of the backward modelling in this work is explicitly given (for EIV BR) in Section 3.6.4. In the following section, the ideas of the backward modelling here are carried over to nonparametric regression.

3.6.3 Inverse problems in nonparametric regression

Consider some jointly observed data (Y_i, x_i) where it is supposed that these sets of data are related by some function f with some additive random error ϵ_i in the Y-direction, that is,

$$Y_i = f(x_i) + \epsilon_i, \tag{3.6.3.1}$$

and suppose further that the function f is to be estimated using a nonparametric model. That is, the relationship f is not estimated by a fixed, predetermined relationship, rather the functional relationship between the variables is estimated without a predetermined parametric assumption. As in the previous subsections, suppose also there is some desired response value Y^* for which some corresponding explanatory variable value X^* is of interest, with the equation

$$Y^* = f(X^*) + \epsilon^*, \tag{3.6.3.2}$$

where ϵ^* is some random error. This defines the calibration problem in nonparametric regression.

The overall idea from the parametric regression case, that is, to identify how the forward model predicts the response variable(s) then invert these predictions, carries over to the nonparametric regression case. The key difference is that the *process* of predicting the response variable(s) in the nonparametric regression case effectively requires the observed data with which the function f is estimated (in other words, the observed data are required as 'training' data). For the parametric case of linear regression, for example, once the model coefficients $\beta_0, \beta_1, \ldots$, have been estimated, the observed data are not required to predict the response variable(s) for some value(s) of the explanatory variable(s). On the other hand, the Gaussian process regression predicts the response variable(s) using the GP posterior, which is based on the conditional distribution given the observed ('training') data.

For the same reasons as previously, the backward modelling for nonparametric regression is considered in a Bayesian setting. Therefore, it is still the case that the posterior distribution of X^* given Y^* (and the estimated model) is considered, which is given by

$$p(X^*|\boldsymbol{\theta}, Y, X, Y^*) \propto p(Y^*|\boldsymbol{\theta}, Y, X, X^*)p(X^*),$$

where $\boldsymbol{\theta}$ represents the hyperparameters of the nonparametric model, and $p(X^*)$ is the prior distribution of X^*. Effectively, the key difference is that Y and X is dropped from the conditioning in the predictive distribution $p(Y^*|\boldsymbol{\theta}, Y, X, X^*)$ in the parametric regression case, while also noting that the posterior distribution $\boldsymbol{\theta}$ in the parametric case contains the model coefficients (whereas here, $\boldsymbol{\theta}$ represents the vector of hyperparameters within the distribution(s) in the nonparametric model).

Having discussed the concept of the backward modelling more generally, for both parametric and nonparametric regression cases, the method for carrying out the backward modelling in this work is discussed in Section 3.6.4 for the parametric regression and Section 3.6.7 for the nonparametric regression (specifically for the EIV GP).

3.6.4 Backward model methodology for parametric regression

Suppose now the forward modelling process has been carried out, and the best-fitting forward model M_{best} (with respect to overall model form) has been identified and fitted, with the fitted model given by $\widehat{M}_{\text{best}}$. In this section, it is assumed that M_{best} is a parametric regression model.

For the purpose of simplification, suppose the relationship between the response variable and the explanatory variable in M_{best} is given by the simple linear model with errors-in-variables on both the response variable and the explanatory variable, so

$$M_{\text{best}} : \tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_i + \epsilon_i,$$

where $\epsilon_i \sim \text{N}(0, \tau_\epsilon)$. Within the forward modelling process, the joint posterior distribution

$$p(\boldsymbol{\theta}|Y, X)$$

has been estimated using MCMC, where $\boldsymbol{\theta} = (\beta_0, \beta_1, \boldsymbol{\tau}, \tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}})$, with $\boldsymbol{\tau} = (\tau_\epsilon, \tau_{\tilde{X}}, \tau_\eta, \tau_\delta)$. Also, the matrices Y and X represent the observed data (in the form of replicate measurements on multiple groups). This joint posterior distribution can be summarised using a large number S of posterior samples.

With the backward model for the simple linear model defined as in Equation 3.6.1.1, assume that some desired response value $Y^* = y^*$ is chosen, for which a corresponding explanatory variable value is to be suggested as a candidate for $X^* = x^*$ to be able to produce the desired response value y^*. The backward model of interest is therefore the model in Equation 3.6.1.1, with Y^* fixed at y^*. Note that, the precision parameters τ_η and τ_δ are not required in the backward model, as measurement error is not of interest. The precision term $\tau_{\tilde{X}}$ could be implemented depending on the choice of prior distribution for X^*, which is discussed further in Section 3.6.5.

In order to find appropriate values for the explanatory variable, a prior distribution is placed on X^*, which is then updated by the forward model posterior distribution $p(\boldsymbol{\theta}|Y, X)$ to give the posterior distribution $p(X^*|Y^*, \boldsymbol{\theta})$. This is carried out using MCMC. As noted above, the backward model is fitted a large number of times, and for each fitting of the backward model, a different sample from the joint posterior distribution of $p(\boldsymbol{\theta}|Y, X)$ is used, to take into account the uncertainty in the forward modelling process. Then, for each fitting of the backward model, a posterior distribution is found for X^*, from which a small set of samples is taken, and each set of samples from each fitting is then combined to create a complete posterior distribution for X^*.

The algorithm for the backward model process with the simple linear model with desired response value $Y^* = y^*$ is then as follows:

1. Jointly sample S draws from the posterior distribution $p(\boldsymbol{\theta}|Y, X)$, the forward modelling posterior distribution.

2. Assume some prior distribution for X_b^* (discussion of possible priors is given in Section 3.6.5). The subscript b is introduced to indicate that for each joint posterior sample $b = 1, \ldots, S$, a posterior distribution is estimated for X_b^*.

3. For each $b = 1, \ldots, S$, the backward model for the simple linear model is fitted, i.e.
$$y^* = \beta_{0,b} + \beta_{1,b} X_b^* + \epsilon^*,$$
where $\epsilon^* \sim \mathrm{N}(0, \tau_{\epsilon,b})$ and the values $(\beta_{0,b}, \beta_{1,b}, \tau_{\epsilon,b})$ are the b^{th} joint random sample from the S random samples of the joint posterior distribution from the forward modelling process. The fitting is done using MCMC, leading to the posterior $p(X_b^*|Y^* = y^*, \boldsymbol{\theta}_b)$.

4. Take a small random sample from each posterior $p(X_b^*|Y^* = y^*, \boldsymbol{\theta}_b)$, and combine these samples to get the complete posterior distribution $p(X^*|Y^* = y^*, \boldsymbol{\theta})$.

The posterior mode of $p(X^*|Y^* = y^*, \boldsymbol{\theta})$ is chosen as the candidate value as it is sensible to want to maximise the posterior $p(X^*|Y^* = y^*, \boldsymbol{\theta})$. Other summary metrics such as median and mean are not suitable; for example, a multimodal posterior of $p(X^*|Y^* = y^*, \boldsymbol{\theta})$ (which can occur for more complicated linear predictors) would likely lead to the mean and median both being low density values in the posterior, therefore one would be choosing a very unlikely value to be sampled from the posterior as the 'best guess' of X^* given Y^*.

The following section discusses possible prior distributions for X^*.

3.6.5 Prior distribution on X^*

In the case of the simple linear model, the variable X^* is 1-dimensional. In terms of which prior distribution could be placed on X^*, there are multiple possibilities, depending on the interpretation of X^*. From the forward model, posterior distributions were found for each $\tilde{X}_i$, which represents the true value for the explanatory variable for group i. The new corresponding variable X^* could be thought of as a new group, for which the true value of the response is given by $Y^* = y^*$, in which case, the variable X^* could be treated the same as $\tilde{X}_i$, that is, the same prior distribution could be used. This is then the normal distribution with mean μ_X and precision $\tau_{\tilde{X},b}$, i.e.,
$$X^* \sim \mathrm{N}(\mu_X, \tau_{\tilde{X},b}),$$
where $\tau_{\tilde{X},b}$ is a random sample from the posterior distribution of $\tau_{\tilde{X}}$.

The issue with this prior is the inability to extrapolate values for X^*, for example, in the case where $Y^* = y^*$ is an extreme value, i.e., when y^* is either larger than or smaller than all of the true response values $\tilde{Y}_i$. Take the case where the value y^* is larger than all $\tilde{Y}_i$, and the simple linear model is the best-fitting forward model. Suppose all posterior values $\beta_{1,b}$ are positive, corresponding to straight lines with a positive gradient defining the relationship between the response variable and the explanatory variable. The larger y^* becomes, the larger the corresponding value $X^* = x^*$ becomes, and using the normal distribution centred at the mean of the observed data with a between-materials variability based on the true values from the forward model $\tilde{X}_i$, the further x^* heads into the upper tail of the prior distribution on X^*, where the density is small. Although changes could be made to the parameters for the prior distribution in order to facilitate the extrapolation case, it is hard to justify which values to choose for the mean and the precision of the normal distribution.

A different distribution to use which can cope with this and still be appropriate for interpolation is simply the uniform distribution. This gives each value in its range the same density, and the limits of the uniform distribution can be set wide enough to account for both finding values for X^* within the minimum and maximum values of the true values $\tilde{X}_i$ and for finding values larger than the maximum $\tilde{X}_i$ and smaller than the minimum $\tilde{X}_i$. The application of the backward modelling in this work considers only the uniform distribution (see Chapter 5).

3.6.6 Extending the backward model

The backward modelling process for the simple linear model described in Section 3.6.4 can be adjusted to account for any relationship between some response variable(s) and some finite number of explanatory variables. For simplicity, the case with two explanatory variables is considered in this section, which can be scaled up to higher numbers of explanatory variables, and include any combination of interaction and polynomial terms. So, consider the parametric model with one response variable, two explanatory variables, and with measurement error on all variables as in Equation 3.3.1.8, i.e.

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_{1,i} + \beta_2 \tilde{X}_{2,i} + \epsilon_i.$$

Suppose the forward model has been fitted for this relationship, with S samples summarising the joint posterior distribution

$$p(\boldsymbol{\theta}|Y, X_1, X_2),$$

with $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \tau_\epsilon, \tau_\eta, T_{\tilde{X}}, T_\delta, \tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2)$. Furthermore, the matrices X_1 and X_2 represent the observed data for the first and second explanatory variables respectively.

The backward model to be considered now is

$$y^* = \beta_{0,b} + \beta_{1,b} X^*_{1,b} + \beta_2 X^*_{2,b} + \epsilon^*,$$

where $\epsilon^* \sim \mathrm{N}(0, \tau_{\epsilon,b})$ and the values $(\beta_{0,b}, \beta_{1,b}, \beta_{2,b}, \tau_{\epsilon,b})$ are the b^{th} joint random sample from the S random samples of the joint posterior distribution from the forward modelling process.

Another consideration needs to be made for the prior distribution on the corresponding explanatory variables $\boldsymbol{X}^*_b = (X^*_{1,b}, X^*_{2,b})$. The preference here again is to use a uniform distribution. The uniform distribution can be defined over any measurable region in space, so there are many options for placing a uniform prior over $\boldsymbol{X}^*_b$. A uniform 'box' is a straightforward option ('box' meaning the volume or n-dimensional volume space bounded by a cube or n-dimensional hypercube), which is defined simply by placing a uniform distribution on each of the two explanatory variables, leading to $X^*_{1,b} \sim \mathrm{U}(c_1 - \rho_{X_1}, c_1 + \rho_{X_1})$ and $X^*_{2,b} \sim \mathrm{U}(c_2 - \rho_{X_2}, c_2 + \rho_{X_2})$ (for example, if $c_1 = c_2 = 0$ and $\rho_{X_1} = \rho_{X_2} = 1$, the box is the area bounded by a square centred at the origin with side length 2).

Alternatively, a uniform 'disk' (or, for more than two explanatory variables, 'ball') could be used as the prior distribution for $(X^*_{1,b}, X^*_{2,b})'$, which can be created in two ways, either using rejection sampling or with polar coordinates. Suppose the uniform 'disk' from which the samples will be taken has radius ρ and is centred at (c_1, c_2). For some horizontal axis x_1 and vertical axis x_2, the equation of this circle is given by $(x_1 - c_1)^2 + (x_2 - c_2)^2 = \rho_X^2$, and so to draw samples for $(X^*_{1,b}, X^*_{2,b})$ from within this circle, rejection sampling can be used by specifying that the samples must satisfy the inequality $(X^*_{1,b} - c_1)^2 + (X^*_{2,b} - c_2)^2 \leq \rho_X^2$.

To use polar coordinates to sample values from within the same circle, define the probability $P(R < r)$ to be the probability of choosing a point within a circle of radius r for some $r < \rho_X$ (in other words, imagine one circle with radius ρ_X, and another circle with radius r where $r < \rho_X$, and this defines the probability of picking a point from the circle of radius r which lies within the circle of radius ρ_X). This probability is equal to $\frac{\pi r^2}{\pi \rho_X^2} = \frac{r^2}{\rho_X^2}$. Using a probability integral transform, the variable R' is defined to be $R' = \frac{R^2}{\rho_X^2}$, which can be rearranged to give $R = \rho_X \sqrt{R'}$. Placing a uniform distribution as the prior on R' over the range [0,1], samples from this prior are then transformed to produce values of R, and for the other polar coordinate θ,

a uniform prior over the range $(0, 2\pi]$ can be used, and these points are transformed to produce the samples $X_1^* = R\cos(\theta) + c_1, X_2^* = R\sin(\theta) + c_2$.

The uniform 'disk' can be taken a step further, with unique radii for each dimension of $\boldsymbol{X}^*$ leading to the area of space bounded by an ellipse (or, for more than two explanatory variables, by a hyperellipsoid). For an ellipse centred at (c_1, c_2) with semi-major axis ρ_{X_1} and semi-minor axis ρ_{X_2}, and some horizontal axis x_1 and vertical axis x_2, the equation of this ellipse is given by $\frac{(x_1-c_1)^2}{\rho_{X_1}^2} + \frac{(x_2-c_2)^2}{\rho_{X_2}^2} = 1$. To draw samples for (X_1^*, X_2^*) from within this ellipse, rejection sampling can be used by specifying that the samples must satisfy the inequality $\frac{(X_1^*-c_1)^2}{\rho_{X_1}^2} + \frac{(X_2^*-c_2)^2}{\rho_{X_2}^2} \leq 1$. With polar coordinates, the variable R from above becomes two-dimensional (so $R = (R_1, R_2)$), such that $X_1^* = R_1\cos(\theta) + c_1, X_2^* = R_2\sin(\theta) + c_2$, with $R_1 = \rho_{X_1}\sqrt{R'}$, $R_2 = \rho_{X_2}\sqrt{R'}$ and $R' \sim \mathrm{U}(0, 1)$. As the number of explanatory variables in the linear predictor increases, it is clear that the preferred choice to sample from a uniform ball (or ellipsoid) is to use rejection sampling.

The backward model can then be further extended by adding some random variable Z to the desired response variable Y^*. The idea behind the variable Z is to enlarge/widen the target for the desired response, which is appropriate for when the desired response need not be produced *exactly*, rather it is important to find values of the explanatory variables to produce a response value within a certain range of the desired response. For example, suppose the desired response value for tapped density that has been suggested is 0.55g/ml, so $y^* = 0.55$, and it is also stated that values within 0.02g/ml of $y^* = 0.55$ are equally as good. With this information, define the random variable Z such that $Y' = Y^* + Z$, and place the uniform distribution U(-0.02,0.02) on Z. Then, values of X^* are found which can produce response values Y' within the range [0.53,0.57]. More generally, the random variable takes the distribution $\mathrm{U}(-\rho_Z, \rho_Z)$ for the purpose of finding values of X^* to produce values of the response within the desired range $[y^* - \rho_Z, y^* + \rho_Z]$.

The final extension considered here is that of a multivariate response model, which accomplishes the ultimate goal of the PhD of optimising the explanatory variable(s) in order to produce desired values for multiple response variables simultaneously. The benefit to fitting the backward model with a multivariate response, as opposed to fitting multiple univariate backward models for each response variable, is that the relationship between the response variables is captured in the multivariate response model.

The backward model for the multivariate response is demonstrated here with two response variables and two explanatory variables. That is, for some

fixed vector of desired response values $\boldsymbol{Y}^* = \boldsymbol{y}^* = (Y_1^* = y_1^*, Y_2^* = y_2^*)$, the backward modelling process with bivariate response and two explanatory variables then follows this algorithm (assuming the joint posterior distribution of $\boldsymbol{\theta}$ has been estimated in the forward modelling):

1. Jointly sample S draws from the posterior distribution $p(\boldsymbol{\theta}|Y_1, Y_2, X_1, X_2)$ from the forward modelling posterior distribution.

2. Assume a prior distribution for $\boldsymbol{X}_b^* = (X_{1,b}^*, X_{2,b}^*)$—a uniform 'disk' or 'box' over an appropriate range, as described above.

3. For each $b = 1, \ldots, S$, the backward model

$$\begin{pmatrix} Y_1^* \\ Y_2^* \end{pmatrix} = \begin{pmatrix} \beta_{0,1,b} & \beta_{1,1,b} & \beta_{2,1,b} \\ \beta_{0,2,b} & \beta_{1,2,b} & \beta_{2,2,b} \end{pmatrix} \begin{pmatrix} 1 \\ X_{1,b}^* \\ X_{2,b}^* \end{pmatrix} + \begin{pmatrix} \epsilon_1^* \\ \epsilon_2^* \end{pmatrix}$$

   is fitted, where $\boldsymbol{\epsilon}^* = (\epsilon_1^*, \epsilon_2^*)' \sim \mathrm{N}_2(\mathbf{0}, T_{\epsilon,b})$, and the values $\boldsymbol{\beta}_b = (\beta_{0,1,b}, \ldots, \beta_{2,2,b})$ and the matrix $T_{\epsilon,b}$ are the joint b^{th} random sample from S random samples the joint posterior distribution leading to the posterior $p(\boldsymbol{X}_b^*|\boldsymbol{Y}^* = \boldsymbol{y}^*, \boldsymbol{\theta}_b)$.

4. Take a small random sample from each posterior distribution $p(\boldsymbol{X}_b^*|\boldsymbol{Y}^* = \boldsymbol{y}^*, \boldsymbol{\theta}_b)$, and combine these samples to get a complete posterior distribution $p(\boldsymbol{X}^*|\boldsymbol{Y}^* = \boldsymbol{y}^*, \boldsymbol{\theta})$.

As in the case of a univariate desired response, the variable Z can be introduced as a widening/enlargement of the target for the desired response. In the bivariate response case, the variable Z becomes a bivariate random variable, defined as $\boldsymbol{Z} = (Z_1, Z_2)$. The prior on $\boldsymbol{Z}$ is again uniform, and the discussion of the uniform 'disk' and 'box' for the corresponding explanatory variables $\boldsymbol{X}^*$ carries over to $\boldsymbol{Z}$ in the same sense. For example, the powder manufacturer may suggest that the the ideal powder flow specifications are to have a tapped density (the first response variable) in the range of 0.53 and 0.57 g/ml and an angle of repose (the second response variable) in the range of 21 and 24 degrees, suggesting a uniform prior for $\boldsymbol{Z}$ over the volume space within the ellipse centred at $(y_1^* = 0.55, y_2^* = 22.5)$ with semi-major axis $\rho_{Z_1} = 0.02$ and semi-minor axis $\rho_{Z_2} = 1.5$.

The backward modelling methodology for the Gaussian process regression model is now considered below.

3.6.7 Backward model methodology for GP regression

The backward model methodology for the GP more or less coincides with the parametric regression, but, for the Gaussian process, the methodology requires posterior samples of the true values $\tilde{\boldsymbol{Y}}$ and $\tilde{\boldsymbol{X}}$ from the forward modelling, due to the reliance of 'training' data in order to predict the response variable (in the forward modelling sense) for some 'test' data. The backward model for the EIV GP is detailed here, with one input variable and one response variable.

Suppose a Gaussian process has been fitted in a Bayesian setting in order to estimate the functional relationship between some response variable and some input variable, where both variables are represented by some true values for groups $i = 1, \ldots, n_g$. The observed data are noisy observations of these true values. As a reminder, the multivariate normal distribution relating the variables is therefore

$$\tilde{\boldsymbol{Y}} \sim \mathrm{N}_{n_g}(\boldsymbol{m}(\tilde{\boldsymbol{X}}), V(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}})), \tag{3.6.7.1}$$

where $\boldsymbol{m}(\tilde{\boldsymbol{X}}) = (m(\tilde{X}_1), \ldots, m(\tilde{X}_{n_g}))' = \boldsymbol{\alpha} \in \mathbb{R}^{n_g}$ is the mean vector, and $V(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}) = K(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}) + \sigma_\epsilon^2 I_{n_g}$ is the $n_g \times n_g$ covariance matrix. The matrix $K(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}})$ represents the covariance kernel matrix, where each element is an assessment of some covariance kernel k, taking pairs of true values $\tilde{X}_i$ and $\tilde{X}_{i'}$ as inputs, for $i, i' = 1, \ldots, n_g$. The covariance kernel of choice is the squared exponential, defined here as

$$k(\tilde{X}_i, \tilde{X}_{i'}) = \sigma_k^2 \exp\left(-\frac{(\tilde{X}_i - \tilde{X}_{i'})^2}{2l^2}\right), \tag{3.6.7.2}$$

with σ_k^2 representing some signal variance. In contrast, the parameter σ_ϵ^2 represents some noise variance.

Fitting the EIV GP model to the true values, as well as estimating the true values, leads to a posterior distribution for the parameter vector

$$\boldsymbol{\theta} = (\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}}, \sigma_\epsilon, \sigma_k, l, \alpha, \tau_{\tilde{X}}, \tau_\eta, \tau_\delta). \tag{3.6.7.3}$$

For simplicity, the notation for the hyperparameter vector is reintroduced, given by

$$\boldsymbol{\phi} = (\sigma_\epsilon, \sigma_k, l, \alpha), \tag{3.6.7.4}$$

since the precision parameters are not required in the conditional distribution, i.e., the EIV GP posterior. Suppose further that some desired response value $Y^* = y^*$ is suggested by an expert, and the aim is to optimise the input

variable by choosing a value that is most likely to produce $Y^* = y^*$ given the posterior distribution of $\boldsymbol{\theta}|Y, X$. The backward model for the EIV GP with one input and one response then follows the following algorithm:

1. Jointly sample S draws from the joint posterior distribution for $\boldsymbol{\theta}$ defined in Equation 3.6.7.3, estimated in the forward modelling.

2. Assume a prior distribution for X_b^* (the input variable corresponding to the desired response variable Y^*).

3. For each $b = 1, \ldots, S$, consider the joint normal distribution

$$\begin{pmatrix} \tilde{\boldsymbol{Y}}_b \\ Y^* \end{pmatrix} \sim \mathrm{N}_{n_g+1}\left(\begin{pmatrix} \boldsymbol{m}_b(\tilde{\boldsymbol{X}}_b) \\ m_b(X_b^*) \end{pmatrix}, \begin{pmatrix} V_b(\tilde{\boldsymbol{X}}_b, \tilde{\boldsymbol{X}}_b) & \boldsymbol{k}_b(\tilde{\boldsymbol{X}}_b, X_b^*) \\ \boldsymbol{k}_b(X_b^*, \tilde{\boldsymbol{X}}_b) & v_b(X_b^*, X_b^*) \end{pmatrix} \right).$$

   Note that the subscript b for the mean vectors and block matrices in the covariance matrix indicates that the function used to produce the vector/matrix is informed by the joint b^{th} posterior sample for the parameters required in the function (also, the vectors of true values $\tilde{\boldsymbol{Y}}_b$ and $\tilde{\boldsymbol{X}}_b$ are the joint b^{th} posterior sample of the true values).

4. Derive the conditional distribution (i.e., the EIV GP posterior) of $Y^*|\tilde{\boldsymbol{Y}}_b, \tilde{\boldsymbol{X}}_b, \boldsymbol{\phi}_b, X_b^*$ from the above joint distribution by estimating the quantities m_b^* and v_b^* (the mean and variance of the conditional distribution), given by

$$m_b^* = m_b(X_b^*) + \boldsymbol{k}_b(X_b^*, \tilde{\boldsymbol{X}}_b) V_b(\tilde{\boldsymbol{X}}_b, \tilde{\boldsymbol{X}}_b)^{-1} (\tilde{\boldsymbol{Y}}_b - \boldsymbol{m}_b(\tilde{\boldsymbol{X}}_b))$$

   and

$$v_b^* = v_b(X_b^*, X_b^*) - \boldsymbol{k}_b(X_b^*, \tilde{\boldsymbol{X}}_b) V_b(\tilde{\boldsymbol{X}}_b, \tilde{\boldsymbol{X}}_b)^{-1} \boldsymbol{k}_b(\tilde{\boldsymbol{X}}_b, X_b^*).$$

5. With the prior distribution of choice on X_b^*, the prior is updated given this conditional distribution and the forward model posterior samples $\boldsymbol{\theta}_b$ (using MCMC) to give the posterior $p(X_b^*|Y^* = y^*, \tilde{\boldsymbol{Y}}_b, \tilde{\boldsymbol{X}}_b, \boldsymbol{\phi}_b)$.

6. Take a small random sample from each posterior distribution $X_b^*|Y^* = y^*, \tilde{\boldsymbol{Y}}_b, \tilde{\boldsymbol{X}}_b, \boldsymbol{\phi}_b$ and combine these samples to get a complete posterior distribution $X^*|Y^* = y^*, \tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}}, \boldsymbol{\phi}$

The algorithm follows the same process as in the previous sections; the difference here is that predictions of the response variable (in the forward modelling sense) are carried out using the EIV GP posterior, whereas the

response variable in the EIV BR models can be predicted directly using the parametric relationship between the response and the input variables. That is, in the backward modelling, the aim is to invert the process of predicting the response variable, in order to learn about the behaviour of the input variable(s) given the output variable(s).

Having covered the statistical methods that are implemented in this work, the following chapter now considers applying these methods, to both simulated data and real-world data. In the simulated data cases, the modelling process is fine-tuned by considering standardisation of the observed data, and by finding appropriate prior distributions to be able to recover the underlying relationships on which simulations are based. With the real-world data, the forward modelling process is fitted to the powder flow and powder bed deposition case study, described in Section 2.2, to find the appropriate relationship between tapped density and angle of repose (the response variables) and the powder properties from the powder rheometer (the input variables).

Chapter 4

Investigating powder flow

In this chapter, the forward modelling process is investigated in multiple ways; this involves rigorous testing on simulated data as well as fitting multiple forward models to the real data. This chapter is split into two sections; in the first section, the two methods are in a sense 'developed' by considering some simulated data, and in the second section, these developed models are tested on the real data with the aim of demonstrating a legitimate and careful approach to answering modelling questions. The development of the method using the simulated data is investigated in two ways—ensuring the model is finding the correct relationship defined in the simulation (recovering parameter values as high-density values in the posterior distribution), and ensuring the MCMC is performing effectively and efficiently by considering the mixing and convergence of multiple parallel chains. In order to satisfy these two aspects, adjustments can be made to certain parts of the modelling (such as prior distributions, MCMC tuning parameters, rescaling the data; the scaling of the data in this work is discussed later).

The example scenario of the simulation needs consideration. The data structure in the model should align with that of the real data for which you want to prepare the model—in this work, since the observations are measured with error, the simulated data will be noisy observations of 'true', unobservable values. The magnitude of measurement error in the simulation, the number of replicate measurements, and the number of groups, are all aspects of the modelling which can be altered in the simulation in order to see the effect these have on the effectiveness of the modelling, and so are not necessarily identical to those aspects in the real data. In the case of the real data, there are seven groups (materials), and three replicate measurements on each group (for each variable apart from the response variable angle of

repose, which has five replicate measurements). It is easy to imagine that an investigation into another aspect of additive manufacturing with the same data structure as here (measurement error with replicate observations) could lead to different magnitudes of measurement error, which could have an impact on the modelling. It is clear that the more groups in the modelling, the easier it would be to recover the relationship between the variables in the model (supposing the chosen linear predictor was justified). In essence, the data structure of the simulations aligns with the structure observed in the real data, but other aspects of the modelling have some alterations, in some cases to confirm the modelling is working as intended, and in other cases to test the capabilities of the model.

It should be reiterated that the aim of using simulated data is to ensure the model is performing as intended, as opposed to finding out, for example, the ideal number of replicate observations or ideal number of groups in order for the model to perform to a certain level. It is *not* the aim of the simulations to duplicate the specific example of real data (that will be considered here) precisely, as it is very unlikely that any future data sets would align exactly (in terms of magnitude of measurement error, number of replicate measurements, and so on) with the data set considered here. Setting up the modelling so that it performs best for this data set would not correspond to setting up the modelling to deal with a variety of possible data sets. The aim of the simulations is also *not* to investigate ideal values for the number of groups and number of replicate observations on the groups for the purpose of designing an experiment, although in some cases, it becomes clear that the number of materials may need to be increased in order to capture more complicated linear predictors.

As previously discussed in Section 3.1, the modelling in this work is considered in a Bayesian setting, requiring prior and hyperprior distributions to be placed on the model parameters of interest. These distributions can be designed to incorporate any level of information that might be available for that parameter prior to considering any data. In the case of uninformative prior distributions, a very small amount of information is put into the prior distribution, so that the posterior distribution for the parameters is largely dictated by the data at hand. On the other end of the spectrum, informed prior distributions could be used, which contain a significant amount of information about the parameter (often obtained from an expert) before considering any data. In this chapter, models with uninformative priors and models with informed priors are both fitted to simulated data, to demonstrate the potential different outcomes of these two possible extents of prior information. Once the real data is considered, the models are fitted using

informed priors, to demonstrate the process of obtaining expert information and a possible way to build informed priors using that expert information.

Due to the complexity of the errors-in-variables problem and consideration of Bayesian methods, model-fitting code is not readily available. For this reason, bespoke code to fit these models has been written by the author. Some examples of the models (written in JAGS) are provided in Appendix F.

Before working through the simulated data examples, the scaling of the observations is discussed briefly.

4.1 Scaling the observed data

A very common practice in statistical modelling is to scale the data before fitting any model. The reasons for this are:

1. to make the model accessible to any data set,
2. to make collaboration easier,
3. to ensure numerical instability in parameter estimates is minimised (particularly in a Bayesian setting while taking into account uncertainty), which can occur if the data for different variables are sufficiently different in magnitude.

Before carrying out the scaling of the data, there were some issues with numerical instability in this work, which prompted the need to consider data scaling. To illustrate the issues: tapped density measurements were initially measured in kg/m^3, with observed values measured roughly around 500kg/m^3. On the other hand, conditioned bulk density measurements are recorded in g/ml, with observed values measured roughly around 0.45g/ml. The scale of these values are very different, but these variables both measure a form of density, and with the same scaling produce very similar values. If the tapped density values are then converted to g/ml, its observed values measure roughly around 0.5g/ml, which compares much better in terms of scale with conditioned bulk density. The observed measurements of the variable tapped consolidation vary roughly from 250mJ to 1000mJ, so this magnitude is much larger than for conditioned bulk density and now tapped density.

Consequently, the data here has been standardised by scaling onto the range [0,1] (in some cases of simulated data, some response variable values are negative, in which case the data are scaled onto the range [-1,1]). This

has been carried out by dividing the data by an appropriate power of 10. The reason for this is to avoid using the data itself 'for normalisation'. That is, the common practice of subtracting the mean and dividing by the standard deviation is not followed here, as this can lead to ambiguity as to how to proceed with new data (should a rescaling take place to take into account the new data, or shoud the original scaling be applied to the new data).

This scaling onto the range [0,1] allows for suitable adjustments to be made to 'conventional' uninformative prior distributions, and in turn, provides the suitable modelling environment for multiple situations. These adjustments are exemplified by the EIV Gaussian process regression modelling, for which estimation of hyperparameters is tricky, and attention is required for finding appropriate hyperprior distributions.

4.2 Errors-in-variables Bayesian regression on simulated data

Firstly, the errors-in-variables Bayesian regression (EIV BR) models are developed by considering some simulated data (SD). Fitting the model to simulated data where the relationship between the variables is known helps to confirm that the model can correctly estimate the relationship between variables. Moreover, any structural adjustments to the model-fitting process, whether it be reparameterising or changing the parameters of prior distributions, can be trialled in this simulation setting, where the underlying truth is known.

Particularly for the Bayesian regression modelling, compared with the Gaussian process modelling, the forward modelling process is more long-winded, as candidate Bayesian regression models can be fitted to the same data, each differing with their assumed predetermined relationship. For this reason, multiple simulated-data cases are considered, each with differing relationships to be estimated.

For all cases, the data structure in the simulation represents the general data structure as in the real data, with replicate measurements on multiple groups, assumed to be noisy observations of some 'true', unobservable value.

4.2.1 Simple linear model

In this first case, some data is simulated for the purpose of fitting a simple linear model with measurement error. Taken from Sections 3.2.3 and 3.3, a reminder of the simple linear EIV BR model follows. This model is defined

by

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_i + \epsilon_i, \tag{4.2.1.1}$$

where the value $\tilde{Y}_i$ represents the 'true', unobservable value of the response variable for material i (where $i = 1, \ldots, n_g$), the value $\tilde{X}_i$ represents the 'true', unobservable value of the response variable for material i, and the value ϵ_i represents some random error. That is, it is assumed that a straight line represents the relationship between the response and explanatory variables, and this relationship is estimated using 'true', unobservable values, because of the measurement error on the observed data. The observed data are of the form

$$Y_{i,j} = \tilde{Y}_i + \eta_{i,j} \tag{4.2.1.2}$$

and

$$X_{i,k} = \tilde{X}_i + \delta_{i,k} \tag{4.2.1.3}$$

for the response variable and explanatory variable respectively, where the subscript j denotes the j^{th} replicate measurement of the response variable for material i, and similarly the subscript k denotes the k^{th} replicate measurement of the response variable for material i. The terms $\eta_{i,j}$ and $\delta_{i,k}$ also represent random error terms, capturing the within-materials variability (due to both measurement error and variation due to sampling). Using the mean and precision to define the normal distribution, these random error terms in the model take the distributions

$$\epsilon_i \sim \mathrm{N}(0, \tau_\epsilon),$$

$$\eta_{i,j} \sim \mathrm{N}(0, \tau_\eta),$$

and

$$\delta_{i,k} \sim \mathrm{N}(0, \tau_\delta).$$

The true values for the explanatory variable assume the normal distribution

$$\tilde{X}_i \sim \mathrm{N}(\mu_X, \tau_{\tilde{X}}),$$

where the precision parameter $\tau_{\tilde{X}}$ is defined to be the between-materials precision. Note that, without prior information, it is assumed that μ_X is fixed at 0.5, which is the midpoint of range onto which the data is scaled.

The specification of the prior and hyperprior distributions is

$$\begin{aligned}
\beta_0 &\sim \mathrm{N}(0, \tau_{\beta_0}),\\
\beta_1 &\sim \mathrm{N}(0, \tau_{\beta_1}),\\
\tilde{X}_i &\sim \mathrm{N}(\mu_X, \tau_{\tilde{X}}),\\
\tau_\epsilon &\sim \mathrm{Gamma}(a_\epsilon, b_\epsilon),\\
\tau_\eta &\sim \mathrm{Gamma}(a_\eta, b_\eta),\\
\tau_\delta &\sim \mathrm{Gamma}(a_\delta, b_\delta),\\
\tau_{\tilde{X}} &\sim \mathrm{Gamma}(a_{\tilde{X}}, b_{\tilde{X}}),\\
\tau_{\beta_0} &\sim \mathrm{Gamma}(a_{\beta_0}, b_{\beta_0}),\\
\tau_{\beta_1} &\sim \mathrm{Gamma}(a_{\beta_1}, b_{\beta_1}).
\end{aligned} \tag{4.2.1.4}$$

The gamma distributions above are parameterised by shape and rate. In the case of uninformative priors distributions, a gamma distribution on a precision parameter (assuming the likelihood of the data is normally distributed) takes values of 0.001 for both the shape and the rate. Informed gamma prior distributions are discussed further when considering the real data.

The simulated data is discussed now. The number of groups chosen for the simulation is 13, with the reason being this is appreciably more than for the real data (seven materials), meaning the estimate of the model parameters are likely to be better with the simulated data. This provides the model with more information, with the idea being to check the feasibility of the modelling approach (structure, appropriate priors)—if successful, this allows for a reduction in the number of materials to a level which represents the real-life problem. The true values for the explanatory variable, $\tilde{X}_i$, are chosen. The specific values are not of importance, but need to be chosen and recorded so that the estimate of the true values in the model can be checked. These are arbitrarily taken to be $\tilde{X}_i = i$ for $i = 1, \ldots, 13$. The replicate observations $X_{i,k}$ for each true value $\tilde{X}_i$ are simulated from a normal distribution centred at $\tilde{X}_i$ with a precision value of 100 (a standard deviation of 0.1). The number of replicate observations is seven (for both the explanatory variable and the response variable), again with the idea that it would provides the model with a sufficiently large amount of information in order to estimate measurement error and the respective true values. The values of the intercept β_0 and slope β_1 in the linear model are also noted to be 3 and 5 respectively. The true values for the response variable, $\tilde{Y}_i$, are then produced using the simple linear relationship

$$\tilde{Y}_i = 3 + 5\tilde{X}_i + \epsilon_i,$$

where the precision of the model error ϵ_i, τ_ϵ, is set equal to 4. Following this, the observed data for the response variable, $Y_{i,j}$, are simulated from a normal distribution with mean $\tilde{Y}_i$ and precision 100.

The observed data for the two variables are then scaled onto the range [0,1], so the data for both variables is divided by 100. The simple linear errors-in-variables model is then fitted, with the observed data above updating the prior distributions defined in Equation 4.2.1.4, given also the relationships defined in Equations 4.2.1.1, 4.2.1.2 and 4.2.1.3. If not otherwise specified, the sampler is run with four parallel chains, each with an adaptation phase of length 1000, a burn-in of length 25000, then 20000 samples are stored having taken every 10^{th} sample from 200000. Any required changes to the MCMC tuning parameters will be noted when appropriate.

Based on the stored samples from the MCMC output, the potential scale reduction factor is calculated for each random variable for the purpose of assessing the convergence of the MCMC output (from now on, 'potential scale reduction factor' is abbreviated to 'PSRF'). As discussed in Section 3.1.4, any estimated value larger than 1.1 is considered to be too large, in which case the MCMC output has not converged to the posterior distribution. The value in each case (for each variable as well as the multivariate value) is 1.00 (to 2 decimal places), suggesting the MCMC algorithm has achieved convergence to the posterior distribution. Furthermore, the effective sample size for each variable is calculated, which helps determine whether there is a high level of autocorrelation between posterior samples. With reference again back to Section 3.1.4, the accepted minimum estimate for effective sample size is $5m$ (also assuming convergence), where m is twice the number of parallel chains. In this case with four parallel chains, the accepted minimum effective sample size estimate is 40 (after summing the effective sample sizes from each chain). From the four parallel chains, the smallest calculated value of effective sample size is 13460, which is a clear indication that there are no issues with autocorrelation with the posterior samples. For simulations in subsequent sections, the MCMC summary statistics relating to convergence and mixing are discussed more briefly in cases where both convergence and sufficient mixing are found.

The posterior distributions can now be checked. Firstly, the true values of the explanatory variable $\tilde{X}_i$ and of the response variable $\tilde{Y}_i$ are checked, compared with their observed data, and show the true values are appropriate. These plots can be found in Appendix D.1.

The following plot in Figure 4.1 shows the posterior densities for the measurement error *standard deviation* terms $\sigma_\delta = \sqrt{\frac{1}{\tau_\delta}}$ (left) and $\sigma_\eta = \sqrt{\frac{1}{\tau_\eta}}$

(right), where each posterior sample of the precisions τ_δ and τ_η has been converted into a standard deviation value (for ease of understanding). These plots demonstrate that the model has captured slightly larger levels of measurement error than was used in the simulation, with the standard deviation for the measurement error after scaling being 0.001. Evidently from Figure 4.1, the estimated values of posterior standard deviation measurement error are still small, which is appropriate.

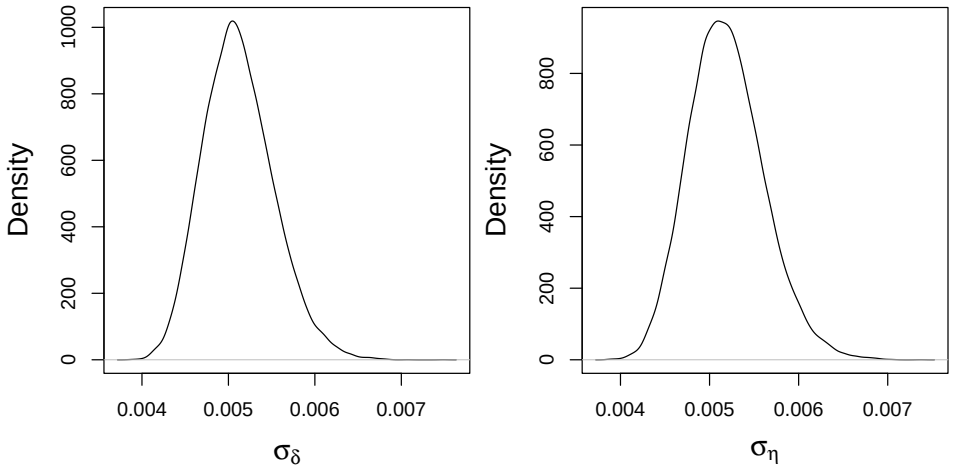

Figure 4.1: Posterior densities for σ_δ (left) and σ_η (right) having rearranged each posterior sample of τ_δ and τ_η into standard deviation values. The 'chosen' values of each parameter are 0.001, which is just outside the plotted range in both cases.

The posterior density of $\sigma_{\tilde{X}} = \sqrt{\frac{1}{\tau_{\tilde{X}}}}$, the between-groups standard deviation, is plotted in Figure 4.2, again with the posterior samples of $\tau_{\tilde{X}}$ converted into standard deviations. The plot indicates that the between-materials standard deviation is considerably larger than the measurement error standard deviation (comparing with the top plot of Figure 4.1), which is to be expected, for two reasons. Firstly, the chosen value of the measurement error standard deviation (after scaling) is much smaller than any rough estimate of the between-materials precision. Secondly, given that μ_X is fixed at 0.5, and the model has to recover the true value of $\tilde{X}_1$ (which takes the chosen value of 0.01 after scaling), the posterior density for $\sigma_{\tilde{X}}$ would logically have a rough lower bound of around 0.25, using the 2σ rule. This looks to be the case.

The final posterior densities to be investigated visually are those of the

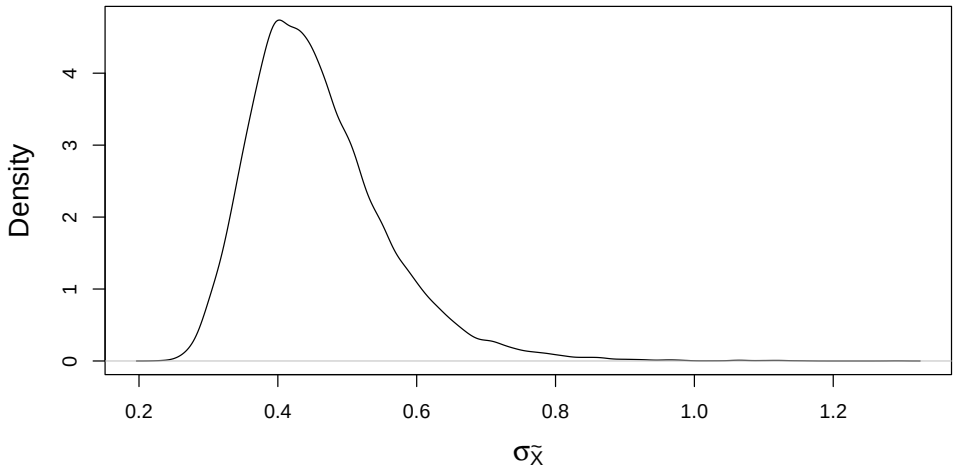

Figure 4.2: Posterior density for $\sigma_{\tilde{X}}$, having rearranged each posterior sample of $\tau_{\tilde{X}}$ into a standard deviation value.

slope and intercept of the model, β_0 and β_1 respectively. Their posterior densities are displayed in Figure 4.3, along with their 'chosen' values for the simulation. It is clear that the model is estimating the relationship effectively, with the 'chosen' values (after scaling) for β_0 and β_1 being high-density values in their respective posterior densities.

Having inspected the posterior densities for the parameters of interest, the fitted values of the model can also be inspected. As discussed in Section 3.4.2, there are several fitted value plots to be considered, comparing the posterior distribution of $\tilde{Y}_i$ with the observed data $Y_{i,j}$, comparing the posterior distribution of $\tilde{X}_i$ with the observed data $X_{i,k}$, and comparing the distribution of the fitted values

$$\widehat{Y}_{i,s} = \beta_{0,s} + \beta_{1,s}\tilde{X}_{i,s}$$

with the posterior distribution of $\tilde{Y}_i$. Also of interest is a plot of the fitted line produced by the model, which is given in terms of a prediction interval (along with confidence ellipses for the joint marginal posterior distribution of $\tilde{Y}_i$ and $\tilde{X}_i$). These plots are given in Figure 4.4.

Note that, on the axis labels, the notations of $\tilde{Y}$ and $\tilde{X}$ are represented by Yt and Xt respectively. The model is confident in the estimation of the true values given the credible intervals for each true value (top-left plot of Figure 4.4). The intervals being effectively centred on the line $\tilde{Y} = Y$ along with the narrow credible intervals implies the true values have been

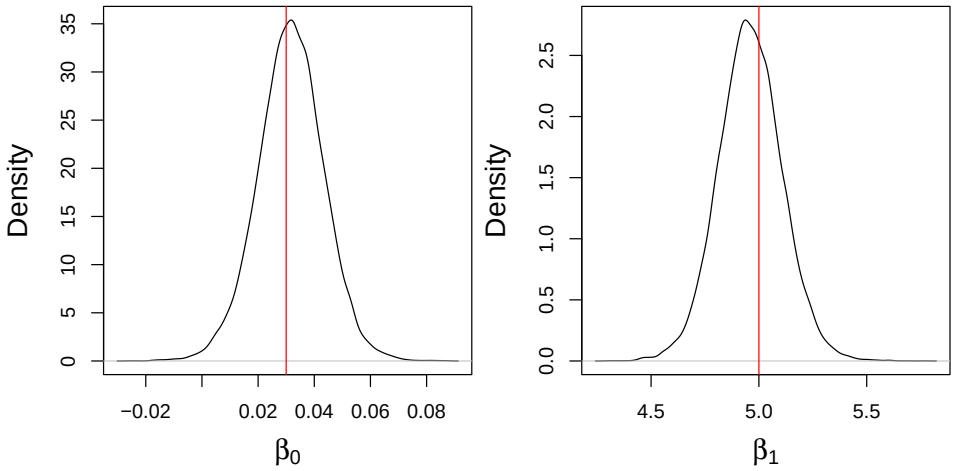

Figure 4.3: Posterior densities for β_0 (left) and β_1 (right), along with the 'chosen' values for each parameter from the simulation.

recovered well in the model.

In contrast to the true values on the response variable, whose credible intervals are almost not visible in the plot, the credible intervals are slightly more visible on the true values of the explanatory variable, again due to the ranges on the axes, which are in this case relatively smaller. The top-right plot in Figure 4.4 clearly indicates the true values for the explanatory variable have been recovered well in the model, with the 95% credible intervals centred on the line $\tilde{X} = X$.

Finally, it is unsurprising that the model has estimated the straight line relationship very well, as demonstrated in the bottom-left plot of Figure 4.4. Recall from Section 3.4.2 that the observed data is omitted from this plot, since the observed measurement $Y_{i,j}$ does not have a one-to-one correspondence with the observed measurement $X_{i,k}$. The fitted values of the response variable given in the bottom-right plot of Figure 4.4 clearly show that the model has effectively predicted the response variable at each true value of the explanatory variable.

Following the checks of the MCMC output (with respect to convergence and mixing), plots of the marginal posterior densities of parameters of interest, plots of fitted values, and a plot of the fitted model, it is clear that the modelling is working effectively. The true values from the simulation, and the relationship between the response variable and the explanatory variable, have been recovered from the underlying simulation.

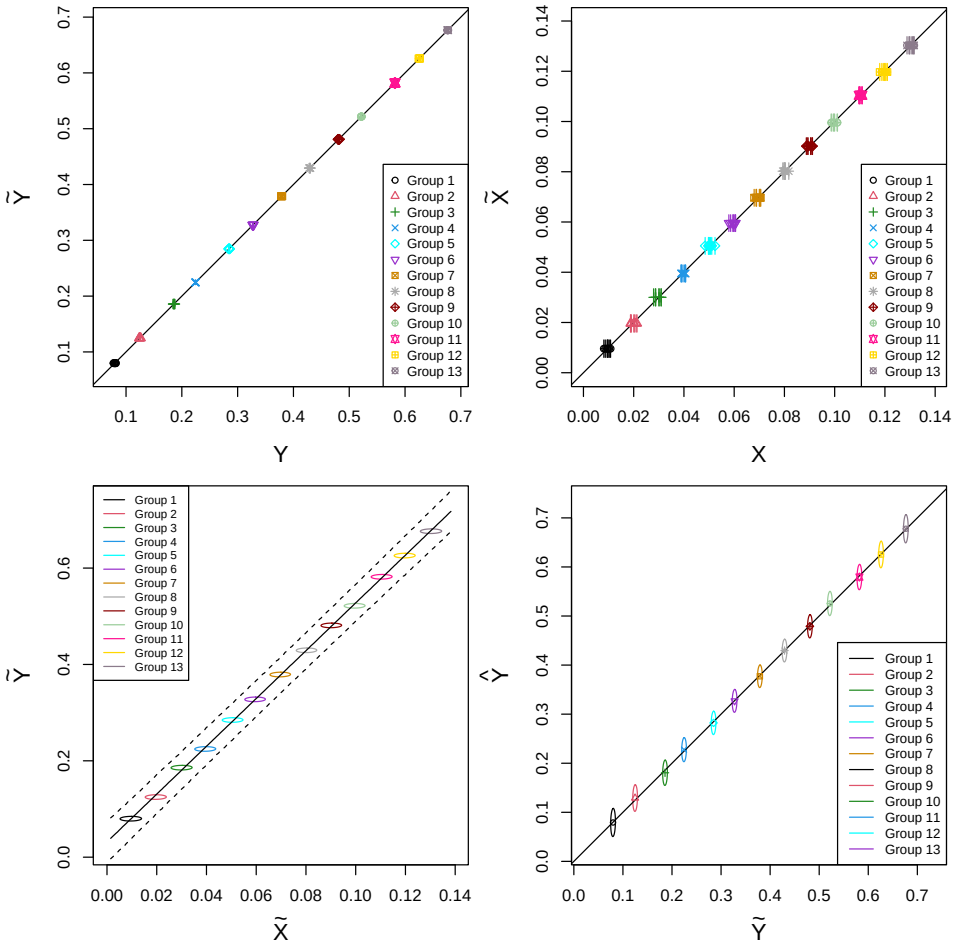

Figure 4.4: Four plots considering the various 'fitted values' plots to be considered in the EIV BR. The top-left plot compares the posterior distribution of each $\tilde{Y}_i$ with the observed data $Y_{i,j}$; the top-right plot compares the posterior distribution of each $\tilde{X}_i$ with the observed data $X_{i,k}$; the bottom-left plot provides the fitted model; the bottom-right plot compares the posterior distribution of the fitted values $\widehat{Y}_i$ with the true values $\tilde{Y}_i$.

4.2.2 Cubic model

Having considered the most simple one-covariate model, a brief investigation of some data simulated from a cubic model is considered, to demonstrate that the modelling can cope with this additional complexity. Moreover, the cubic model provides further complexities in the backward modelling scenario, which is considered in Section 5.1.2. Again, the aim of the simulation is to show that, if some response variable and explanatory variable from real data were to have a cubic relationship, then this modelling would be able to recover the relationship effectively. For the errors-in-variables case, the cubic model is therefore given by

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_i + \beta_{11} \tilde{X}_i^2 + \beta_{111} \tilde{X}_i^3 + \epsilon_i.$$

Once more, the same 'chosen' true values and observed data for the explanatory variable are used for this simulation. The chosen values of β_0, β_1, β_{11} and β_{111} are given by 30, 10, 20 and -1.5 respectively; these are selected carefully to ensure that the simulated data for the response variable are not negative, and after scaling cover a large range of [0,1]. The 'chosen' true values of the response variable are then simulated using the equation

$$\tilde{Y}_i = 30 + 10\tilde{X}_i + 20\tilde{X}_i^2 - 1.5\tilde{X}_i^3 + \epsilon_i,$$

with the chosen true values inputted for $\tilde{X}_i$, and ϵ_i taking a normal distribution with mean 0 and precision 4. The observed data for the response variable, $Y_{i,j}$ are simulated from a normal distribution with mean $\tilde{Y}_i$ and precision 1.

The cubic model is then fitted using the simulated observed data for $Y_{i,j}$ and $X_{i,k}$ as described above. The convergence to the posterior distribution and sufficient levels of mixing are both confirmed using PSRF and effective sample size estimates respectively.

For brevity, the plots regarding estimates of any aspect of the cubic model are restricted to the posterior densities of β_0, β_1, β_{11} and β_{111} (which are found in Figure 4.5), and the fitted values comparisons given in Figure 4.6, where the top-left plot compares the posterior $\tilde{X}_i$ with the respective observed data, the top-right plot compares the posterior $\tilde{Y}_i$ with the respective observed data, the bottom-left plot provides the fitted cubic model and the bottom-right plot shows the posterior fitted values $\widehat{Y}_i$ against the posterior true values $\tilde{Y}_i$.

Once more, the posterior densities for the model coefficients (see Figure 4.5) seem generally quite appropriate. It seems clear that the 'chosen' values

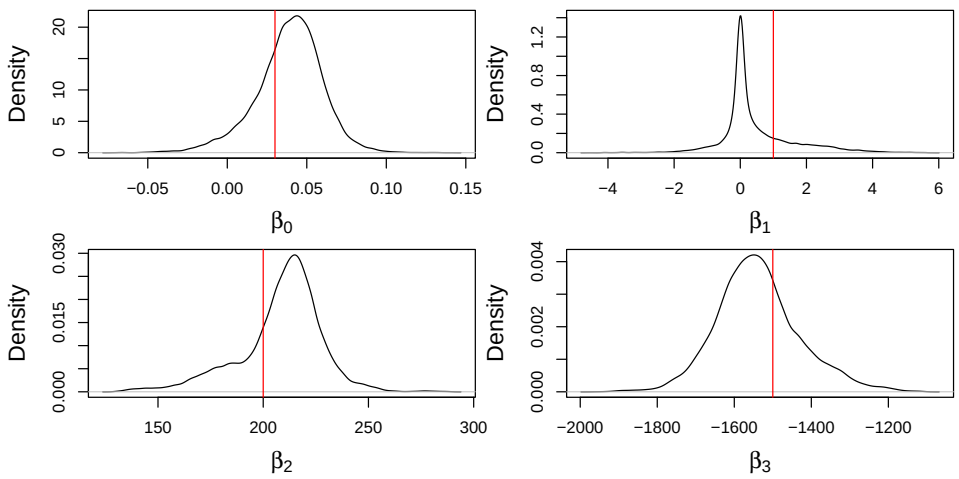

Figure 4.5: Posterior densities for β_0 (top left), β_1 (top right), β_{11} (bottom left) and β_{111} (bottom right) from the cubic model fitted to simulated data, along with the 'chosen' values (after scaling) for each parameter from the simulation.

for these parameters in the simulation (vertical red lines in the plot) are high-density points in posterior densities for β_0 and β_{111}. The chosen value for β_{11} looks to be about half the density of the posterior mode for the parameter. Finally, the posterior density for β_1 seems to be centred at 0, and the corresponding chosen value for this parameter is a low-density value in the posterior. As these models become more complex, i.e., there are more parameters to be estimated, it is unsurprising that the chosen values used to build the simulation are not the highest density values in the posterior. Moreover, as more terms are included in the linear predictor, more possible 'solutions' can be found for estimating the relationship between the response variable and the explanatory variable.

The comparison between the marginal posteriors of the true values on the response variable and the corresponding observed data (see top-left plot of Figure 4.6) clearly shows that the $\tilde{Y}_i$ have been estimated well in the model, and the same can be said for the estimates of $\tilde{X}_i$ (top-right plot of Figure 4.6). The plot of the fitted model with its corresponding uncertainty, given in the bottom-left plot of Figure 4.6, clearly shows that the model has estimated the relationship effectively, despite chosen true values not being recovered as posterior modes for each model coefficient. The fitted values of $\widehat{Y}_i$ are considered in the bottom-right plot of Figure 4.6, where their

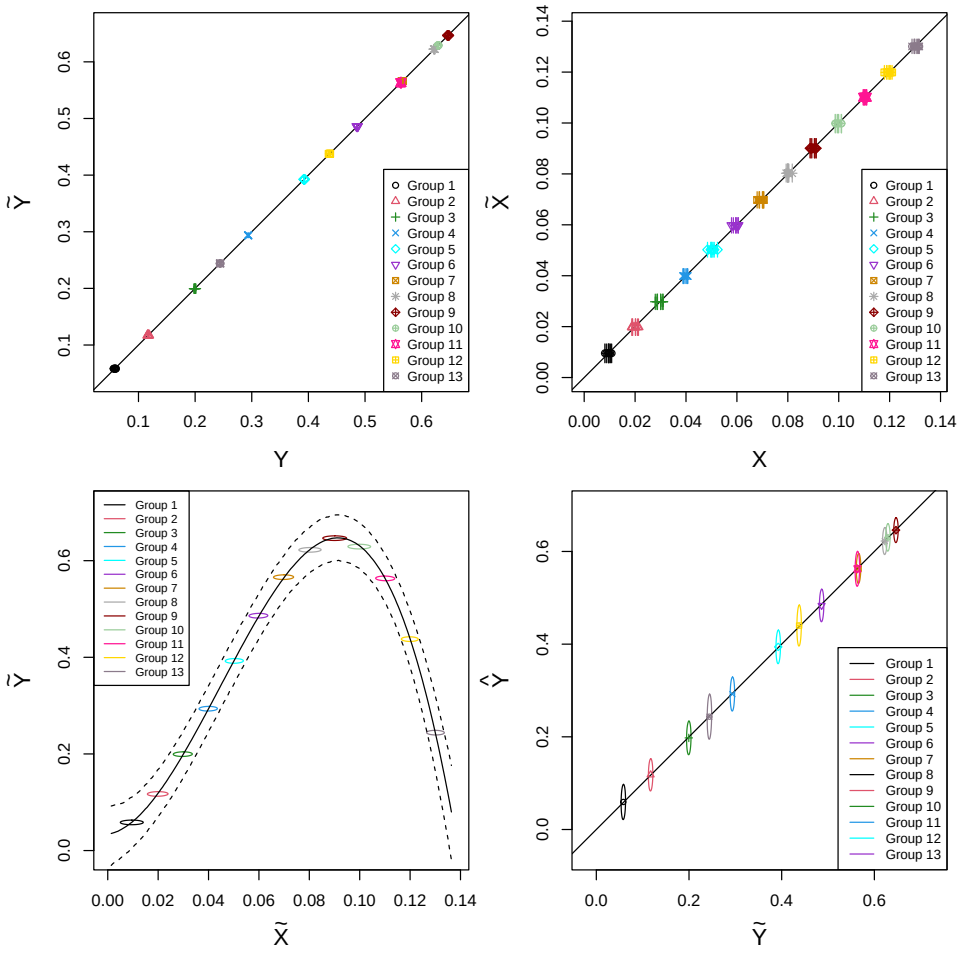

Figure 4.6: Four plots comparing the 'fitted' values of the model: the top-left plot compares the posterior $\tilde{Y}_i$ with the respective observed data, the top-right plot compares the posterior $\tilde{X}_i$ with the respective observed data, the bottom-left plot provides the fitted cubic model and the bottom-right plot shows the posterior fitted values $\widehat{Y}_i$ against the posterior true values $\tilde{Y}_i$.

joint distributions with their corresponding posterior distributions of $\tilde{Y}_i$ are plotted. With these distributions centred on the line $\widehat{Y} = \tilde{Y}$, this shows the model has been fitted well, since the true values of the response variable also correspond well with their respective observed data. It is clear that there is slightly more uncertainty in the fitted values compared to what has been seen in previous simulations, which is justified by the increase of complexity of the model.

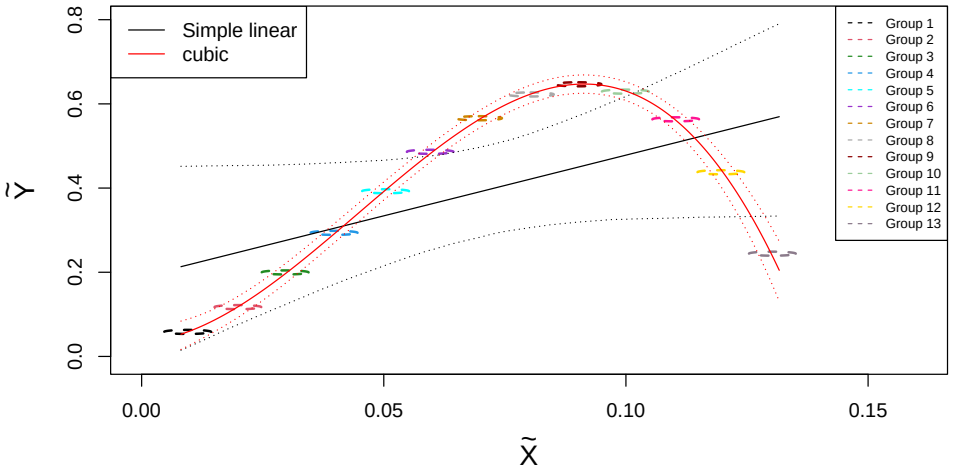

Figure 4.7: A plot of two 'fitted' models produced by errors-in-variables Bayesian regression, with the black lines corresponding to the simple linear model, and red lines corresponding to the cubic model. Also displayed are joint 95% credible ellipses for the true values of each group (taken from the simple linear model), as well as a 95% prediction intervals for the fitted lines produced by the two models.

An opportunity is presented here to demonstrate the use of LOO-CV-IC, explained in Section 3.4.2, for model comparison. Using the same simulated data to which the cubic model has been fitted, a simple linear model is fitted. The fits of these two models to the cubic simulation data can be found in Figure 4.7. It is clear that the visual comparison between the model fits shows the cubic model provides a much better model fit in terms of the quality of the predictions of the response variable than the simple linear model, which is to be expected. To confirm this with PSIS-LOO-CV-IC, we find a mean estimate of -66.7 for the cubic model, with a standard error of 1.5, whereas the mean estimate for the simple linear model is given by -1.7, with a standard error of 7.1. This is clear evidence that the cubic model

is a more appropriate fit, since the mean estimate of PSIS-LOO-CV-IC is smaller for the cubic model.

4.2.3 Two explanatory variables, correlated measurement error

This section concerns fitting models to simulated data for one response variable and two explanatory variables. Specifically, the model fits in this section assume correlation between the measurement error terms for the two explanatory variables; more details on this are given below. Model fits to simulated data where there are no assumptions of correlation between measurement error terms on the explanatory variable are considered in Section 4.2.4. The case of correlated measurement error on the explanatory variable is discussed first since it is prevalent in the real data considered later in the chapter. The errors-in-variables Bayesian regression with two explanatory variables is defined as

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_{1,i} + \beta_2 \tilde{X}_{2,i} + \epsilon_i, \tag{4.2.3.1}$$

where the value $\tilde{X}_{1,i}$ represents the 'true', unobservable value of the first explanatory variable for material i, and $\tilde{X}_{2,i}$ represents the true, unobservable value of the second explanatory variable for material i. Therefore, this model assumes there is some additive relationship between the two explanatory variables and the response variable, with neither polynomial nor interaction terms included in the linear predictor. Extensions of this model with interaction and polynomial terms are considered with simulated data in Section 4.2.5.

The correlated measurement error is now discussed (see Section 3.3.1 for further detail). The values X_{1,i,k_1} and X_{2,i,k_2} represent the observed data for the first and second explanatory variables respectively. In this setting with two explanatory variables, there are two possibilities—either the observed data X_{1,i,k_1} and X_{2,i,k_2} have been measured under identical circumstances (in the real data in this work, this corresponds to the replicates k_1 and k_2 being the same subsample of powder, for $k_1 = k_2$), or, the observed data have not been measured under identical circumstances. For the real data in this work, both of these outcomes are possible, and so both models are 'tested' using simulated data.

The true values for the two explanatory variables take the normal distribution

$$\begin{pmatrix} \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \end{pmatrix}, T_{\tilde{X}} \right),$$

where the matrix $T_{\tilde{X}}$ represents the precision matrix of the distribution. Given that it will always be plausible that the explanatory variables could have some dependence on each other, this is factor into the modelling by assuming this joint distribution. Moreover, a Wishart prior distribution is assumed for the precision matrix $T_{\tilde{X}}$, with degrees of freedom 2. The scale matrix is discussed in conjunction with the scale matrix on the measurement error precision T_δ following the adjustments to the prior specification. The observed data are of the form

$$\begin{pmatrix} X_{1,i,k_1} \\ X_{2,i,k_2} \end{pmatrix} = \begin{pmatrix} \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \end{pmatrix} + \begin{pmatrix} \delta_{1,i,k_1} \\ \delta_{2,i,k_2} \end{pmatrix},$$

with the assumption of correlated measurement error leading to the joint distribution

$$\begin{pmatrix} \delta_{1,i,k_1} \\ \delta_{2,i,k_2} \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, T_\delta \right),$$

where T_δ represents the precision matrix for the measurement error vector.

The simulated data is now discussed. The first explanatory variable takes the same 'chosen' true values as the sole explanatory variable considered in Sections 4.2.1 to 4.2.2. The second explanatory variable takes the 'chosen' true values

$$\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_{13}) = (1, 3, 4, 5, 7, 8, 7, 4, 3, 2, 5, 6, 9).$$

These true values guarantee a different scaling of the second explanatory variable to the first (simply for demonstration purposes), and guarantee that the explanatory variables are not a linear combination of one another. The observed data for the explanatory variables are simulated with nonzero correlation, in line with the model being investigated. Therefore, the observations X_{1,i,k_1} and X_{2,i,k_2} are simulated from a bivariate normal distribution with mean $(\tilde{X}_{1,i}, \tilde{X}_{2,i})'$ and covariance matrix

$$\begin{pmatrix} 0.01 & 0.006 \\ 0.006 & 0.01 \end{pmatrix},$$

meaning the marginal standard deviations are 0.1 (taking the square root of the variances on the diagonal), and the correlation between the observations X_{1,i,k_1} and X_{2,i,k_2} given $k_1 = k_2$ is 0.6 (the off-diagonal element of the 2×2 covariance matrix is the product of the marginal standard deviations and the correlation parameter). The true values for the response variable are simulated from the relationship

$$\tilde{Y}_i = 3 + 8\tilde{X}_{1,i} - 5\tilde{X}_{2,i} + \epsilon_i, \tag{4.2.3.2}$$

with the model error ϵ_i having mean 0 and precision 4 (a standard deviation 0.5). That is, the 'chosen' values for the model coefficients in the simulation are $\beta_0 = 3$, $\beta_1 = 8$ and $\beta_2 = -5$. The observed data for the response variable $Y_{i,j}$, are then simulated from the normal distribution with mean $\tilde{Y}_i$ and precision 1. The observed data for the response variable and the explanatory variables are then scaled onto the range [0,1], by dividing the response variable observations by 100, the first explanatory variable by 100, and the second explanatory variable by 10.

The specification of the prior and hyperprior distributions for this model (two explanatory variables, with jointly-simulated observations of the two explanatory variables) is an extension of that specified in Equation 4.2.1.4, with the inclusions of $\beta_2 \sim \mathrm{N}(0, \tau_{\beta_2})$ and $\tau_{\beta_2} \sim \mathrm{Gamma}(a_{\beta_2}, b_{\beta_2})$, and the replacements of $\tilde{X}_i \sim \mathrm{N}(\mu_X, \tau_{\tilde{X}})$, $\tau_\delta \sim \mathrm{Gamma}(a_\delta, b_\delta)$ and $\tau_{\tilde{X}} \sim \mathrm{Gamma}(a_{\tilde{X}}, b_{\tilde{X}})$ with $\tilde{\boldsymbol{X}}_i \sim \mathrm{N}_2(\boldsymbol{\mu}_X, T_{\tilde{X}})$, $T_\delta \sim \mathrm{Wishart}(S_\delta, \nu_\delta)$ and $T_{\tilde{X}} \sim \mathrm{Wishart}(S_{\tilde{X}}, \nu_{\tilde{X}})$ respectively. Note that $\tilde{\boldsymbol{X}}_i = (\tilde{X}_{1,i}, \tilde{X}_{2,i})'$, $\boldsymbol{\mu}_X = (\mu_{X_1}, \mu_{X_2})'$, and the parameters of the Wishart distribution are the scale matrix, followed by the degrees of freedom.

Suitable scale matrices S_δ and $S_{\tilde{X}}$ are required. Given the Wishart distribution is a multivariate equivalent to the gamma distribution, a possible attempt to satisfy this requirement is to try to replicate some standard properties of the gamma distributions used for the priors of τ_δ and $\tau_{\tilde{X}}$ in the model with one explanatory variable. The gamma distribution with shape 0.001 and rate 0.001 has mean 1 and variance 1000. Suppose $\nabla \sim \mathrm{Wishart}(\nabla_0, \nu)$. The mean and variance of the diagonal elements $\nabla_{i,i}$ are given by $\mathrm{E}(\nabla_{i,i}) = \nu\nabla_{0_{i,i}}$ and $\mathrm{Var}(\nabla_{i,i}) = \nu\left(\nabla^2_{0_{i,i}} + 2\nabla_{0_{i,i}}\right)$. Setting the expectation equal to 1 then gives $\nabla_{0_{i,i}} = \frac{1}{\nu}$, and substituting this into the variance while setting that to 1000 gives

$$\nu\left(\frac{1}{\nu^2} + 2\frac{1}{\nu}\right) = 1000 \implies \frac{1}{\nu} + 2 = 1000 \implies \nu = \frac{1}{998},$$

which is not possible, since $\nu \geq 2$. In the literature, Gelman et al. (2013) states that a noninformative distribution is obtained as $\nu \to 0$, so the best option is to set the degrees of freedom to be 2. In this case, clearly both requirements cannot be satisfied; satisfying the requirement of the mean being 1 suggests $\nabla_{0_{i,i}} = \frac{1}{2}$, and satisfying the requirement of the variance

being 1000 suggests

$$\begin{aligned} 1000 &= 2\left(\nabla_{0_{i,i}}^2 + 2\nabla_{0_{i,i}}\right) \\ &= 2\left((\nabla_{0_{i,i}} + 1)^2 - 1\right) \\ &\implies 500 = (\nabla_{0_{i,i}} + 1)^2 - 1 \\ &\implies \sqrt{501} - 1 = \nabla_{0_{i,i}} \implies \nabla_{0_{i,i}} = 21.38 \text{ (to 2 d.p.).} \end{aligned}$$

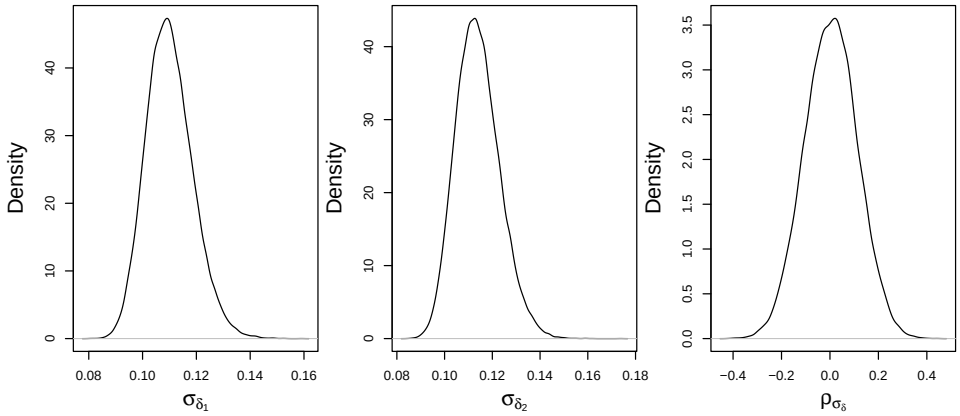

Figure 4.8: Posterior densities of the standard deviation and correlation parameters corresponding to the precision matrix of the measurement error of the two explanatory variables T_δ (the left plot is the standard deviation on the measurement error for the first explanatory variable, the middle plot is the standard deviation on the measurement error for the second explanatory variable, and the right plot is the correlation between the measurement error on both explanatory variables). This corresponds to the model with I_2 as the scale matrices for the Wishart priors on $T_{\tilde{X}}$ and T_δ.

A preferred approach is to carry out a sensitivity analysis with different scale matrices for S_δ and $S_{\tilde{X}}$ and see how this affects the posterior distribution estimated in the model. Following Gelman et al. (2013), the degrees of freedom parameter for both Wishart distributions is set to 2. The first attempt at fitting the model is under the assumption that $S_\delta = S_{\tilde{X}} = I_2$, where the off-diagonal element for the prior is assumed to be 0, which is the most logical value given no further evidence (*a priori*) for a nonzero value.

The model is fitted using the same MCMC tuning parameters discussed in Section 4.2.1. Firstly, in order to achieve convergence to the posterior distribution according to PSRF, 40000 posterior samples were stored from

the MCMC (with the same burn-in and thinning parameters as previously stated). In this case, the posterior samples mix sufficiently well, and so the posterior distribution is now investigated.

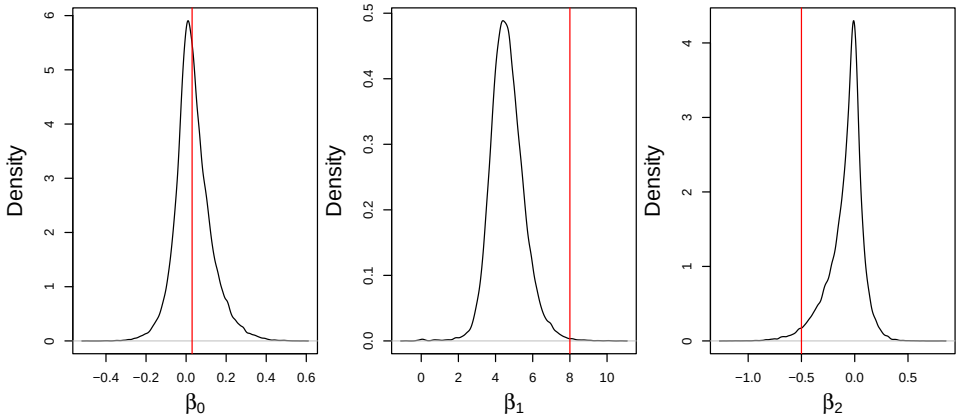

Figure 4.9: Posterior densities for β_0 (left), β_1 (middle) and β_2 (right) from the model with two explanatory variables fitted to simulated data, along with the 'chosen' values (after scaling) for each parameter from the simulation. This corresponds to the model with I_2 as the scale matrices for the Wishart priors on $T_{\tilde{X}}$ and T_δ.

The posterior densities of the random error parameters are firstly considered. Some manipulation of the posterior samples for the precision matrix T_δ for the measurement error of the two explanatory variables in this simulation is helpful, in order to provide a more simple comparison between the 'chosen' values for these parameters in the simulation. As noted earlier, the observed data for the explanatory variables were simulated with marginal standard deviations of 0.1 and correlation 0.6. After scaling the data, the marginal standard deviation for the first explanatory variables becomes 0.001 (since the data are divided by 100); for the second explanatory variable this becomes 0.01 (since the data are divided by 10). The correlation remains unchanged since it is invariant under any linear transformation. In order to recover the standard deviations and correlation from the precision matrix, each matrix posterior sample, say $T_{\delta,s}$ for $s = 1, \ldots, S$ (with S being a large number of joint posterior samples) is inverted to give the respective covariance matrix, with the diagonal elements giving the measurement error variance for each explanatory variable, and the off-diagonal element giving the product of the standard deviations for each explanatory variable and the correlation. Note that the diagonal elements of the precision matrix are

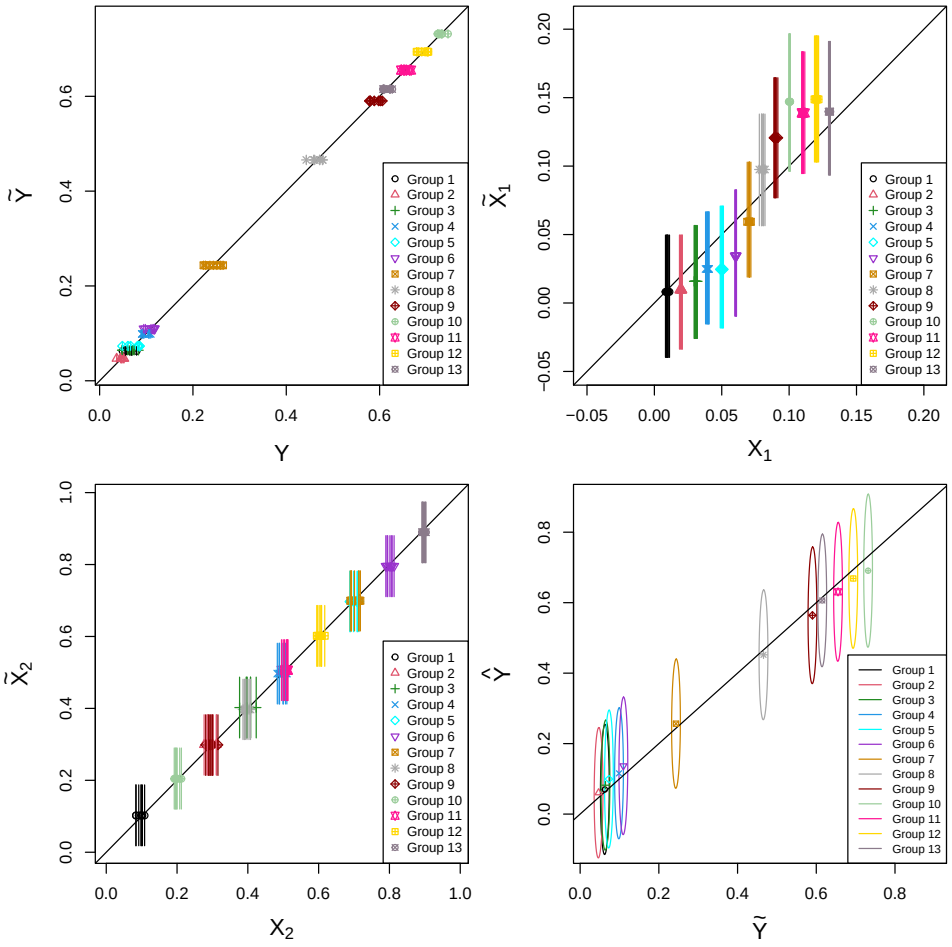

Figure 4.10: Four plots of the fitted values of the two-covariate model with correlated measurement error and $S_{\tilde{X}} = S_\delta = I_2$. The top-left plot compares the posterior densities of $\tilde{Y}_i$ with the corresponding observed data $Y_{i,j}$ for each material i; the top-right plot compares the posterior densities of $\tilde{X}_{1,i}$ with its respective observations X_{1,i,k_1} for each i; the bottom-left plot compares the posterior densities of $\tilde{X}_{2,i}$ with its respective observations X_{2,i,k_2} for each i; the bottom-right plots compares the posterior fitted values $\widehat{\tilde{Y}}_i$ with the posterior true values $\tilde{Y}_i$ for each i.

not the corresponding marginal precisions for each explanatory variable, so taking the reciprocal of these values does not correspond to a variance term.

Following the manipulation of the posterior precision matrix samples to be able to extract the marginal standard deviations and correlation, the density of these samples of these terms are plotted in Figure 4.8. For each plot, the 'chosen' value for the simulation for each of these parameters is not able to be plotted as it does not lie within the range of the posterior samples. The model believes there is most likely to be very little correlation between the measurement errors (the density in the right plot is centred at 0) albeit with a fair amount of uncertainty, and the model has overestimated the measurement error for both explanatory variables, by a factor of 100 for the first explanatory variable and a factor of 10 for the second.

The posterior densities for the model coefficients β_0, β_1 and β_2 are now considered in Figure 4.9. This clearly shows there is an issue in the model, with only the intercept term having its chosen value as a high-density value in the posterior distribution.

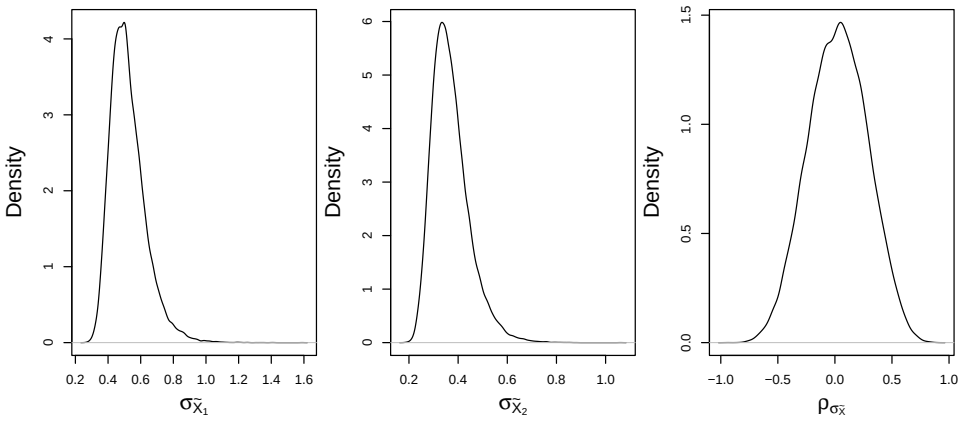

Figure 4.11: Posterior densities of the standard deviation and correlation parameters corresponding to the precision matrix $T_{\tilde{X}}$ (the left plot is the standard deviation on the for the first explanatory variable, the middle plot is the standard deviation on the second explanatory variable, and the right plot is the correlation between the two explanatory variables). This corresponds to the model with I_2 as the scale matrices for the Wishart priors on $T_{\tilde{X}}$ and T_δ.

Further inspection of the fitted values plots is required, given in Figure 4.10. The top-left plot compares the posterior densities of $\tilde{Y}_i$ with the corresponding observed data $Y_{i,j}$ for each material i; the top-right plot compares

the posterior densities of $\tilde{X}_{1,i}$ with its respective observations X_{1,i,k_1} for each i; the bottom-left plot compares the posterior densities of $\tilde{X}_{2,i}$ with its respective observations X_{2,i,k_2} for each i; the bottom-right plots compares the posterior fitted values $\widehat{Y}_i$ with the posterior true values $\tilde{Y}_i$ for each i.

	Value	$S_\delta = I_2$	$S_\delta = 10I_2$	$S_\delta = 100I_2$	$S_\delta = 1000I_2$
β_0	0.03	$[-0.133, 0.247]$	$[-0.098, 0.150]$	$[-0.029, 0.083]$	$[-0.004, 0.060]$
β_1	8	[3.046, 6.523]	[5.840, 8.795]	[7.240, 8.589]	[7.600, 8.336]
β_2	-0.5	$[-0.487, 0.174]$	$[-0.656, -0.124]$	$[-0.595, -0.376]$	$[-0.557, -0.437]$
σ_{δ_1}	0.001	[0.095, 0.128]	[0.030, 0.040]	[0.010, 0.013]	[0.003, 0.004]
σ_{δ_2}	0.01	[0.098, 0.134]	[0.032, 0.044]	[0.013, 0.018]	[0.009, 0.012]
σ_ϵ	0.005	[0.023, 0.197]	[0.019, 0.095]	[0.015, 0.047]	[0.012, 0.032]

Table 4.1: Centred 95% credible intervals for parameters of interest in the two-explanatory-variable model, given four model fits with different scale matrices S_δ on the measurement error precision matrix prior. The table demonstrates the improvement in the estimates of these parameters as the scale matrix S_δ is increased, with respect to the accuracy of the estimate and how much variability there is in the estimate.

An issue with the model becomes evident when considering the comparison between the posterior true values for the explanatory variables and their respective observed data, with the issue stemming from the first explanatory variable (top-right plot of Figure 4.10). For all true values, it is clear that the marginal posteriors of $\tilde{X}_{1,i}$ are not centred at the observed data, and the uncertainty in these marginal posteriors is large—the range of the chosen true values is from 0.01 to 0.13 (after scaling), and the 95% credible intervals for each marginal posterior are roughly spread by 0.09, which makes up 75% of the spread of the true values. Given how well the model is able to estimate the true values for the second explanatory variable (bottom-left plot of Figure 4.10, where the true values are centred at the observations, with a more suitable 95% credible interval for the posteriors of $\tilde{X}_{2,i}$ than in the case just described), it seems clear that the similar estimates for the marginal standard deviations on the measurement error (see the left and middle plots in Figure 4.8) are too similar. The comparison between the

fitted values $\widehat{Y}_i$ and $\tilde{Y}_i$ in the bottom-right plot of Figure 4.10 adds further clarity to the model issues; for each i, the variability in the fitted values is quite appreciable, particularly relative to the spread of the of the posterior for each true value. In some cases, such as $i = 9$ (Group 9 on the plot), the vertical diameter of the credible ellipse appears to be around 0.4. Consequently, an adjustment to the scale matrix S_δ for the Wishart prior on T_δ seems appropriate.

The posterior density of the precision matrix $T_{\tilde{X}}$ is also worth checking to ensure its posterior is appropriate. In a similar fashion to the treatment of the posterior of T_δ, the marginal standard deviations and correlation are extracted from each posterior sample of $T_{\tilde{X}}$, giving the marginal posterior distribution for each of these parameters, whose densities are then plotted in Figure 4.11. The posterior densities all seem appropriate, given the true values for each explanatory variable and the fixed prior mean of 0.5 for each explanatory variable.

Given the relative difference between the posterior distribution of the marginal standard deviations for the measurement error and the 'chosen' values for these parameters in the simulation, the most logical adjustment to make in the model is to change the scale matrix S_δ, with larger diagonal elements required. This is investigated here for the scaled identity matrices $10I_2$, $100I_2$ and $1000I_2$, with the results summarised in Table 4.1 with centred 95% credible intervals for relevant parameters and their respective 'chosen' values in the underlying simulation. It is evident that the changes to the scale matrix S_δ are required and help to improve the estimates of several model parameters, with larger scaled identity matrices improving the model fit.

A final plot comparing the distribution of the fitted values $\widehat{Y}_i$ with the marginal posteriors of $\tilde{Y}_i$, in the case of $S_\delta = 1000I_2$, is given in Figure 4.12. Comparing the fitted values here with those found in Figure 4.10 (bottom-right plot, where the model is fitted with $S_\delta = I_2$) shows the clear improvement in the predictions of the response variable when adjusting the scale matrix for the measurement error precision prior.

It is noted that increasing the scale matrix further to give $S_\delta = 10000I_2$ provides a marginal improvement in the model fit compared with $S_\delta = 1000I_2$, with a 95% credible interval for the model error standard deviation of [0.0111,0.02844]. With the model fitted with $S_\delta = 1000I_2$ performing sufficiently well, this is preferred in order to reduce uncertainty in the measurement error precision.

Therefore, this investigation of the scale matrix S_δ for the Wishart prior on the measurement-error precision matrix T_δ demonstrates that, for the

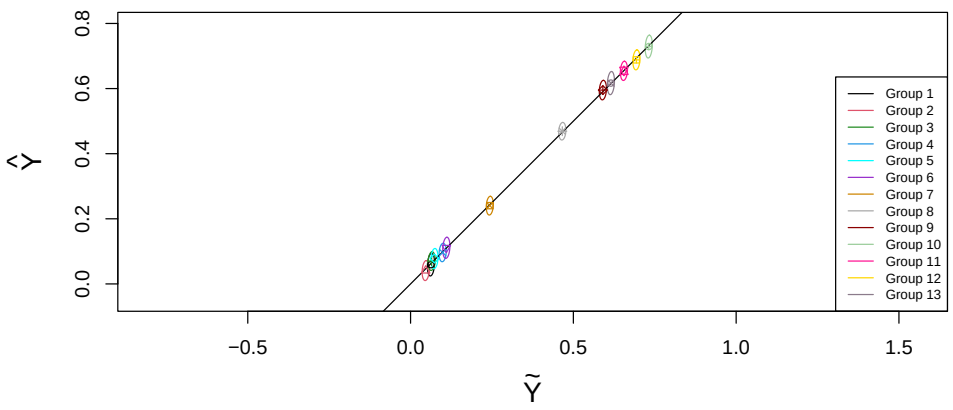

Figure 4.12: A plot comparing the distribution of the fitted values $\widehat{Y}_i$ with the corresponding marginal posterior distributions of $\tilde{Y}_i$ in the case of the simulation for the model with two explanatory variables, where a 95% credible ellipse is plotted for the fitted values distribution and marginal posterior of each material i. This corresponds to the model with I_2 and $1000I_2$ as the scale matrices for the Wishart priors on $T_{\tilde{X}}$ and T_δ respectively.

data scaling utilised here, the scaled identity matrix $1000I_2$ is appropriate for being able to estimate suitable measurement error precision values in this modelling structure. Going forward, $S_\delta = 1000I_2$ is utilised for measurement-error precision matrices in future simulations. Moreover, the above investigation demonstrates how to suitably adjust the modelling process with the use of simulated data, where the prior distributions can be tweaked so that the posterior estimates provide values that are expected given the underlying simulation. The desire here is to ensure the modelling process is working effectively before consideration of real data.

4.2.4 Two explanatory variables, correlated measurement error not assumed

This section considers the same simulated data as considered in Section 4.2.3, while not making any assumptions on correlation between the measurement error terms of the explanatory variables.

As mentioned in Section 4.2.3, in the case where the observed data for the two explanatory variables are not measured under identical circumstances, then the elements of the measurement error vector are marginally

distributed, i.e.

$$\delta_{1,i,k_1} \sim \mathrm{N}(0, \tau_{\delta_1})$$

and

$$\delta_{2,i,k_2} \sim \mathrm{N}(0, \tau_{\delta_2}).$$

The prior specification from the previous section is adjusted to take this into consideration, with $\tau_{\delta_1} \sim \text{Gamma}(a_{\delta_1}, b_{\delta_1})$ and $\tau_{\delta_2} \sim \text{Gamma}(a_{\delta_2}, b_{\delta_2})$ replacing $T_\delta \sim \text{Wishart}(S_\delta, \nu_\delta)$. The shape parameters a_{δ_1}, a_{δ_2} and rate parameters b_{δ_1}, b_{δ_2} take values of 0.001, to give an uninformative gamma prior distribution for each measurement error precision.

	'Chosen' value	a_{δ_1}, a_{δ_2}, b_{δ_1}, $b_{\delta_2} = 0.001$	$S_\delta = 1000I_2$, $S_{\tilde{X}} = I_2$
β_0	0.03	[−0.0088, 0.0639]	[−0.0037, 0.0599]
β_1	8	[7.5338, 8.3881]	[7.6001, 8.3357]
β_2	−0.5	[−0.5635, −0.4257]	[−0.5573, −0.4372]
σ_{δ_1}	0.001	[0.0044, 0.0059]	[0.0031, 0.0042]
σ_{δ_2}	0.01	[0.0096, 0.0131]	[0.009, 0.0124]
σ_ϵ	0.005	[0.0125, 0.0345]	[0.0119, 0.0316]

Table 4.2: Centred 95% credible intervals for parameters of interest in the two-explanatory-variable model, given two model fits: (1) with no assumption of correlation between the measurement errors on the explanatory variables, uninformative gamma prior distributions on the measurement errors, (2) assuming a correlation between the measurement errors on the explanatory variables, and with $S_\delta = 1000I_2$ in that case. The table demonstrates that the assumption of the correlation, and an appropriate scale matrix for the Wishart prior, gives slightly tighter credible intervals around the estimates.

The model described above, with the above prior specification, is fitted to the simulated data described in Section 4.2.3. Convergence to the posterior distribution and sufficient levels of mixing are confirmed by PSRF and effective sample size estimates. The posterior distributions of the parameters of interest from this model (with no assumption of correlated measurement error) are compared with those from the model that assumes correlated

measurement error (with $S_\delta = 1000I_2$) in Table 4.2. On the whole, the differences in the credible intervals for each parameter between the two model fits are minimal. This indicates that the modelling process is performing equally well in both cases. With the real data, it is possible that a model can be fitted with multiple explanatory variables, some of which having correlated measurement error, and some not. An example of having correlated measurement error and uncorrelated measurement error being considered for three explanatory variables is found in Section 4.2.6.

4.2.5 Two explanatory variables with interaction and quadratic terms

The final simulated-data example for the models with two explanatory variables, and the final example for Bayesian regression with a univariate response, considers a linear predictor with a quadratic term for each explanatory variables and an interaction term between the explanatory variables.

In this first attempt of fitting this model to simulated data, the true values and observed data for the explanatory variables are taken from Section 4.2.3 and utilised here. The true values $\tilde{Y}_i$ are simulated from the linear model

$$\tilde{Y}_i = 36 + 28\tilde{X}_{1,i} - 11\tilde{X}_{2,i} - 4\tilde{X}_{1,i}^2 + 5\tilde{X}_{1,i}\tilde{X}_{2,i} + 6\tilde{X}_{2,i}^2 + \epsilon_i, \qquad (4.2.5.1)$$

with values of ϵ_i simulated from a normal distribution with mean 0 and precision 4. The observed data $Y_{i,j}$ are subsequently simulated from a normal distribution with mean $\tilde{Y}_i$ and precision 1. Furthermore, the observed data for the response variable and both explanatory variables are scaled onto the range [0,1], by dividing the observed data for the response variable by 1000, the observed data for the first explanatory variable by 100, and the observed data for the second explanatory variable by 10.

The prior specification for this model is an extension of that detailed in Section 4.2.3 with the inclusion of the model coefficients β_{11}, β_{22}, and β_{12}, where each take normal distributions with mean 0 and precisions $\tau_{\beta_{11}}$, $\tau_{\beta_{22}}$, and $\tau_{\beta_{12}}$ respectively. These precisions take gamma distributions with shape and rate equal to 0.001. Note that with the assumption of correlated measurement error between the explanatory variables, and following the investigation of prior setting in Section 4.2.3, $S_\delta = 1000I_2$. The scaled 'chosen' values for the model coefficients are given by $\beta_0 = 0.036$, $\beta_1 = 2.8$, $\beta_2 = -0.11$, $\beta_{11} = -40$, $\beta_{12} = 5$ and $\beta_{22} = 0.6$, where the parameters β_{22} and β_{12} refer to the model coefficients for the squared true values of the

second explanatory variable, and the interaction between the true values on both explanatory variables, respectively.

Fitting the model with typical MCMC tuning parameters as previously noted, it is observed that the adaptation phase is not sufficient in length in order to provide optimal sampling behaviour for the MCMC algorithm. This warning suggests that, at the end of the adaptation phase of length 1000, the MCMC algorithm has not reached its optimal sampling behaviour. The recommendation in Plummer et al. (2016) is to start over the algorithm, and set up the adaptation phase to be longer. The issue is resolved with an adaptation phase of 3000. Moreover, 30000 posterior samples were sufficient for the MCMC output to have converged to the posterior distribution and to have sufficiently mixed.

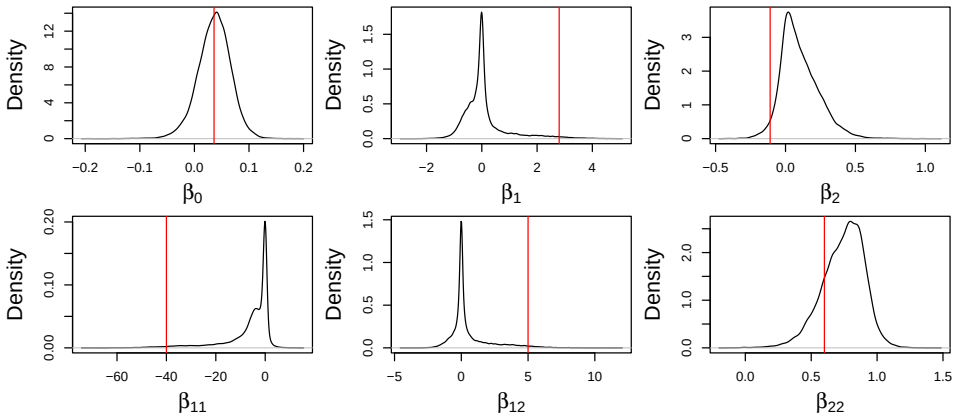

Figure 4.13: Posterior densities for β_0 (top left), β_1 (top middle), β_2 (top right), β_{11} (bottom left), β_{12} (bottom middle) and β_{22} (bottom right) from the model with two explanatory variables, quadratic and interaction terms fitted to simulated data, along with the 'chosen' values (after scaling) for each parameter from the simulation. The model assumes uninformed prior distributions.

The posterior distribution is now examined. It is noted that there are no issues with the posteriors that have been recovered for the true values $\tilde{Y}_i$, $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$ (plots comparing these with respective observed data are omitted). The posterior model coefficients are of interest to determine whether the underlying values of the linear predictor are recovered. These are given in Figure 4.13.

These plots clearly demonstrate that the majority of model coefficients have not been recovered, with the only estimate looking to be appropriate

being the intercept term, corresponding to the top left plot. While the remaining densities are not appropriate, it is clear that each scaled 'chosen' value exists in each marginal posterior density, as each plot contains the red vertical line corresponding to the scaled chosen value. For example, looking more closely at the marginal posterior density of β_{11}, where the scaled chosen value is -40, there are 43 posterior samples which lie within 1% of this scaled chosen value. Extracting the posterior samples for the remaining model coefficients that correspond to these 43 posterior samples of β_{11}, the means for each parameter given these samples are 0.0229, 2.70, -0.0365, -40.0, 5.21 and 0.514, which are clearly more representative of the scaled chosen values. In this sense, this is an effective check to show that the model has been coded correctly, with the scaled chosen values at least attainable in this posterior distribution. The issues with the current marginal posterior densities still need resolving.

While the process would surely work better with more groups (i.e. providing the model with more true values, thus more information about the relationship), this may not be a realistic possibility in a real-life scenario. It is more preferable to investigate the possibility of using more informative prior distributions, which will, in a sense, restrict the parameter space by removing implausible values for some parameters, and also expand the parameter space by including plausible values for other parameters. The following subsection delves into the process of building 'weakly informed' priors, defined here to be priors that are informed by the data scaling and its implications.

Weakly informative priors for simulated data

The uninformative prior distributions for the model coefficients and precision parameters previously assumed are insufficient for recovering the more complicated two-covariate model with a full quadratic linear predictor. These priors are therefore appropriately adjusted, given the scaling of the data.

Since the data is forced onto the range [0,1], there is a high probability that a given true value exists in this range (after scaling). Therefore, the probability statement

$$P(0 \leq \tilde{X}_{1,i}, \tilde{X}_{2,i} \leq 1) = 0.95,$$

is appropriate, that is, the probability that the true values for the explanatory variables lies between 0 and 1 is 95%. A between-materials precision can be implied for the explanatory variables using the fact that, for any

normally distributed random variable $X \sim \mathrm{N}(\mu, \sigma^2)$, where σ^2 represents the variance parameter, it is true that

$$P(-1.96\sigma \leq X \leq 1.96\sigma) \approx 0.95.$$

In other words, 95% of the density of the random variable X lies between $[-1.96\sigma, 1.96\sigma]$. Combining this probability statement and the probability relating to the simulation example above, an estimate of the between-materials standard deviation can be derived by equating the ranges, thus $3.92\sigma_{\tilde{X}} = 1$ (adding the subscript $\tilde{X}$ to clarify this standard deviation relates to $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$). Therefore, $\sigma_{\tilde{X}} = \frac{1}{3.92}$, and so an estimate of the precision parameter $\tau_{\tilde{X}}$ is given by $\tau_{\tilde{X}} = 3.92^2 = 15.3664$. Since the true values of the input variables are jointly distributed, a Wishart prior distribution is assumed for the between-materials precision matrix $T_{\tilde{X}}$, whose scale matrix $S_{\tilde{X}}$ is the diagonal matrix

$$S_{\tilde{X}} = \begin{pmatrix} 15.3664 & 0 \\ 0 & 15.3664 \end{pmatrix},$$

and degrees of freedom $\nu_{\tilde{X}}$ is assumed to be 2.

Considered now is the measurement-error precision matrix T_δ for the input variables. Note that this is investigated here despite the sensitivity analysis from Section 4.2.3 identifying $S_\delta = 1000I_2$ as suitable. The reason for this is to be consistent with the process for $T_{\tilde{X}}$, and so that there is an alignment between the measurement-error precision prior for one explanatory variable, τ_δ, and the measurement-error precision prior for multiple explanatory variables (assuming correlated measurement error), T_δ.

Analogously to $T_{\tilde{X}}$, the degrees of freedom of the Wishart prior for T_δ is assumed to be 2, so the scale matrix S_δ is of interest. Appropriate values of the diagonal elements of S_δ required further assumptions, specifically relating to the scale matrix $S_{\tilde{X}}$ for the between-materials precision. In order to carry this out, the single input variable case is considered, which is analogous to the procedure carried out for the real powder data in Section 4.4.3. That is, appropriate gamma prior distributions for τ_δ and $\tau_{\tilde{X}}$ are required.

An estimate of the mean of $\tau_{\tilde{X}}$, 15.3664, has been found above, and so

$$\frac{a_{\tilde{X}}}{b_{\tilde{X}}} = 15.3664,$$

where $a_{\tilde{X}}$ and $b_{\tilde{X}}$ represent the shape and rate of the gamma distribution respectively. The mean is used here as the point estimate as this is

easy to interpret and gives a simple relationship between the shape and rate parameters. Similarly, the shape and rate of the gamma prior for the measurement-error precision τ_δ are given by a_δ and b_δ respectively. With no mean assumption for τ_δ, an alternative assumption is sought. The first assumption is that the shape parameters $a_{\tilde{X}}$ and a_δ are equal; that is to say that the distributions for these parameters 'behave' in the same way, but on different scales (that is, the behaviour of their densities are the same, with the rate parameter determining how spread out the behaviour is). Furthermore, the shape parameters are assumed to be equal to 1 – this is chosen because of the property that the gamma distribution with shape equal to 1 is an exponential distribution, which is a suitable simplification (moreover, smaller values of shape correspond to significant density at values close to 0, whereas larger shape values correspond to more density closer to the mean, which is a stronger assumption that is avoided here to preserve the idea of 'weakly informed'). Having assumed that $a_{\tilde{X}} = a_\delta = 1$, it is therefore implied that $b_{\tilde{X}} = \frac{1}{15.3664}$, since the mean is equal to 15.3664. One final assumption is required, to help provide an estimate for b_δ, that

$$P(\tau_{\tilde{X}} \leq \tau_\delta) = 0.99,$$

i.e., that the between-materials precision is smaller than the measurement-error precision with probability 0.99. Using simulation, a suitable value for b_δ that satisfies this equation is 6.57×10^{-4}. This provides a mean estimate for τ_δ given by $\frac{a_\delta}{b_\delta} = \frac{1}{6.57 \times 10^{-4}} = 1522.07$, which is used as the diagonal elements for the scale matrix S_δ (notably, this is similar to that value identified in Section 4.2.3). The prior distribution for the measurement-error precision for the response variable is the equivalent gamma distribution from the measurement-error precision gamma distribution for the explanatory variable, i.e., $\tau_\eta \sim \text{Gamma}(a_\eta, b_\eta)$, with $a_\eta = 1$ and $b_\eta = 6.57 \times 10^{-4}$ (since we scale the data onto the same range for both outputs and inputs).

The model error precision is now of interest. The main aim for finding a suitable prior here is to allow for larger values of the model error precision to be found, and to reduce the probability for smaller values of model error precision (the justification is demonstrated after having chosen the weakly informed prior). The between-materials precision for the explanatory variable true values was estimated above by 15.3664, which can be thought of here as a lower bound for model error precision, since the response variables true values are also scaled onto the range [0,1]. This can be derived by considering the simplest possible model, where

$$\tilde{Y}_i = \beta_0 + \epsilon_i,$$

where the prior mean estimate of β_0 would be given by 0.5 (the midpoint of the range [0,1]), and the model error precision must be large enough to ensure that true values of the response at the end points [0,1] are plausibly fitted with this model. In other words, the true values of the response could exist at the endpoints of the range, so the precision must be at least 15.3664 in order to achieve that.

Any further assumptions about the model error precision are hard to justify. It would be not implausible for the model error precision to be similar to that of the measurement error precision, so suppose that the model error precision were $\tau_\epsilon = 1522.07$. Then, the model error standard deviation would be $\sigma_\epsilon = \sqrt{\frac{1}{1522.07}} = 0.0256$. With respect to previous simulation examples, this model error standard deviation seems appropriate (see the final column of Table 4.2, which compares model fits of the previous simulation with just two inputs, with 0.0256 lying within the centred 95% credible interval for model error standard deviation). Therefore, in effect, a gamma distribution for the model error precision is desired such that

- $P(\tau_\epsilon < 15.3664)$ is small,
- $P(15.3664 < \tau_\epsilon < 100)$ is not too small,
- $P(\tau_\epsilon > 1522.07)$ is not too small.

The first bullet point roughly equates to values of 15.3664 for model error precision are not possible with the current scaling, so should be minimised. The second bullet point roughly equates to poor model fits should be possible with this prior. Finally, the third bullet point roughly equates to good model fits should be possible with this prior. With the gamma distribution, it is difficult to minimise the first probability while maintaining the roughly desired probabilities for the other two bullet points. With some experimentation, the gamma distribution with shape $a_\epsilon = 0.2$ and rate $b_\epsilon = 2 \times 10^{-4}$ seems reasonable. The probability statements above are evaluated using these parameter values to give

- $P(\tau_\epsilon < 15.3664) = 0.342$,
- $P(15.3664 < \tau_\epsilon < 100) = 0.154$,
- $P(\tau_\epsilon > 1522.07) = 0.182$.

Comparing this with the uninformative prior distribution used in previous sections, $u \sim \text{Gamma}(0.001, 0.001)$, which gives

- $P(u < 15.3664) = 0.996$,
- $P(15.3664 < u < 100) = 0.00179$,
- $P(u > 1522.07) = 9.69 \times 10^{-5}$,

the improvement in making these densities more appropriate is clear.

Adjustments to the priors for the model coefficients are now discussed. The prior mean of 0 for the slope coefficients remains, but the prior mean for the intercept term is changed from 0 to 0.5 (since 0.5 is the midpoint of [0,1]). The precision parameters are now fixed instead of assuming a hyperprior distribution, to reduce the number of parameters to be estimated in the model. The intercept precision is fixed to be 0.0001, as it should be straightforward to learn about the intercept term with multiple explanatory variables also included in the linear predictor.

The precision parameters for the slope terms require more attention, ensuring that the prior is appropriately restricted, but being vague enough to not exclude plausible values. Suppose the response variable, after scaling, sufficiently covers the range of [0,1], and suppose the input variable only covers the values [0,0.1]; this is the most extreme case that can occur with this scaling method (using the maximum observation of the input variable to decide the power of 10 by which the data is divided). This extreme case provides the largest possible value the slope term could take in this scaling, which is derived by finding the line which connects the point with minimum values for the response and the input, (0,0), with the maximum values, (0.1,1). The slope of this line is just given by the difference in the response values over the difference in the input values, leading to $\beta_{\max} = \frac{1-0}{0.1-0} = 10$. Logically, the minimum value of the slope term is $\beta_{\min} = -10$. Making the appropriate assumption that the slope term is normally distributed, the probability statement

$$P(-10 \leq \beta_{\text{slope}} \leq 10) = 0.9$$

can be derived, assuming that the probability that the slope term lies between -10 and 10 is 0.9. A smaller probability is assumed here, compared with the 0.95 assumed for the above cases, since the precision parameter here is fixed, and so less certainty seems appropriate. The width of the interval is equated to $3.29\sigma_{\beta_{\text{slope}}}$, since

$$P(-1.645 \leq X \leq 1.645) \approx 0.9,$$

and so an estimate of the standard deviation of the slope parameter is $\frac{20}{3.29} = 6.08$ (to 3 s.f.), and a precision estimate is given by $\frac{1}{6.08^2} = 0.0271$ (to 3 s.f.).

Finally, attention is required for the slope term for quadratic and interaction terms. Note that, given the scaling onto the range [0,1], that squaring or taking the product of two true values provides a smaller range. If the smallest range of the input true values of [0,0.1] is taken, then range of the corresponding squared values is given by [0,0.01], and as such the corresponding slope terms have a range of [-100,100]. The standard deviation of these slope terms (corresponding to quadratic and interaction terms) is then given by $\frac{200}{3.29} = 60.79$, implying a precision of $\frac{1}{60.79^2} = 2.71 \times 10^{-4}$.

In summary, the following weakly informed prior distributions are now assumed in this simulation example:

$$\begin{aligned} \tau_\eta &\sim \text{Gamma}(1, 6.57 \times 10^{-4}), \\ \tau_\epsilon &\sim \text{Gamma}(0.2, 2 \times 10^{-4}), \\ \beta_0 &\sim \text{N}(0.5, 0.0001), \\ \beta_1, \beta_2 &\sim \text{N}(0, 0.0271), \\ \beta_{11}, \beta_{12}, \beta_{22} &\sim \text{N}(0, 2.71 \times 10^{-4}), \end{aligned} \tag{4.2.5.2}$$

with $S_{\tilde{X}} = 15.3664 I_2$ and $S_\delta = 1522.07 I_2$.

	σ_{δ_1} (=0.001)	σ_{δ_2} (=0.01)	σ_η (=0.01)	σ_ϵ (=0.005)
Uninformed priors	[0.0032, 0.0043]	[0.0090, 0.0123]	[0.0064, 0.0088]	[0.0154, 0.0480]
Weakly informed priors	[0.0026, 0.0035]	[0.0088, 0.0121]	[0.0058, 0.0079]	[0.0058, 0.0180]

Table 4.3: Centred 95% credible intervals for standard deviation parameters of interest in the full-quadratic, two-explanatory-variable model, comparing models fitted with uninformed priors and weakly informed priors. It is shown that, while the measurement-error standard deviations remain relatively unchanged between these models, the model error standard deviation is smaller in the weakly informed model, suggesting an improved model fit.

The model with two covariates and a full quadratic linear predictor is fitted again to the simulated data given in Section 4.2.5. Running the model with the same MCMC tuning parameters (notably an adaptation phase of length 3000), there are no issues with convergence and mixing in the MCMC output.

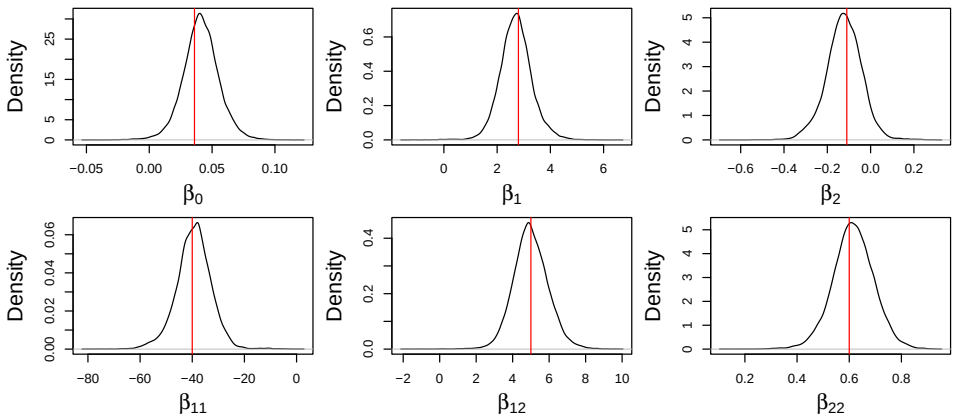

Figure 4.14: Posterior densities for β_0 (top left), β_1 (top middle), β_2 (top right), β_{11} (bottom left), β_{12} (bottom middle) and β_{22} (bottom right) from the model with two explanatory variables, quadratic and interaction terms fitted to simulated data, along with the 'chosen' values (after scaling) for each parameter from the simulation. The model assumes the weakly informed prior distributions discussed above.

For comparison with the fitted model given uninformed priors, Table 4.3 provides 95% centred credible intervals for parameters of interest. Notably, the measurement-error standard deviation posteriors compare well between the two models. More importantly, the model error standard deviation for the weakly informed model is smaller than in the previous case, providing an indication that the model fit has improved.

The plot in Figure 4.14 provides the marginal posterior densities for each model coefficient, demonstrating the model fit is greatly improved with the adjustments to the prior specification.

To conclude this section, plots of the fitted values from the uninformed-priors case (left plot) and the weakly-informed priors case (right plot) are given in Figure 4.15 for comparison. An evident improvement in the fit of the model is demonstrated in the case of considering more appropriate (weakly informed) prior distributions. In the following section, multivariate regression models are investigated with simulated data to continue to develop the modelling process.

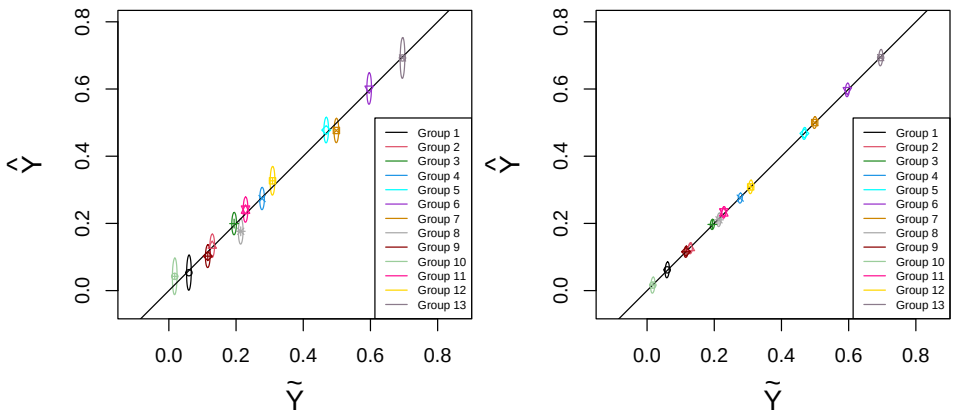

Figure 4.15: Two plots, each comparing the posterior fitted values for each group with their respective posterior true values, using 95% joint credible ellipses. The left plot corresponds to the case of uninformed prior distributions, where the model coefficients were not recovered effectively in the posterior distribution, whereas the right plot provides the fitted values from the weakly informed prior case, where the model coefficients are recovered effectively, leading to an improvement in the fitted models.

4.2.6 Multivariate regression models

The modelling continues with consideration of multivariate regression modelling. Specifically, only bivariate regression models are considered in this work, but extensions to higher dimensions are possible with relative ease. The key motivation for fitting multivariate regression models, as opposed to fitting multiple univariate regression models, is to take into account any correlation between the response variables. This is performed by building a multivariate model error vector $\boldsymbol{\epsilon}$ which takes a joint normal distribution, with mean 0 and precision matrix T_ϵ.

Firstly, the multivariate regression modelling is 'tested' here using simulated data. To demonstrate the capabilities of the modelling further, a bivariate regression model with three explanatory variables is considered here (extending from two explanatory variables, which has been investigated in Sections 4.2.3 and 4.2.4). That is, the model to be investigated is given by

$$\begin{pmatrix} \tilde{Y}_{1,i} \\ \tilde{Y}_{2,i} \end{pmatrix} = \begin{pmatrix} \beta_{0,1} & \beta_{1,1} & \beta_{2,1} & \beta_{3,1} \\ \beta_{0,2} & \beta_{1,2} & \beta_{2,2} & \beta_{3,2} \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \\ \tilde{X}_{3,i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{pmatrix},$$

where the true, 'unobservable' values of the first and second response variables are given by $\tilde{Y}_{1,i}$ and $\tilde{Y}_{2,i}$ respectively, and the model error vector is given by $\boldsymbol{\epsilon}_i = (\epsilon_{1,i}, \epsilon_{2,i})'$.

The setup of the simulated data is now considered. The true values for the first two explanatory variables are those used in Section 4.2.3, where $\tilde{X}_{1,i} = i$ for $i = 1, \ldots, 13$ and $\tilde{\boldsymbol{X}}_2 = (1, 3, 4, 5, 7, 8, 7, 4, 3, 2, 5, 6, 9)$. The observed data for these two explanatory variables, given by X_{1,i,k_1} and X_{2,i,k_2} for $i = 1, \ldots, n_g$ (with $n_g = 13$), $k_1 = 1, \ldots, n_{c_1}$ (with $n_{c_1} = 7$) and $k_2 = 1, \ldots, n_{c_2}$ (with $n_{c_2} = 7$), are taken to be the same as in Section 4.2.3, notably with the explanatory variable data assumed to have jointly distributed measurement error.

The true values for the third explanatory variable, given by $\tilde{X}_{3,i}$ for $i = 1, \ldots, n_g$, are

$$\tilde{\boldsymbol{X}}_3 = (10, 90, 20, 80, 30, 70, 40, 60, 50, 55, 45, 65, 35),$$

whose corresponding observed data, X_{3,i,k_3} are simulated such that the measurements are independent to those of the first and second explanatory variables, with $k_3 = 1, \ldots, n_{c_3}$ (with $n_{c_3} = 5$, to demonstrate that the model can cope with different numbers of replicate measurements). In summary, the observed data for the explanatory variables are of the form

$$\begin{pmatrix} X_{1,i,k_1} \\ X_{2,i,k_2} \end{pmatrix} = \begin{pmatrix} \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \end{pmatrix} + \begin{pmatrix} \delta_{1,i,k_1} \\ \delta_{2,i,k_2} \end{pmatrix}$$

and

$$X_{3,i,k_3} = \tilde{X}_{3,i} + \delta_{3,i,k_3},$$

with the assumptions that

$$\begin{pmatrix} \delta_{1,i,k_1} \\ \delta_{2,i,k_2} \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, T_\delta \right)$$

and

$$\delta_{3,i,k_3} \sim \mathrm{N}_1(0, \tau_{\delta_3}),$$

with δ_{3,i,k_3} independent from $(\delta_{1,i,k_1}, \delta_{2,i,k_2})'$. The measurement error standard deviations are chosen to be 0.1 for the first two explanatory variables, with correlation of 0.6 (that is, the simulated observations are drawn from a bivariate normal distribution with the 'chosen' true values as the mean, and covariance matrix built on marginal standard deviations of 0.1 and correlation 0.6). For the third explanatory variable, the 'chosen' measurement

error standard deviation is 1. Note that the true values for each explanatory variable and for each i are jointly distributed, that is

$$\begin{pmatrix} \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \\ \tilde{X}_{3,i} \end{pmatrix} \sim \mathrm{N}_3 \left(\begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \mu_{X_3} \end{pmatrix}, T_{\tilde{X}} \right).$$

The 'chosen' true values for the model coefficients are given by $\beta_{0,1} = 20$, $\beta_{1,1} = 11$, $\beta_{2,1} = -5$ and $\beta_{3,1} = 9$ for the first response variable, and $\beta_{0,2} = 130$, $\beta_{1,2} = -55$, $\beta_{2,2} = -103$ and $\beta_{3,2} = 637$ for the second response variable. The 'chosen' true values for both response variables, denoted by $\tilde{Y}_{1,i}$ and $\tilde{Y}_{2,i}$ respectively, are jointly simulated from a normal distribution with mean equal to the linear predictor given by

$$\begin{pmatrix} 20 + 11\tilde{X}_{1,i} - 5\tilde{X}_{2,i} + 9\tilde{X}_{3,i} \\ 130 - 55\tilde{X}_{1,i} - 103\tilde{X}_{2,i} + 63\tilde{X}_{3,i} \end{pmatrix}$$

and covariance matrix

$$\begin{pmatrix} 0.25 & 0.1 \\ 0.1 & 0.25 \end{pmatrix},$$

which is built from marginal standard deviations of 0.5 and 0.5, and a correlation of 0.4. Finally, the observed data Y_{1,i,j_1} and Y_{2,i,j_2}, for $j_1 = 1, \ldots, n_{r_1}$ (with $n_{r_1} = 7$) and $j_2 = 1, \ldots, n_{r_2}$ (with $n_{r_2} = 7$), are simulated from the 'chosen' values for the response variables, where

$$Y_{1,i,j_1} = \tilde{Y}_{1,i} + \eta_{1,i,j_1}$$

and

$$Y_{2,i,j_2} = \tilde{Y}_{2,i} + \eta_{2,i,j_2},$$

with $\eta_{1,i,j_1} \sim \mathrm{N}(0, \tau_{\eta_1})$ and $\eta_{2,i,j_2} \sim \mathrm{N}(0, \tau_{\eta_2})$ representing random measurement error. The 'chosen' precisions for these measurement error terms are 1 and 10 respectively. Note that the measurement error terms on the response variables are assumed to be independent here.

The specification of the prior and hyperprior distributions is now provided below. It is anticipated here, given that the linear predictor is less complicated than in the full quadratic two-covariate univariate regression model in Section 4.2.5, that the uninformative prior distributions are sufficient for recovering the underlying relationship in the posterior distribution.

If they prove to be insufficient, appropriate adjustments will be considered.

$$\begin{aligned}
\beta_{d_X,d_Y} &\sim \mathrm{N}(0, \tau_{\beta_{d_X,d_Y}}), \text{ for } d_X = 0,\ldots,3,\ d_Y = 1,2,\\
\tilde{\boldsymbol{X}}_i &\sim \mathrm{N}_3(\boldsymbol{\mu}_X, T_{\tilde{X}}),\\
T_\epsilon &\sim \mathrm{Wishart}(S_\epsilon, \nu_\epsilon),\\
\tau_{\eta_1} &\sim \mathrm{Gamma}(a_{\eta_1}, b_{\eta_1}),\\
\tau_{\eta_2} &\sim \mathrm{Gamma}(a_{\eta_2}, b_{\eta_2}),\\
T_\delta &\sim \mathrm{Wishart}(S_\delta, \nu_\delta),\\
\tau_{\delta_3} &\sim \mathrm{Gamma}(a_{\delta_3}, b_{\delta_3}),\\
T_{\tilde{X}} &\sim \mathrm{Wishart}(S_{\tilde{X}}, \nu_{\tilde{X}}),\\
\tau_{\beta_{d_X,d_Y}} &\sim \mathrm{Gamma}(a_{\tau_{\beta_{d_X,d_Y}}}, b_{\tau_{\beta_{d_X,d_Y}}}), \text{ for } d_X = 0,\ldots,3,\ d_Y = 1,2.
\end{aligned} \tag{4.2.6.1}$$

As in the previous subchapters, the following values are assumed within the 'uninformed' prior and hyperprior specification:

- the shape and rate parameters of the gamma distributions take the value 0.001,
- the mean of the explanatory variable true values for material i, $\tilde{\boldsymbol{X}}_i = (\tilde{X}_{1,i}, \tilde{X}_{2,i}, \tilde{X}_{3,i})'$, is set equal to 0.5 for each explanatory variable, so $\boldsymbol{\mu}_X = (\mu_{X_1}, \mu_{X_2}, \mu_{X_3}) = (0.5, 0.5, 0.5)$,
- Given the discussion in Section 4.2.3, the scale matrix S_δ for the Wishart prior distribution on the measurement error precision for the explanatory variables T_δ is set equal to $1000I_2$, with the degrees of freedom, ν_δ set equal to 2
- the scale matrix $S_{\tilde{X}}$ and degrees of freedom $\nu_{\tilde{X}}$ for the Wishart prior distribution on the between-materials precision matrix $T_{\tilde{X}}$ are set equal to I_3 and 3 respectively,
- the scale matrix S_ϵ and the degrees of freedom ν_ϵ for the Wishart prior distribution on the model error precision matrix T_ϵ are set equal to I_2 and 2 respectively.

The model is fitted with the typical MCMC tuning parameters with no issues relating to convergence and mixing in the MCMC output. The MCMC output is thus considered to be the posterior distribution, which is now examined.

Parameter	σ_{δ_1}	σ_{δ_2}	σ_{δ_3}	σ_{η_1}	σ_{η_2}
'Chosen' value	0.001	0.01	0.01	0.001	0.001
95% credible interval	[0.00317, 0.00433]	[0.00904, 0.0124]	[0.0101, 0.0149]	[0.00446, 0.00611]	[0.00445, 0.00613]

Table 4.4: Centred 95% credible intervals for measurement error standard deviation parameters of interest in the bivariate response model with three covariates. The table demonstrates the model capably allocates measurement error.

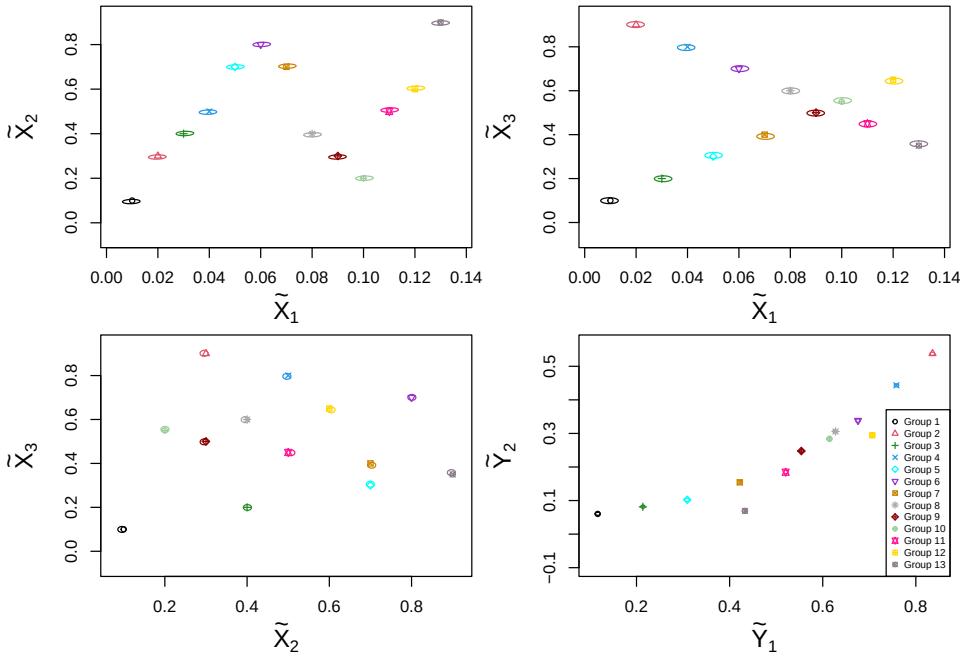

Figure 4.16: Three plots of each pair of explanatory variables showing 95% joint credible ellipses over the posterior samples for each material i, and a fourth plot providing the same summary for the true values of the response variables. The top-left plot shows the joint 95% credible ellipse for $(\tilde{X}_{1,i}, \tilde{X}_{2,i})'$ for $i = 1, \ldots, n_g$, the top-right plot showing the equivalent for $(\tilde{X}_{1,i}, \tilde{X}_{3,i})'$, the bottom-left plot showing the equivalent for $(\tilde{X}_{2,i}, \tilde{X}_{3,i})'$, and the bottom-right plot showing the equivalent for $(\tilde{Y}_{1,i}, \tilde{Y}_{2,i})$.

Converting the posterior samples of the measurement error precisions into measurement error standard deviations (that is, inverting the measurement error precisions τ_{δ_3}, τ_{η_1} and τ_{η_2} and taking the square root, and extracting the marginal standard deviations from T_δ) provides information as to whether the model has appropriately accounted for measurement error. The information in Table 4.4 corresponds to 95% credible intervals for these measurement error standard deviation posterior samples. While only the measurement error standard deviation for the second explanatory variable lies within the respective credible interval, the magnitude of the values within the other credible intervals is similar to the underlying measurement error standard deviations. This suggests that the model allocates the measurement error appropriately.

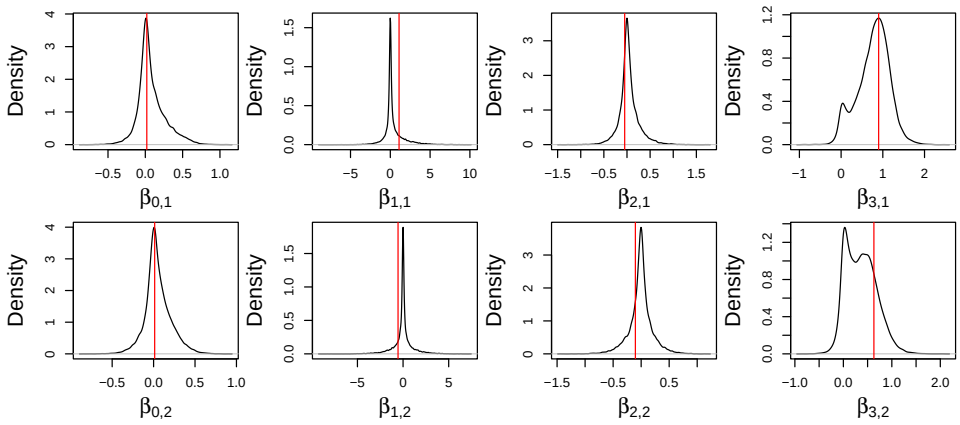

Figure 4.17: Plot of the marginal posterior densities of the model coefficients in the bivariate response model with three covariates, with red vertical lines representing the 'chosen' values for each model coefficient. Working through the plots by rows, the marginal densities appear in the following order: (1) for $\beta_{0,1}$, (2) for $\beta_{1,1}$, (3) for $\beta_{2,1}$, (4) for $\beta_{3,1}$, (5) for $\beta_{0,2}$, (6) for $\beta_{1,2}$, (7) for $\beta_{2,2}$, (8) for $\beta_{3,2}$.

One way to explore visually how well the true values have been estimated is to consider the joint posterior true values for pairs of variables, for each material (group) i – these can be displayed with joint 95% credible ellipses. These plots are provided in Figure 4.16: the top-left plot showing the joint 95% credible ellipse for $(\tilde{X}_{1,i}, \tilde{X}_{2,i})'$, the top-right plot showing $(\tilde{X}_{1,i}, \tilde{X}_{3,i})'$, the bottom-left plot showing $(\tilde{X}_{2,i}, \tilde{X}_{3,i})'$, and the bottom-right plot showing $(\tilde{Y}_{1,i}, \tilde{Y}_{2,i})$. Within each plot, the respective joint true values from the underlying simulation (for each group i) are also plotted, to demonstrate

that the model correctly captures the true values for each variable and each group.

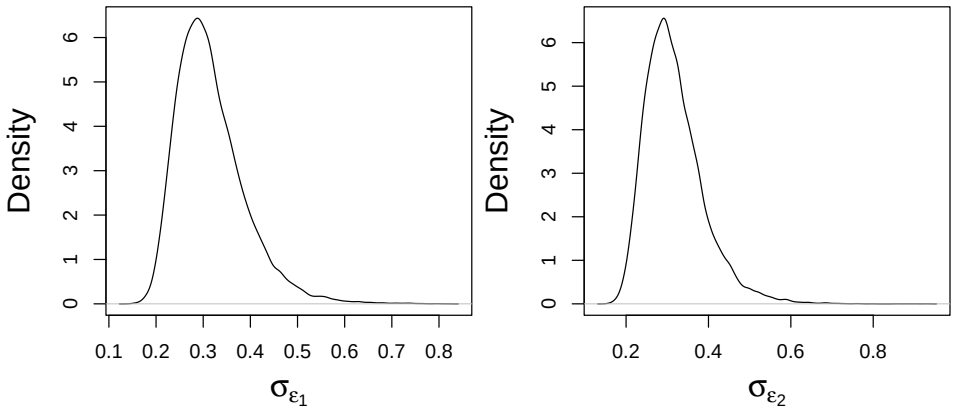

Figure 4.18: Posterior densities of the marginal standard deviations extracted from the model error precision matrix T_ϵ (with the model error standard deviation for the first response variable on the left, and for the second on the right), providing evidence that the Wishart$(I_2, 2)$ prior distribution for T_ϵ is not capable of finding suitable posterior densities for the corresponding standard deviation parameters.

A check of the posterior densities of the model coefficients $\beta_{0,1}, \beta_{1,1}, \ldots, \beta_{3,2}$ is carried out visually in Figure 4.17, and demonstrates some problems with the fitted model. It is hard to argue whether any model coefficient parameter has been estimated well—there are some cases, such as $\beta_{0,1}$ and $\beta_{0,2}$ (first column), where the 'chosen' true value appears to be well centred in the posterior density. In other cases, such as $\beta_{3,1}$ and $\beta_{3,2}$ (last column), there appears to be multimodality in the marginal posterior densities, which alludes to the model not recovering the underlying relationship appropriately, since it is expected that these posterior densities for the model coefficients are unimodal (and appear to follow a normal distribution), and the linear predictor is a relatively simple additive linear model. Finally, there are cases which show that the model is not estimating the parameter well to any extent, with the 'chosen' true value appearing to be a low-density value in the posterior density (see $\beta_{1,1}$, $\beta_{1,2}$, second column). A similar outcome is found in the model with one response variable, and a full quadratic linear predictor in two explanatory variables, from Section 4.2.5. The solution to this problem was to adjust the prior distributions for some of the parameters given the data is scaled onto [0,1].

While this could be carried out here, it is noted from Table 4.3 that the key difference between the uninformed-prior fit and the weakly informed prior fit (for the case of a full quadratic model with two input variables) was the improvement in the model error standard deviation, having adjusted the gamma prior distribution for τ_ϵ. The posterior density of the standard deviations of the model error (extracted from the posterior precision matrix samples T_ϵ) are provided in Figure 4.18. Note that, in the simplest model with just an intercept term, and scaling the data onto the range [0,1], there is an upper bound on the model error standard deviation of roughly 0.25, supposing the intercept is estimated to be 0.5, and true values of the response variable exist close to the endpoint of the range [0,1] (so, using the 2σ rule, a standard deviation of $\sigma_\epsilon = 0.25$ means 95% of the data lie within [0,1]). It is clear from Figure 4.18 that the model error standard deviation being estimated is far too large, comparing with the 0.25 upper bound suggested above.

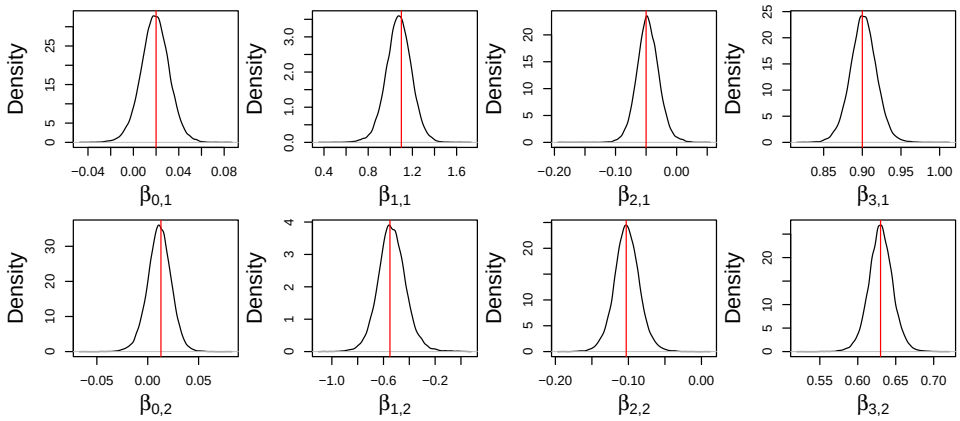

Figure 4.19: Plot of the marginal posterior densities of the model coefficients in the bivariate response model with three covariates, having fitted the model with the scale matrix $S_\epsilon = 1000I_2$ for the Wishart prior on the model error precision matrix T_ϵ. Red vertical lines represent the 'chosen' value of the model coefficient for the simulation. Working through the plots by rows, the marginal densities appear in the following order: (1) for $\beta_{0,1}$, (2) for $\beta_{1,1}$, (3) for $\beta_{2,1}$, (4) for $\beta_{3,1}$, (5) for $\beta_{0,2}$, (6) for $\beta_{1,2}$, (7) for $\beta_{2,2}$, (8) for $\beta_{3,2}$.

Given the sole issue appears to be the prior distribution for the model error precision matrix, an adjustment to this prior is considered. Having adjusted the scale matrix S_δ for the Wishart prior distribution T_δ from I_2 to $1000I_2$ in previous examples, this simple adjustment is trialled here for

the scale matrix S_ϵ, with the belief that the prior for measurement error is not too dissimilar to the prior for model error (larger values of model error precision must be accessible in the prior). If a more rigorous process is required, this will be carried out (for example, see the weakly informed prior distribution example from the previous section).

Having fitted the bivariate response model with three input variables to the corresponding simulation, with the sole adjustment of $S_\epsilon = 1000I_2$, the plots of the marginal posterior densities of the model coefficients, analogous to those in Figure 4.17 where $S_\epsilon = I_2$, are provided in the current case in Figure 4.19. These indicate the clear improvement in capturing the underlying relationship due to the adjustment to the model error precision prior.

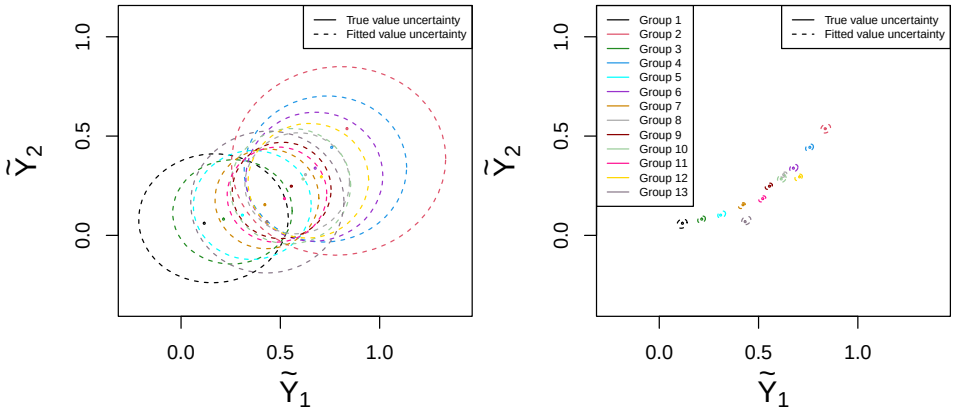

Figure 4.20: Comparison of the two bivariate regression models fits (the left plot corresponding to $S_\epsilon = I_2$, right plot corresponding to $S_\epsilon = 1000I_2$) using the fitted values of the model. In each case, the credible ellipses of the fitted values can be compared with the posterior true values of the response variables. It is clear that the right plot, with a more appropriate scale matrix for S_ϵ, provides a better model fit, with the credible ellipses of the fitted values being much smaller and lying closer to the credible ellipses for the true values.

Finally, these two model fits are compared using the plots of fitted values and true values for the response variables (see Figure 4.20). Each plot contains the joint 95% credible ellipses for the fitted values of the response

variables, given by

$$\begin{pmatrix} \widehat{Y}_{1,i,s} \\ \widehat{Y}_{2,i,s} \end{pmatrix} = \begin{pmatrix} \beta_{0,1,s} & \beta_{1,1,s} & \beta_{2,1,s} & \beta_{3,1,s} \\ \beta_{0,2,s} & \beta_{1,2,s} & \beta_{2,2,s} & \beta_{3,2,s} \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{X}_{1,i,s} \\ \tilde{X}_{2,i,s} \\ \tilde{X}_{3,i,s} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,i,s} \\ \epsilon_{2,i,s} \end{pmatrix},$$

along with the joint 95% credible ellipses for the true values of the response variables. The left plot is for the case of $S_\epsilon = I_2$, and the right plot is for the case of $S_\epsilon = 1000I_2$, with a very clear improvement in the fitted values going from $S_\epsilon = I_2$ to $S_\epsilon = 1000I_2$.

This concludes the section of fitting numerous EIV BR models to simulated data, with varying degrees of difficulty in the estimation of the linear predictor. It was clear that, in the simpler cases, the uninformative prior specification (i.e., with consideration of the scaling of the data) was sufficient for recovering the model coefficients and true values from the underlying simulation. As more explanatory variables were included, as well as more response variables, this uninformative prior specification was insufficient, however appropriate adjustments to the prior specification, which accounted for the scaling of the data, improved the model fits greatly in these more challenging cases. The overall modelling process has been sufficiently trained with the simulated data, and consideration of the real-world data relating to powder behaviour can be explored (see Section 4.4.3). The following section moves onto fitting EIV GP models to simulated data, following a similar process to refining the modelling process here, with more attention to detail on the prior specification, given the nonparametric nature of the problem and the scaling of the data.

4.3 Errors-in-variables Gaussian process on simulated data

Having considered fitting the parametric regression method of EIV BR to various simulated examples, the nonparametric regression method of errors-in-variables Gaussian processes (EIV GP) is now of interest. The obvious key difference between these methods is that the parametric regression assumes a predetermined relationship between the response variable(s) and explanatory variable(s), whereas nonparametric regression does not. Recall that 'nonparametric' in this sense does not mean that no parameters are estimated in the method, rather that there is no fixed relationship between the output(s) and input(s) to be estimated. When it comes to fitting the

model in the parametric case, to a simulated relationship, values can be chosen for the model coefficients given some linear predictor, then a parametric regression given the same linear predictor can be fitted to see if these values can be recovered in the model fit (i.e. in the posterior distribution). In the case of nonparametric regression, there is no predetermined relationship that can be 'recovered' in the posterior distribution. Instead, a good model fit must be determined simply by predicting the response variable(s) given the fitted model at different points (vectors) of the input space. With simulated data, the underlying values of the response(s) can be found at each point (or vector), providing a comparison.

In the following subsection, the Gaussian process regression is investigated with one response variable and one input variable, with three examples of different functions relating the response variable and the explanatory variable. Two extreme simulation examples are firstly considered, with the first of these examples being the basis for adjusting the prior specification, which will then be tested using another extreme example, where adjustments can be made if necessary. Of course, the main aim is to adapt the prior specification, and thus the EIV GP modelling process, to be able to cope with any possible functional relationship. Subsequently, a less extreme example is examined to demonstrate that the EIV GP modelling process deals with an 'intermediate' (and more realistic) case capably. The simulation examples become more complex by including more input variables and more response variables (see Section 4.3.4 for a simulation example with two inputs and one output, and Section 4.3.5 for an example with two inputs and two outputs).

4.3.1 First simulation

The first simulation example is a linear combination of a sine curve and exponential curve that represents the relationship between some response variable and input variable. The simulated data is discussed firstly, followed by a short overview of the EIV GP and the prior specification that is assumed. The explanatory variable has 'chosen' true values of

$$\tilde{\boldsymbol{X}} = (1, 2, 3, \ldots, 13),$$

as also fitted above in the Bayesian regression simulation examples. Observed data $X_{i,k}$, for $i = 1, \ldots, 13$ and $k = 1, \ldots, 7$ are simulated from a normal distribution with mean equal to $\tilde{X}_i$ and standard deviation 0.01. The true values for the response variable, $\tilde{Y}_i$, are simulated from a normal distribution with mean

$$18 + 20\sin(\tilde{X}_i) + \exp(0.2\tilde{X}_i) \tag{4.3.1.1}$$

and standard deviation 0.1. A plot of this function is given in Figure 4.21. Furthermore, observed data $Y_{i,j}$ (for $j = 1, \ldots, 13$) are simulated from a

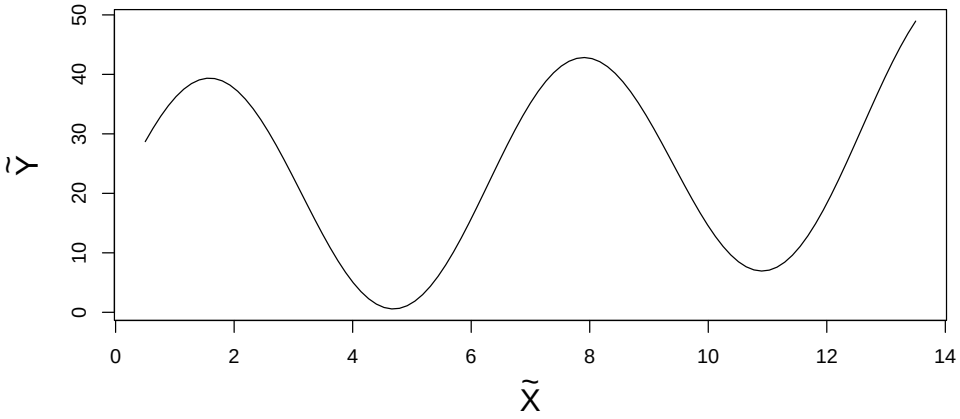

Figure 4.21: A plot of the function $18+20\sin(\tilde{X})+\exp(0.2\tilde{X})$ over the range $\tilde{X} \in [0.5, 13.5]$ (before scaling).

normal distribution with mean $\tilde{Y}_i$ and standard deviation 0.1.

The model that is fitted to this simulated data is the EIV GP regression model, where the multivariate normal distribution is assumed to relate the true values of the response variable, with some (arbitrary) constant mean function and covariance matrix built by a squared exponential covariance kernel. That is, it is assumed that

$$\tilde{\boldsymbol{Y}} = \begin{pmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_{n_g} \end{pmatrix} \sim \mathrm{N}_{n_g}\left(\begin{pmatrix} \alpha \\ \vdots \\ \alpha \end{pmatrix}, \begin{pmatrix} k(\tilde{X}_1, \tilde{X}_1) & \cdots & k(\tilde{X}_1, \tilde{X}_{n_g}) \\ \vdots & \ddots & \vdots \\ k(\tilde{X}_{n_g}, \tilde{X}_1) & \cdots & k(\tilde{X}_{n_g}, \tilde{X}_{n_g}) \end{pmatrix} + \sigma_\epsilon^2 I_{n_g} \right), \tag{4.3.1.2}$$

where the function k represents the squared exponential covariance kernel, which is the covariance kernel of preference in this work, defined as

$$k(\tilde{X}_i, \tilde{X}_{i'}) = \sigma_k^2 \exp\left\{ -\frac{(\tilde{X}_i - \tilde{X}_{i'})^2}{2l^2} \right\},$$

where σ_k^2 represents some output variance, and l is the distance-scaling parameter. Note that the scaled identity matrix $\sigma_\epsilon^2 I_{n_g}$ in the Gaussian process corresponds to the assumption that there is some model error between the true values $\tilde{\boldsymbol{Y}}$ and the functional relationship assumed by the model. Also note that, for this multivariate normal distribution that defines the Gaussian

process regression, its parameters are the mean vector and the covariance matrix (as opposed to the precision matrix). The distribution in Equation 4.3.1.2 represents the *EIV GP prior distribution*, or *EIV GP prior*, where no 'data' (in this case, data refers to the true, unobservable values $\tilde{Y}_i$ and $\tilde{X}_i$) has yet been 'observed'. The *hyperprior distributions* on the hyperparameters of interest, being $\boldsymbol{\phi} = (\alpha, \sigma_k, l, \sigma_\epsilon)$, are updated based on the EIV GP prior. The EIV GP posterior assumes that the data (i.e., the posterior true values) are now observed, with a joint posterior distribution for the hyperparameters and true values for each material i.

The EIV GP posterior can now be computed for each posterior sample s, where it is assumed that $\tilde{\boldsymbol{Y}}_s = (\tilde{Y}_{1,s}, \ldots, \tilde{Y}_{n_g,s})$ and $\tilde{\boldsymbol{X}}_s = (\tilde{X}_{1,s}, \ldots, \tilde{X}_{n_g,s})$ have been jointly observed, and so a distribution for the prediction of the response variable $\tilde{Y}_{\text{new},s}$ (i.e., given joint posterior sample s of the hyperparameters and true values) at some new input value $\tilde{X}_{\text{new}}$ can be carried out using the conditional distribution

$$\tilde{Y}_{\text{new},s} | \tilde{\boldsymbol{Y}}_s, \tilde{\boldsymbol{X}}_s, \boldsymbol{\phi}_s, \tilde{X}_{\text{new}} \sim \text{N}(m_s^*, v_s^*), \tag{4.3.1.3}$$

where

$$m_s^* = \alpha_s + \boldsymbol{k}_s(\tilde{X}_{\text{new}}, \tilde{\boldsymbol{X}}_s) V_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s)^{-1} (\tilde{\boldsymbol{Y}}_s - \boldsymbol{\alpha}_s) \tag{4.3.1.4}$$

and

$$v_s^* = v_s(\tilde{X}_{\text{new}}, \tilde{X}_{\text{new}}) - \boldsymbol{k}_s(\tilde{X}_{\text{new}}, \tilde{\boldsymbol{X}}_s) V_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s)^{-1} \boldsymbol{k}_s(\tilde{X}_{\text{new}}, \tilde{\boldsymbol{X}}_s)^T. \tag{4.3.1.5}$$

Note that the covariance between the new input value $\tilde{X}_{\text{new}}$ and posterior sample s of the true explanatory variable $\tilde{\boldsymbol{X}}_s$ is given by

$$\boldsymbol{k}_s(\tilde{X}_{\text{new}}, \tilde{\boldsymbol{X}}_s)^T = \begin{pmatrix} k_s(\tilde{X}_{\text{new}}, \tilde{X}_{1,s}) \\ \vdots \\ k_s(\tilde{X}_{\text{new}}, \tilde{X}_{n_g,s}) \end{pmatrix},$$

with the subscript s on the covariance kernel function k implying that posterior sample s for the hyperparameters replaces their respective general terms. Moreover, the matrix $V_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s)$ is equivalent to the matrix in the GP prior, but corresponds to posterior sample s for the relevant parameters, i.e.,

$$V_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s) = K_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s) + \sigma_{\epsilon,s}^2 I_{n_g}$$

where

$$K_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s) = \begin{pmatrix} k_s(\tilde{X}_{1,s}, \tilde{X}_{1,s}) & \cdots & k_s(\tilde{X}_{1,s}, \tilde{X}_{n_g,s}) \\ \vdots & \ddots & \vdots \\ k_s(\tilde{X}_{n_g,s}, \tilde{X}_{1,s}) & \cdots & k_s(\tilde{X}_{n_g,s}, \tilde{X}_{n_g,s}) \end{pmatrix}.$$

Finally, the variance at $\tilde{X}_{\text{new}}$ is given by $v_s(\tilde{X}_{\text{new}}, \tilde{X}_{\text{new}}) = \sigma_{k,s}^2 + \sigma_{\epsilon,s}^2$, since $k_s(x, x) = \sigma_{k,s}^2$ for any $x \in \mathbb{R}$. This conditional distribution is built from the joint distribution

$$\begin{pmatrix} \tilde{Y}_{1,s} \\ \vdots \\ \tilde{Y}_{n_g,s} \\ \tilde{Y}_{\text{new},s} \end{pmatrix} \sim \text{N}_{n_g+1} \left(\begin{pmatrix} \alpha_s \\ \vdots \\ \alpha_s \\ \alpha_s \end{pmatrix}, \begin{pmatrix} V_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s) & \boldsymbol{k}_s(X_{\text{new}}, \tilde{\boldsymbol{X}}_s)^T \\ \boldsymbol{k}_s(X_{\text{new}}, \tilde{\boldsymbol{X}}_s) & v_s(X_{\text{new}}, X_{\text{new}}) \end{pmatrix} \right).$$

Moreover, the observed data for both the response variable and the input variable are noisy observations of their respective true values, where the noise is attributed to measurement error, that is,

$$Y_{i,j} = \tilde{Y}_i + \eta_{i,j}$$

and

$$X_{i,k} = \tilde{X}_i + \delta_{i,k},$$

where $\eta_{i,j} \sim \text{N}(0, \tau_\eta)$ and $\delta_{i,k} \sim \text{N}(0, \tau_\delta)$. Note that the parameters of these measurement error normal distributions are the mean and the precision (as opposed to the variance).

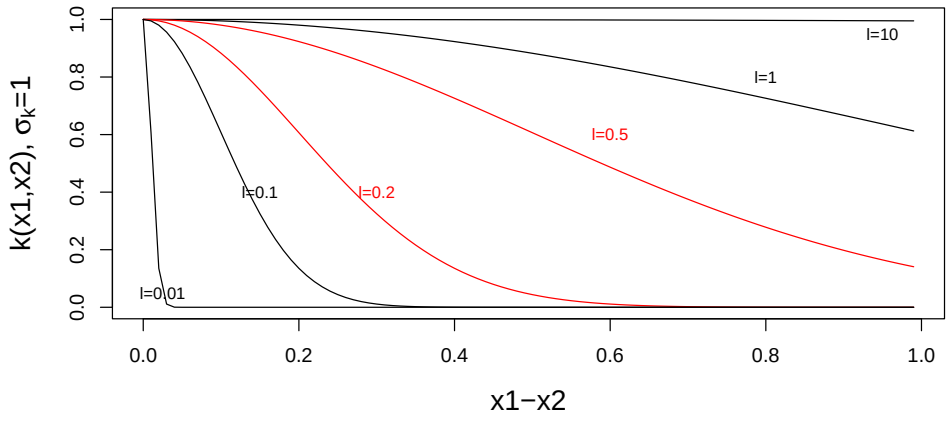

Figure 4.22: Plot of the squared exponential covariance kernel k for different values of the distance-scaling parameter l, with the signal variance σ_k^2 set equal to 1.

The specification of the prior and hyperprior distributions is

$$\begin{aligned} \sigma_\epsilon &\sim \text{Gamma}(a_{\sigma_\epsilon}, b_{\sigma_\epsilon}), \\ \sigma_k &\sim \text{Gamma}(a_{\sigma_k}, b_{\sigma_k}), \\ l &\sim \text{Gamma}(a_l, b_l), \\ \alpha &\sim \text{N}(0, 0.01), \\ \tilde{X}_i &\sim \text{N}(\mu_X, \tau_{\tilde{X}}), \\ \tau_\eta &\sim \text{Gamma}(a_\eta, b_\eta), \\ \tau_\delta &\sim \text{Gamma}(a_\delta, b_\delta), \\ \tau_{\tilde{X}} &\sim \text{Gamma}(a_{\tilde{X}}, b_{\tilde{X}}). \end{aligned} \tag{4.3.1.6}$$

The parameters σ_ϵ and σ_k are standard deviations, which take gamma prior distributions in this prior specification. While the inverse-gamma distribution is the conjugate prior for standard deviation parameters (assuming normally distributed data), the gamma distribution is chosen here due to familiarity. As informed prior distributions are considered for the EIV GP (see Section 4.4.4), the prior distribution is placed on the corresponding precision term τ_ϵ (where $\tau_\epsilon = \frac{1}{\sigma_\epsilon^2}$), to be consistent with the informed prior distribution used in the EIV Bayesian regression. The reason for placing the prior distribution on the standard deviation σ_ϵ here, as opposed to placing a prior distribution on the precision τ_ϵ, was due to the examples in the literature which focus on placing prior distributions on either the variance term of the standard deviation term. Given the scaling of the data, the values that are expected for the standard deviation are small (certainly much smaller than 1), and so placing a prior on the standard deviation instead of the variance is chosen here (since the variance values are then even smaller in this range). The measurement error precisions take gamma prior distributions with shape and rate equal to 0.001.

Finding suitable parameters for the prior distributions of σ_k and l in the covariance kernel requires careful consideration. As described in Section 3.5.1, particularly for the distance-scaling parameter l, these are both largely dependent on the scale of the data, and so they need to be investigated further for this case of data scaling, onto the range [0,1].

The aim of the covariance kernel is to inform the GP about the spatial correlation within the input variable(s). If two input values are close together, then this suggests to the GP (*a priori*) that the corresponding output values are close together. Given the data is scaled onto the range [0,1], and that the squared exponential kernel (ignoring the signal variance σ_k^2) covers the range [0,1] (with a value of 0 indicating that there is no cor-

relation between the output values given the input values, and a value of 1 indicating that there is a large correlation between the output values given the input values), values of l can be investigated that suggest suitable values of the squared exponential covariance kernel, given distances between input values. This is explored in Figure 4.22.

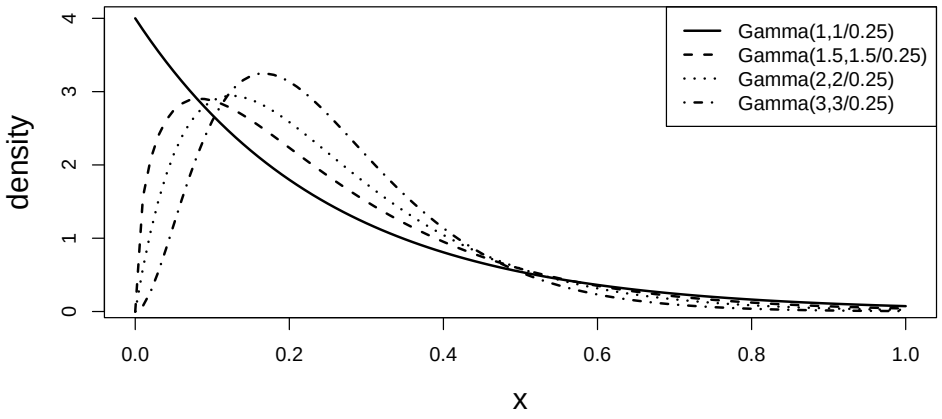

Figure 4.23: Plot of four probability density functions for gamma distributions, each with mean set equal to 0.25.

While values between $l = 0.2$ and $l = 0.5$ (closer to 0.2, given that $l = 0.5$ still does not cover the range of [0,1] for k) seems most appropriate before looking at any data, the prior distribution should also not eliminate the possibility that values such as $l = 1$ and $l = 0.1$ can occur. An ideal prior distribution would therefore have the majority of its density between $l = 0.2$ and $l = 0.5$, with some density at smaller and larger values of l too. Given that it is unlikely for the data to span the entire range of [0,1], and it is also plausible that the data span a much smaller range such as [0,0.2], smaller values of l would be favourable. Some probability density functions of gamma distributions with different shapes and rates are explored in Figure 4.23.

Each gamma density that is plotted in Figure 4.23 has its mean set equal to 0.25, as this seems an appropriate estimate of l before considering any data. The shape parameter is the parameter that changes in each of these probability density functions, with smaller shape values shifting the density closer to 0 (the mode of the distribution with shape 1 is 0), and larger shape values shifting the density closer to the mean value. While the mean seems a sensible estimate, it is sensible to keep the shape parameter fairly small, so that the density is not heavily concentrated at the mean. There is little

to distinguish the densities given the shape values of 1.5, 2 and 3.

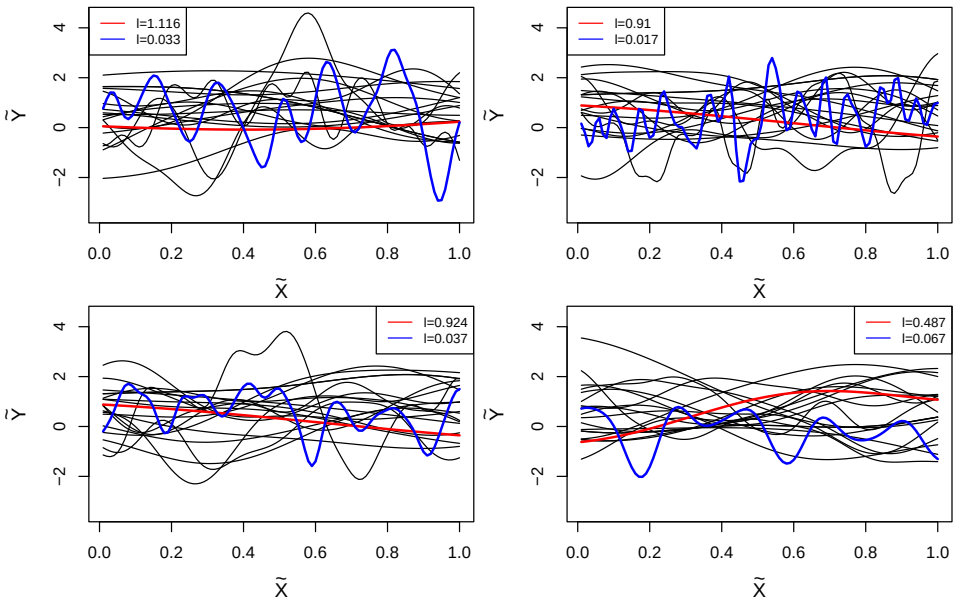

Figure 4.24: Four plots showing GP prior samples for different l, which have been drawn from gamma distributions displayed in Figure 4.23. The top-left plot is built with $l \sim \text{Gamma}(1, \frac{1}{0.25})$, the top-right with $l \sim \text{Gamma}(1.5, \frac{1.5}{0.25})$, the bottom-left with $l \sim \text{Gamma}(2, \frac{2}{0.25})$, and the bottom-right with $l \sim \text{Gamma}(3, \frac{3}{0.25})$. The input variable range is [0.01,1].

It is possible that more information could be gathered by considering plots of samples from a GP prior, with these different shape parameters for the distance-scaling parameter. A first attempt at this is provided in Figure 4.24, where each plot displays 20 samples from the Gaussian process prior, assuming a mean of 0.5 (the midpoint of the range of the data), and covariance matrix informed by the squared exponential covariance kernel. In each case, 20 samples have been simulated from the gamma distributions displayed in Figure 4.23 to provide a value of l, then the squared exponential covariance kernel is evaluated with this sample of l, $\sigma_k^2 = 1$, and 100 values of an input variable ranging through [0.01,1]. Finally, one sample of the GP prior is simulated for each sample of l.

Several observations can be drawn from the plot in Figure 4.24. Firstly, note the samples provide a good variety of behaviours that could be sampled from the GP prior, over this range. Some samples, like those samples highlighted in red (which correspond to the largest sample of l from the

respective gamma distribution) do not change significantly as the value of the input variable changes. On the other hand, when one considers the GP prior samples that correspond to the smallest values of l that are sampled from each prior (highlighted in blue), these show how the GP prior can vary quite appreciably even with relatively small changes in input values. There does not appear to be much difference between the four gamma distributions that have been considered.

A second attempt at sampling from the GP prior is considered below, with the difference being the range of values for the input variable is adjusted to [0.01,0.11]. The first attempt considered the maximum possible range of values an input variable could take, given the scaling of the data. In this second attempt, the input variable takes the minimum possible range of values, to demonstrate how this relates to the values of l that are considered. The analogous plot to Figure 4.24 is provided in Figure 4.25, where the range of the input variable is [0.01,0.11].

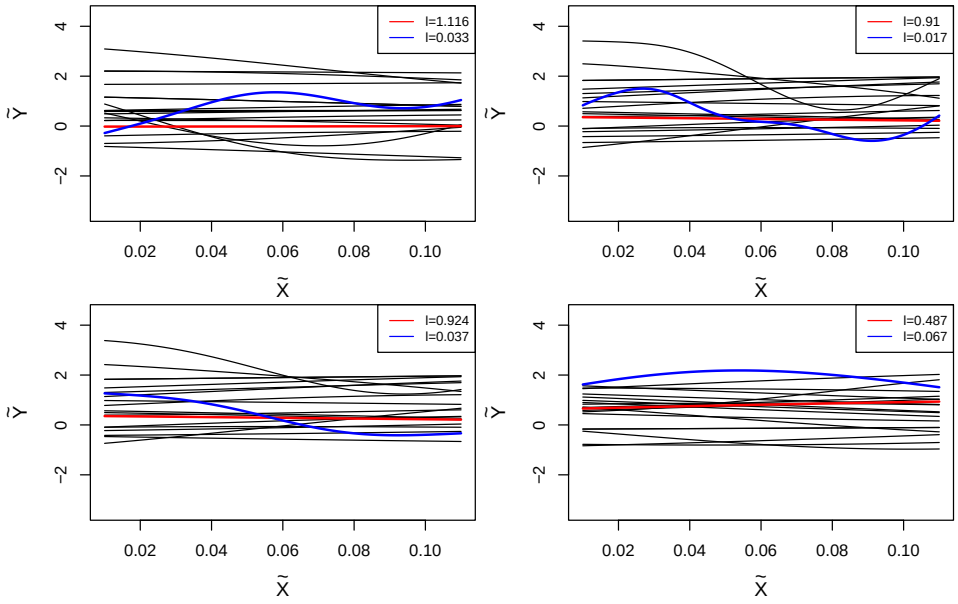

Figure 4.25: The analogous plot to Figure 4.24 now with the input variable range of [0.01,0.11].

The same prior distributions for l, now over the minimum range of input values, shows that the variety of behaviour in the GP prior samples does not carry over from the previous case. Although this case is more extreme, it would suggest that having more density at smaller values of l would be

preferable if the range of input values were small. Since there seems to be little to differentiate the priors with shape parameters 1.5, 2 and 3, the option with shape 2 is chosen, and possible changes can be made (to either the mean or the shape) if necessary.

The prior distribution for σ_k is now discussed. Note that the response variable is scaled onto the range [0,1], and note the range of values that the response variable can take a priori from either Figure 4.24 or Figure 4.25. This indicates that the magnitude of the variance in the GP prior is too large, and should be scaled to be smaller, so the gamma prior distribution for σ_k needs to have almost all of its density below 1. The influence of σ_k is demonstrated in Figure 4.26, with the GP prior simulated analogously to that in Figure 4.25, but with the top-left plot fixing $\sigma_k = 0.75$, the top-right plot at $\sigma_k = 0.5$, the bottom-left plot at $\sigma_k = 0.25$, and the bottom-right plot at $\sigma_k = 0.1$. Within each subplot, horizontal dotted lines at $\tilde{Y} = 0$ and $\tilde{Y} = 1$ are added, which are the limits of the data after scaling.

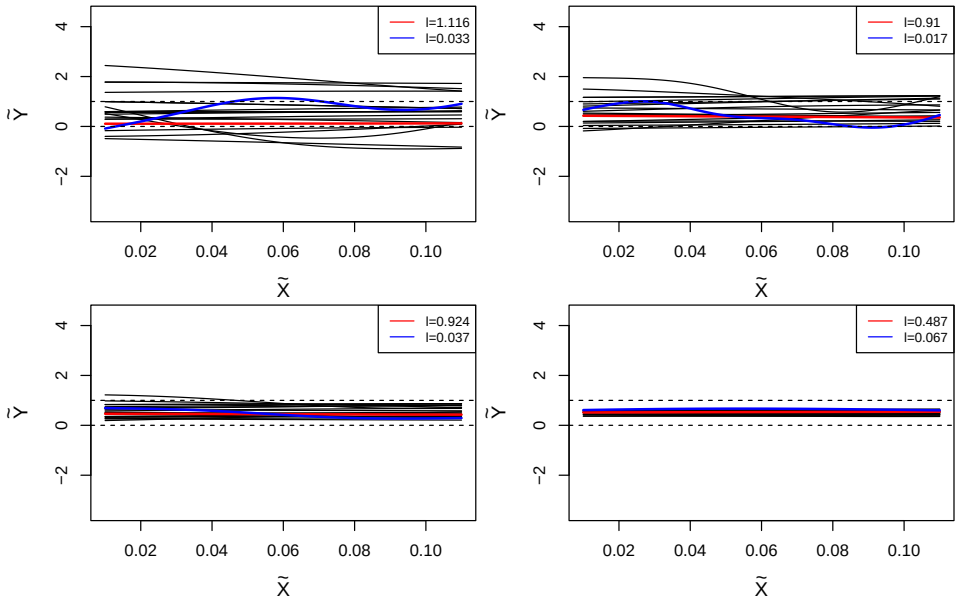

Figure 4.26: The analogous plot to that of Figure 4.25, but with the signal standard deviation term σ_k fixed at different values for each subplot—the top-left plot fixing $\sigma_k = 0.75$, the top-right plot at $\sigma_k = 0.5$, the bottom-left plot at $\sigma_k = 0.25$, and the bottom-right plot at $\sigma_k = 0.1$.

The plots in Figure 4.26 suggest that a value of $\sigma_k = 0.5$ (top-right plot) is too large a value for σ_k—as this value decreases, the range of values the

response variable can take decreases, and the range of values of the response variable at $\sigma_k = 0.5$ exceeds the range for many of the GP prior samples. A value of $\sigma_k = 0.25$ looks appropriate. Values of σ_k too close to 0 shrink the range of the response variable in the GP prior quite excessively (see the bottom-right plot of Figure 4.26), suggesting that a shape value of the gamma prior distribution for σ_k should not be too small, and certainly larger than 1. The mode of the gamma distribution with shape 3 and mean 0.4 (so rate of 7.5) is close to 0.25, so this seems like a good starting point.

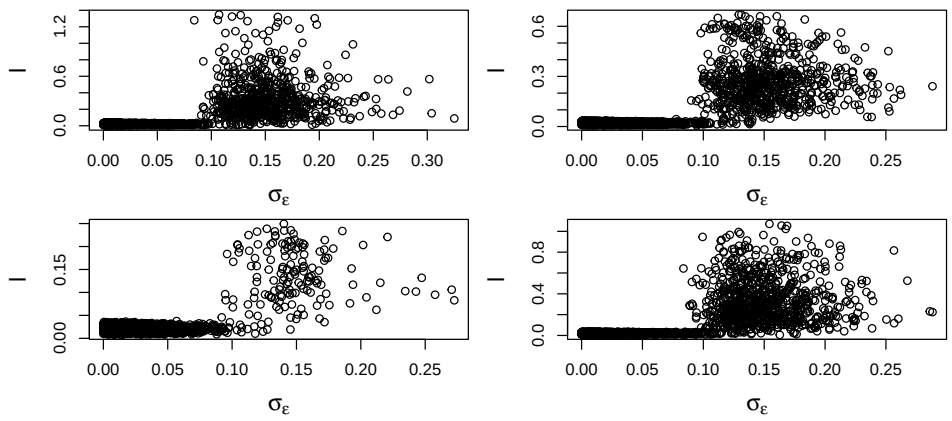

Figure 4.27: Four plots of the joint posterior samples of l against σ_ϵ for each of the parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(1, \frac{1}{0.1})$, $l \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. These show a clear relationship between these two parameters.

The final hyperparameters that need to be fixed are for the gamma prior distribution on the model error standard deviation σ_ϵ. Given the scaling of the data onto the range [0,1], the maximum standard deviation value is roughly 0.25, derived from assuming that 95% of the data lie within 2 standard deviations either side of some mean value, which can be assumed to be 0.5 before considering data. So, the gamma prior distribution for σ_ϵ should have very little density at values greater than 0.25. Moreover, it is expected that σ_ϵ is small, and so a shape parameter of 1 (which gives a mode of 0) is appropriate. Assuming a mean of 0.1 puts the probability of the prior density being larger than 0.25 at 0.082 (to 3 decimal places), which could be made smaller if deemed necessary from simulation examples (i.e.,

thus shifting the mean closer to 0).

The EIV GP is now fitted to the simulated data introduced earlier in Section 4.3.1, with the hyperprior distributions $\sigma_\epsilon \sim \text{Gamma}(1, \frac{1}{0.1})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and $l \sim \text{Gamma}(2, \frac{2}{0.25})$. The MCMC output is checked for convergence to the posterior distribution using PSRF, and shows that the parameters l and σ_ϵ have converged, but are close to 1.1, suggesting that convergence is not secure (mean estimate of 1.06 and 1.05 and upper bound on 95% confidence interval of the estimate of 1.08 and 1.09 respectively). A check of the mixing of the parameters with effective sample size estimates shows that the minimum estimate for any parameter from any chain is 47.9 (to 3 significant figures), corresponding to σ_ϵ. This estimate is quite small, relative to the posterior sample size from each chain of 20000. By definition, the MCMC output has converged, so the samples from each parallel chain can be combined, as they are all samples from the posterior distribution. In the same way, the *effective sample size* estimates can be combined, and provides an estimate of σ_ϵ over the four parallel chains of 682. Given that σ_ϵ is close to not converging, it would suggest that there could be an issue in the model, which is investigated further below. The minimum effective sample size from any chain for the parameter l is 48.0, and the total effective sample size over the four chains is 665—it is noted that the two parameters which seem to be causing issues are σ_ϵ and l.

The joint behaviour of σ_ϵ and l is of interest. Four plots of the joint posterior samples of σ_ϵ and l, for each of the parallel chains, are given in Figure 4.27. It is evident that there is some relationship here between these two parameters—if σ_ϵ is small, this forces l to be small also. This relates back to what is observed in Figure 4.24, where smaller values of l, like those corresponding to GP prior samples highlighted in blue in the plots, correspond to the response variable being sensitive to small changes in the input variable. In this simulation example, given the function that is to be recovered (see Figure 4.21), relatively small values of l are required, particularly with the small range on the input values (which is [0.01,0.13] after scaling). This is also dependent on the model error standard deviation σ_ϵ—if the values of σ_ϵ are larger, this allows for (but does not necessarily restrict to) larger values of l, since more error between the fit and the true values is allowed, which allows for more variability in the behaviour of the GP posterior. It is noted that, for any values of σ_ϵ greater than 0.075, values of l start to vary significantly.

The differences in the variability in the behaviour of the GP posterior is highlighted in Figure 4.28, where two plots of the GP posterior are displayed. This helps to determine whether alterations to the prior specification

are required (it is plausible that the model fit is still appropriate despite the general observation in regards to the posteriors of σ_ϵ and l). The top plot is based on 10 joint posterior samples of the hyperparameters and the true values of both response and explanatory variables, with the restriction that each sample of σ_ϵ must be less than 0.075. The black solid line represents the mean function of the GP posterior from these 10 samples (that is, the mean of each element of $\boldsymbol{m}_s^*$ over the 10 posterior samples, where $\boldsymbol{m}_s^*$ is vectorised and the response variable has been predicted at multiple input values throughout the range [0.01,0.13]), with the black dotted lines representing 95% prediction intervals of the GP posterior from these 10 samples (the prediction intervals calculated by taking the mean of each element of $\boldsymbol{m}_s^*$ and adding ± 1.96 multiplied by the square root of the respective diagonal element from V_s^*). The bottom plot demonstrates the same functions, with the restriction that the 10 joint posterior samples must correspond to samples of σ_ϵ greater than 0.075. The bottom plot also includes the mean of the GP posterior for each posterior sample considered; these are represented by the blue lines and the orange lines.

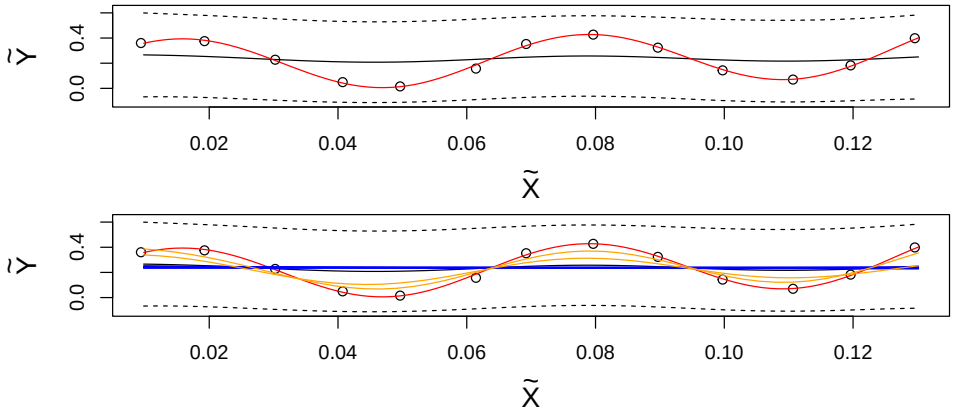

Figure 4.28: Two plots comparing the fit of the GP posterior in cases where $\sigma_{\epsilon,s}$ is smaller than 0.075 (top) and larger than 0.075 (bottom), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(1, \frac{1}{0.1})$, $l \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and uninformed gamma priors for measurement error precision and between-materials precision for the input variable.

The orange lines are noticeably different in behaviour, and provide a better representation of the underlying simulation than the blue lines. For these two posterior samples, the value of the distance-scaling parameter l is given by 0.0266 and 0.0240 (to 3 s.f.). The range of values of l when considering

the blue GP posterior means is [0.243,1.30], which demonstrates that the smaller values of l, for this example, must be found. Clearly, the variability in the GP posterior is much larger for the bottom plot, where σ_ϵ samples are greater than 0.075. In both cases, the red solid line represents the function $\tilde{Y} = 18 + 20\sin(\tilde{X}) + \exp(0.2\tilde{X})$ (after scaling) on which the simulation is based (in other words, the function that we want to recover in the GP posterior). The difference in the behaviour of the GP posterior comparing these two cases is clear, and, more importantly, the fit of the GP posterior is far worse with larger values of σ_ϵ and of l, to the extent that some samples of the GP posterior are a straight line (i.e., not close to representing the underlying simulation). Note that it is not a surprise that the model error standard deviation being smaller corresponds to a better model fit.

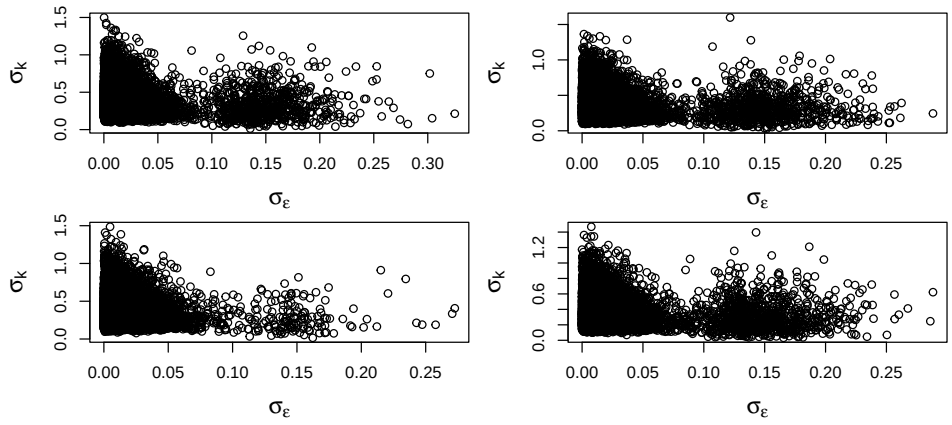

Figure 4.29: Four plots of the joint posterior samples of σ_k against σ_ϵ for each of the parallel chains, corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(1, \frac{1}{0.1})$, $l \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and uninformed gamma priors for measurement error precision and between-materials precision for the input variable.

Before considering the joint behaviour of σ_k and σ_ϵ, it is fair to assume that a relationship could exist between these two parameters as they both capture some variability in the output variable, whether it be due to signal or due to noise. A plot of the joint posterior samples for these two parameters from each of the four parallel chains is given in Figure 4.29.

Consider the joint posterior samples where σ_ϵ is less than 0.075. This indicates a relationship between these two variables hinted above, that as σ_ϵ increases up to this values of 0.075, the typical value of σ_k decreases (interestingly, even for very small σ_ϵ, close to 0, very small values of σ_k can

also be recovered).

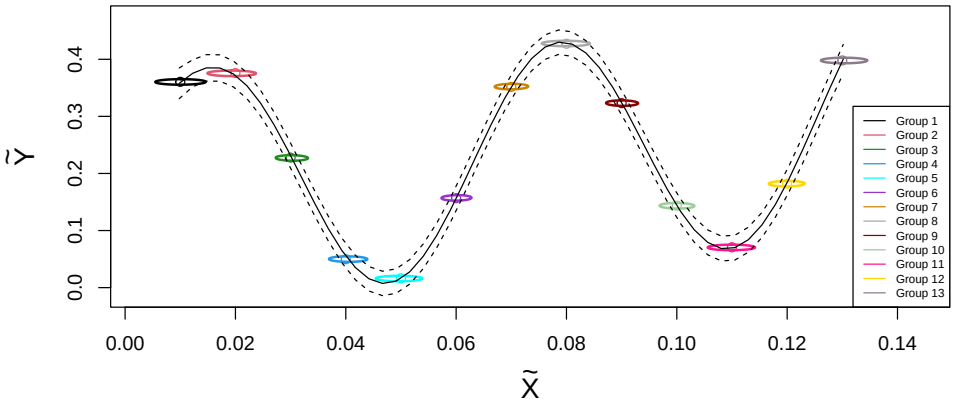

Figure 4.30: A 'fitted values' plot for the EIV GP posterior given the joint posterior distribution of the hyperparameters and true values $(\tilde{Y}_i, \tilde{X}_i)'$, corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. The black dotted line represents the mean GP posterior sample over the joint posterior samples of the hyperparameters, and the black dotted lines represent the mean 95% prediction interval of the GP posterior over the joint posterior samples of the hyperparameters. The uncertainty in the true values of the response variable and input variable are demonstrated by the 95% credible ellipses over the joint marginal posterior distribution of each $(\tilde{Y}_i, \tilde{X}_i)'$. The simulation is built from the function defined in Equation 4.3.1.1.

Possible alterations to either the hyperprior distributions or to the overall parameterisation of the model (or both) could be carried out. Both possibilities have been explored, carrying out various trials with adjustments to the σ_ϵ prior with the initial parameterisation then also with the reparameterisation, as well as adjustments to the uniform priors on the scalars k_l and k_{σ_k} in the reparameterisation. These trials are detailed in Appendix B.1.

Through these trials, it is found that an appropriate prior specification, which fits this particular example well, is with the reparameterisation of $l = k_l\sigma_\epsilon$ and $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, with the prior distributions $k_l \sim \text{U}(0, 60)$, $k_{\sigma_k} \sim \text{U}(4, 200)$ and $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$. This prior specification provides the fitted model (i.e., a plot summarising the GP posterior over the posterior distribution of the hyperparameters) given in Figure 4.30. This plot

indicates that the underlying simulation has been recovered excellently by the EIV GP modelling.

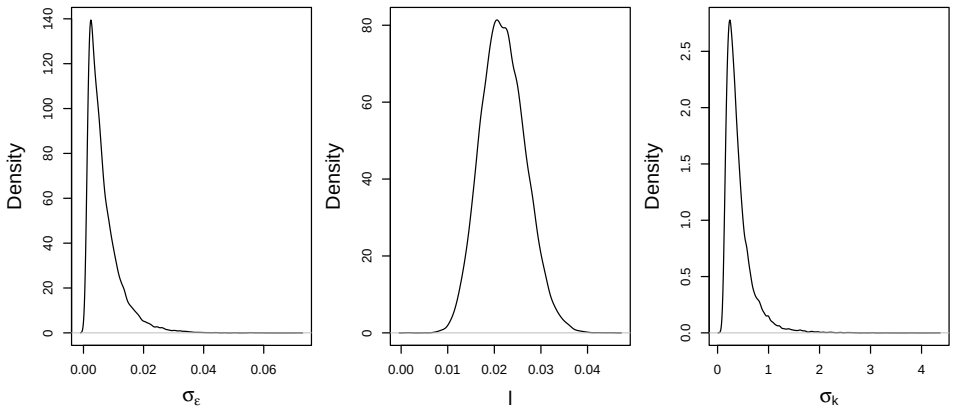

Figure 4.31: Posterior density of σ_ϵ (left plot), l (middle plot) and σ_k (right plot), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. The simulation is built from the function defined in Equation 4.3.1.1.

For comparison with later simulation examples, the posterior densities of σ_ϵ, l and σ_k are provided in Figure 4.31, and the posterior densities of k_l and k_{σ_k} are provided in Figure 4.32. It is noted that the choice of the upper limit of 60 for the uniform prior on the scalar k_l is, to an extent, a compromise which will sacrifice the predictive performance of models where larger values of l are required (i.e., where small changes in the input variable correspond to small changes in the output variable). In these cases, the model will either estimate the model error standard deviation to be larger than necessary, or underestimate of the value of l, which will cause more variance at predictions of the output variable between observations (i.e., true values) of the input variable.

Further simulation examples are now explored, to investigate how appropriate this prior specification, with the reparameterisation, is for other examples, which represent different challenges for the EIV GP modelling.

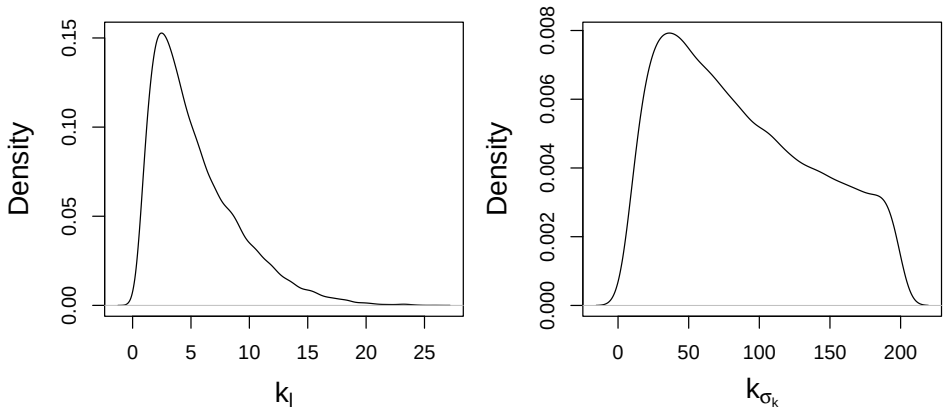

Figure 4.32: Posterior density of k_l (left plot) and k_{σ_k} (right plot), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. The simulation is built from the function defined in Equation 4.3.1.1.

4.3.2 Second simulation

In this section, the EIV GP model is fitted to data simulated using a straight-line relationship. The idea with fitting an EIV GP to a straight line relationship is to test whether the prior specification that has been found appropriate for the example in Section 4.3.1 is also appropriate for other examples which are different in the behaviour of the functional relationships. An example of a straight line relationship, combined with a large range of values for the input variable and a small range of values for the output variable (that is, a straight line with a shallow slope) provides an extreme case where appropriate values of l are likely to be relatively much larger than in the previous example (since small differences in the input variable correspond to small changes in the values of the output variable, roughly speaking). The simulation setup is detailed now to demonstrate this.

The explanatory variable has 'chosen' true values (rounded to 3 significant figures) of

$$\tilde{\boldsymbol{X}} = (0.25, 1.04, 1.83, 2.625, 3.42, 4.21, 5, 5.79, 6.58, 7.38, 8.17, 8.96, 9.75),$$

which are chosen so that, after scaling, the values cover the majority of the possible range of [0,1]. Observed data $X_{i,k}$ for $i = 1, \ldots, 13$ and $k =$

$1, \ldots, 7$ are simulated from a normal distribution with mean equal to $\tilde{X}_i$ and standard deviation 0.01. The true values for the response variable, $\tilde{Y}_i$, are simulated from a normal distribution with mean

$$4 + 0.3\tilde{X}_i \tag{4.3.2.1}$$

and standard deviation 0.1. This function is chosen so that, after scaling, the response variable covers a small range of the possible response variable values (i.e., [0.4,0.6] covers a small subspace of [0,1]). This suggests that large changes in the input variable account for relatively much smaller changes in the response variable. Furthermore, observed data $Y_{i,j}$ are simulated from a normal distribution with mean $\tilde{Y}_i$ and standard deviation 0.1.

The EIV GP model, with the following prior specification, is fitted to the above set of simulated data:

$$\begin{aligned}
\sigma_\epsilon &\sim \text{Gamma}(2, \frac{2}{0.05}),\\
\sigma_k &= k_{\sigma_k}\sigma_\epsilon,\\
l &= k_l\sigma_\epsilon,\\
k_{\sigma_k} &\sim \text{U}(4, 200),\\
k_l &\sim \text{U}(0, 60),\\
\alpha &\sim \text{N}(0, 0.01),\\
\tilde{X}_i &\sim \text{N}(\mu_X, \tau_{\tilde{X}}),\\
\tau_\eta &\sim \text{Gamma}(a_\eta, b_\eta),\\
\tau_\delta &\sim \text{Gamma}(a_\delta, b_\delta),\\
\tau_{\tilde{X}} &\sim \text{Gamma}(a_{\tilde{X}}, b_{\tilde{X}}).
\end{aligned} \tag{4.3.2.2}$$

The measurement error precisions and between-materials precision take uninformative gamma priors with shape 0.001 and rate 0.001. The EIV GP model is fitted to this set of simulated data. The MCMC output is checked for sufficient levels of mixing and convergence to the posterior distribution. The effective sample size estimates show excellent levels of mixing, with the smallest estimate for any parameter from any chain being 11610, and the PSRF shows the MCMC output has converged to the posterior distribution, with the largest upper bound for any parameter being 1.00 to 3 significant figures.

Posterior densities of the key hyperparameters in the model are provided in Figure 4.33. The main comparison here is related to the posterior density of l (middle plot), where the model has found larger values of the parameter

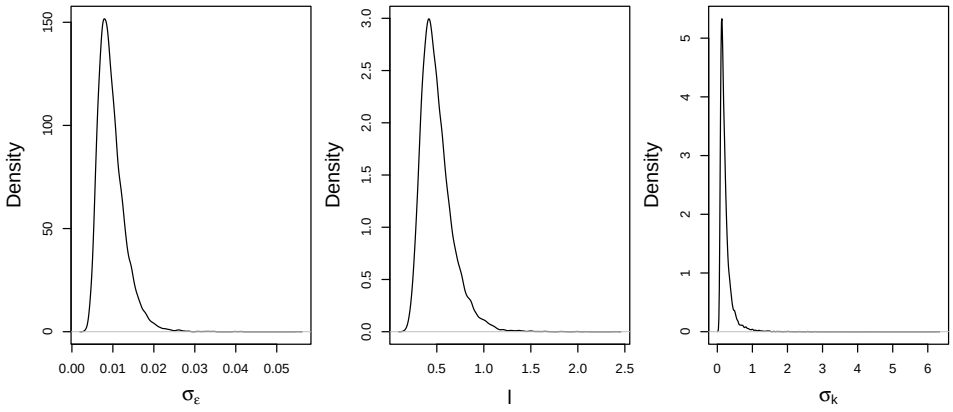

Figure 4.33: Posterior density of σ_ϵ (left plot), l (middle plot) and σ_k (right plot), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. The simulation is built from the function defined in Equation 4.3.2.1.

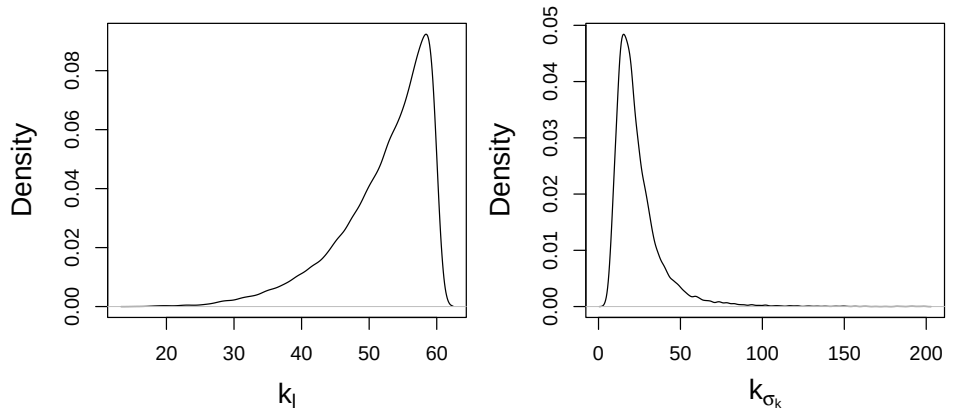

Figure 4.34: Posterior density of k_l (left plot) and k_{σ_k} (right plot), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. The simulation is built from the function defined in Equation 4.3.2.1.

(relative to the previous example, i.e. comparing the middle plot from Figure 4.31 to that of Figure 4.33), which were intended to be required when setting up the example. The signal standard deviation σ_k has a relatively similar posterior density to that of the previous example, with mild differences between the maximum posterior value, and more density concentrated at smaller values of σ_k in this second example (comparing the bottom plot of Figure 4.31 to the bottom plot of Figure 4.33). Finally, the posterior densities of the model error standard deviation σ_ϵ are comparable between Figures 4.31 and 4.33 (top plot).

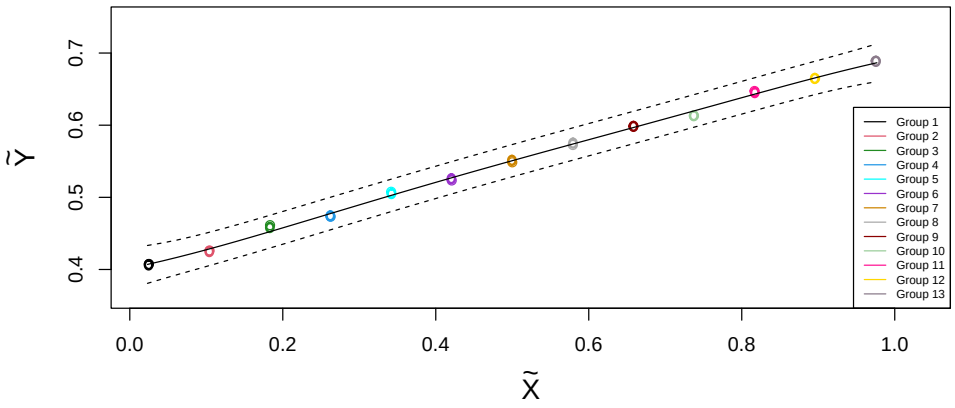

Figure 4.35: A 'fitted values' plot for the EIV GP posterior given the joint posterior distribution of the hyperparameters and true values $(\tilde{Y}_i, \tilde{X}_i)'$, corresponding to the model fitted with the prior specification from Equation 4.3.2.2. This plot is analogous to that of Figure 4.30 for the simulation built from the function defined in Equation 4.3.2.1.

The posterior densities of the scalar values k_l and k_{σ_k} are provided in Figure 4.34. The posterior density of k_l is particularly noteworthy, and not a surprise, given the suspected sacrifice that was discussed while fitting the EIV GP model to the first simulation, where it seemed necessary for the upper bound of the uniform prior for k_l to be restricted to 60. In this simulation case, it was suspected that values of l needed to be large in order to be appropriate, and so it is unsurprising that the majority of the posterior density for k_l is towards the upper limit of 60 for the prior distribution, and there is zero density at values smaller than roughly 10. To this extent, if this simulation was considered in isolation, there would appear to be justification to increase the upper bound of the uniform prior on k_l. With respect to what is observed in the cases of fitting the model to

the first simulation with either $k_l \sim \mathrm{U}(0, 80)$ or $k_l \sim \mathrm{U}(0, 100)$ as the prior, this is not advised. This nonparametric method is most likely to be applied to cases where it is expected (or seems visually clear from preliminary data plotting) that a nonlinear function may relate an output variable to an input variable. In these cases, the nonlinearity is likely to demand smaller values of l, and so the priors have been adjusted accordingly.

The GP posterior plot for this simulation is now provided in Figure 4.35. The GP posterior mean (black solid line) appears to be very suitable. It could be argued that the uncertainty in the GP posterior is larger than desired given the clear relationship between the response variable and input variable, and this has been somewhat compromised by the prior specification which caters for those model fits in which smaller values of l are required.

This prior specification is now explored on the intermediate example to demonstrate that the modelling process has been suitably trained.

4.3.3 Third simulation

Following the two extreme simulation examples considered in the previous two subsections, this simulation example here is developed to be an intermediate example between the two extremes, with a simulation function that should be easier to recover for the EIV GP model. This is due to the combination of the location of the 'chosen' true values of the input variable, and the behaviour of the underlying function. The underlying function is a linear combination of sine and exponential curves, as in Section 4.3.1, but with adjustments to the function, so as to satisfy the reasons stated above.

The 'chosen' data for this simulation are now discussed. The chosen true values (before scaling) for the input variable that are to be recovered in the posterior distribution are (to 3 significant figures)

$$\tilde{\boldsymbol{X}} = (2.50, 2.92, 3.33, 3.75, 4.17, 4.58, 5, 5.42, 5.83, 6.25, 6.67, 7.08, 7.50).$$

The observed data for each group i are simulated from a normal distribution with mean $\tilde{X}_i$ and standard deviation 0.1. The 'chosen' true values $\tilde{Y}_i$ for the output variable are simluated from the normal distribution with mean

$$0.55(80\sin(\tilde{X}_i) + \exp(0.4\tilde{X}_i) + 80) \tag{4.3.3.1}$$

and standard deviation 0.5. This function (without any noise, and before scaling) is plotted in Figure 4.36.

Using the prior specification defined in Equation 4.3.2.2, the EIV GP modelling is fitted to this example. Firstly, the MCMC output is checked

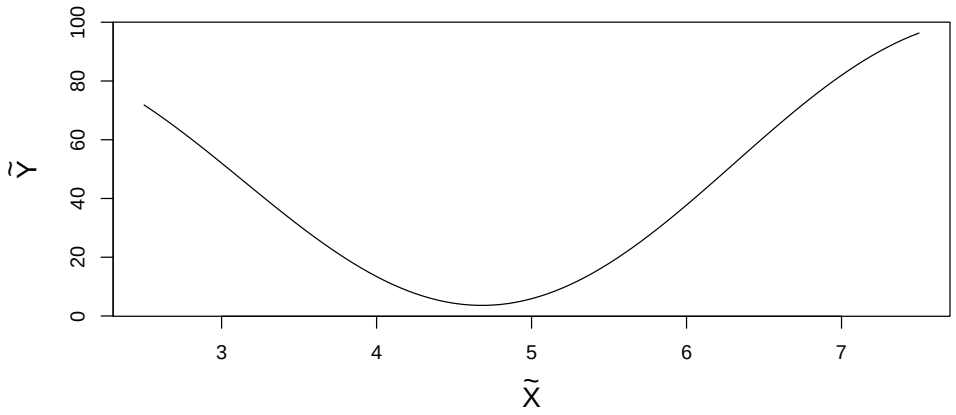

Figure 4.36: A plot of the function $0.55(80\sin(\tilde{X}) + \exp(0.4\tilde{X}) + 80)$ over the range $\tilde{X} \in [2.5, 7.5]$ (before scaling).

for convergence using PSRF, which suggests the largest upper bound for any parameter is 1.00 to 2 decimal places and confirms that the MCMC output has converged to the posterior distribution. The smallest effective sample size of any parameter from any parallel chain is 4652, suggesting that there is sufficient mixing withing the parallel chains and the posterior samples are sufficiently uncorrelated.

The model parameters are now considered. Plots of the posterior densities for σ_ϵ (top plot), l (middle plot) and σ_k (bottom plot) are provided in Figure 4.37. There does not seem to be any issues with the posterior densities here. Comparing the posterior densities here with those in Figure 4.31, it is noted that the values being found for l are significantly larger (by a factor of 10) than those found in the first simulation example, as expected given the function on which this simulation is based. The values of σ_k captured in this simulation fit appear to be slightly larger than those found in the first simulation example.

The posterior densities of the scalars k_l and k_{σ_k} are considered in Figure 4.38. Notably, in comparison with the posterior densities found in the first example (see Figure 4.32), the uniform prior of U(0,60) for k_l is explored to its full extent here, with a significant proportion of posterior density measured between values of 30 and 50, and still some density at values close to 60. In a similar vein, the posterior density of k_{σ_k} in this example has much more density at larger values of k_{σ_k}, compared with the posterior density found in Figure 4.32.

The consideration of the GP posterior plot corresponding to this model

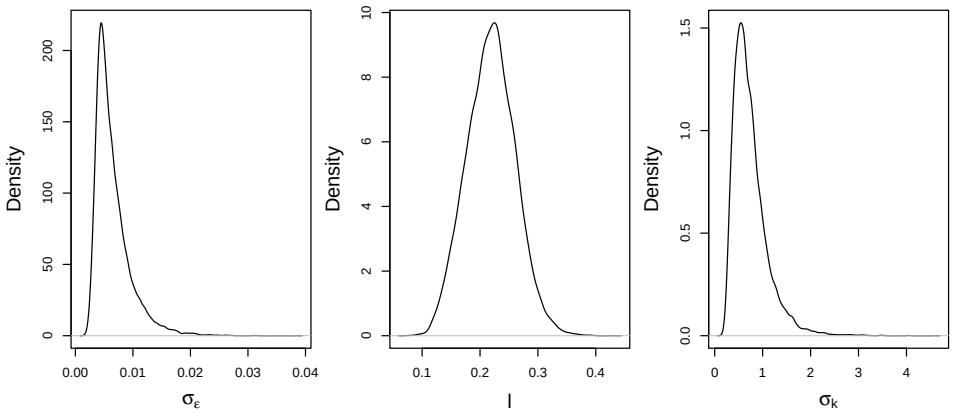

Figure 4.37: Posterior density of σ_ϵ (top plot), l (middle plot) and σ_k (bottom plot), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. The simulation is built from the function defined in Equation 4.3.3.1.

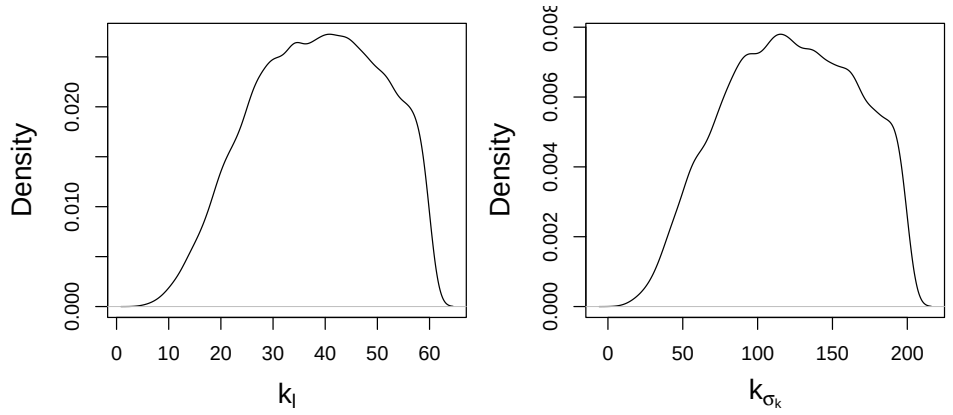

Figure 4.38: Posterior density of k_l (top plot) and k_{σ_k} (bottom plot), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed gamma priors for measurement error precision and between-materials precision for the input variable. The simulation is built from the function defined in Equation 4.3.3.1.

fit, given in Figure 4.39, provides a clear indication that the response variable is well estimated by this EIV GP fit.

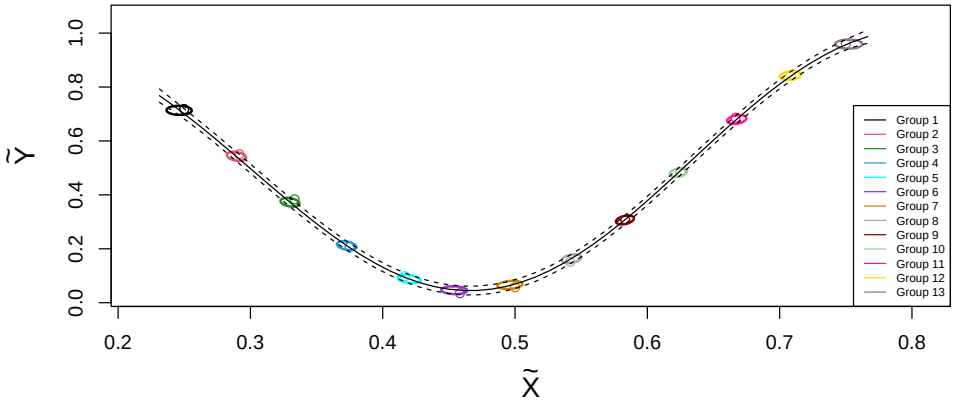

Figure 4.39: A 'fitted values' plot for the EIV GP posterior given the joint posterior distribution of the hyperparameters and true values $(\tilde{Y}_i, \tilde{X}_i)'$, corresponding to the model fitted with the prior specification from Equation 4.3.2.2, corresponding to the simulation built from the function defined in Equation 4.3.3.1.

Although only three examples are considered here, the EIV GP model copes capably with the extremeness (at opposite ends) of two of the examples, which provides confidence that any intermediate examples, like the one demonstrated in this section, should also be fitted well by the EIV GP with this prior specification. At the very least, these examples provide a solid starting point, and possible adjustments (which could be further explored) are introduced and justified with these examples. While the quality of the model fit for these examples is easily assessed by comparing with the underlying simulation, this is clearly not possible when working with real-world data. In the case of working with real data, such as in Section 4.4.2, the method of mixed LOO-CV-IC will be applied to assess the out-of-sample predictive performance of the EIV GP.

In the following section, the EIV GP modelling is extended to consider multiple input variables.

4.3.4 Multiple explanatory variables

The content of Section 3.5.3 is briefly summarised here, to remind the reader of the multi-input EIV GP.

Consider again the error-prone data $Y_{i,j}$, $X_{1,i,k}$ and $X_{2,i,k}$ (that is, the observed replicate measurements $X_{1,i,k}$ and $X_{2,i,k}$ for a given material i and replicate measurement k are carried out in identical experimental conditions, e.g. using the same subsample of powder i), which are of the form

$$Y_{i,j} = \tilde{Y}_i + \eta_{i,j},$$

with $\eta_{i,j} \sim \mathrm{N}(0, \tau_\eta)$, and

$$\begin{pmatrix} X_{1,i,k} \\ X_{2,i,k} \end{pmatrix} = \begin{pmatrix} \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \end{pmatrix} + \begin{pmatrix} \delta_{1,i,k} \\ \delta_{2,i,k} \end{pmatrix},$$

with

$$\begin{pmatrix} \delta_{1,i,k} \\ \delta_{2,i,k} \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, T_\delta \right).$$

The multi-input EIV GP prior assumes that the true values for each group of the response variable are multivariate normally distributed, with some mean vector $m(\tilde{X}) = \boldsymbol{\alpha} = (\alpha, \ldots, \alpha) \in \mathbb{R}^{n_g}$ and some covariance matrix $V(\tilde{X}, \tilde{X}) = K(\tilde{X}, \tilde{X}) + \sigma_\epsilon^2 I_{n_g}$, where

$$K(\tilde{X}, \tilde{X}) = \begin{pmatrix} k(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,n_g}) \\ \vdots & \ddots & \vdots \\ k(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,n_g}) \end{pmatrix},$$

for some covariance kernel function k. Note that the vector $\tilde{\boldsymbol{X}}_{,i} = (\tilde{X}_{1,i}, \tilde{X}_{2,i})'$, i.e. the true values for group i for the two input variables. The preference in this work is to implement a squared exponential automatic relevance determination kernel, denoted here as SE-ARD, which provides each input variable with its own distance-scaling parameter l_d, for $d = 1, 2$, in this case. The signal covariance parameter σ_k^2 is also included in the SE-ARD kernel in this work, giving the function

$$k_{\mathrm{ARD}}(\tilde{\boldsymbol{X}}_{,i}, \tilde{\boldsymbol{X}}_{,i'}) = \sigma_k^2 \exp\left\{ -\frac{1}{2} \sum_{d=1}^{2} \frac{(\tilde{X}_{d,i} - \tilde{X}_{d,i'})^2}{l_d^2} \right\},$$

which is used as the function k to compute the covariance kernel matrix $K(\tilde{X}, \tilde{X})$. Predictions of the response variable $\tilde{Y}_{\mathrm{new}}$ for some new true values $\tilde{\boldsymbol{X}}_{\mathrm{new}} = (\tilde{X}_{1,\mathrm{new}}, \tilde{X}_{2,\mathrm{new}})$ can be found using the EIV GP posterior, given a large random sample of the posterior distribution of the hyperparameters $\boldsymbol{\phi} = (\alpha, \sigma_k, \sigma_\epsilon, l_1, l_2)$ and the existing true values $\tilde{\boldsymbol{Y}}$ and $\tilde{X} = \begin{pmatrix} \tilde{\boldsymbol{X}}_1 & \tilde{\boldsymbol{X}}_2 \end{pmatrix}$. In

this case, the EIV GP posterior given some posterior sample s provides a sample for the distribution of predictions of $\tilde{Y}_{\text{new}}$, given by the conditional distribution

$$\tilde{Y}_{\text{new},s}|\tilde{\boldsymbol{Y}}_s, \tilde{\boldsymbol{X}}_{1,s}, \tilde{\boldsymbol{X}}_{2,s}, \boldsymbol{\phi}_s, \tilde{\boldsymbol{X}}_{\text{new}} \sim \text{N}(m_s^*, v_s^*),$$

where

$$m_s^* = m_s(\tilde{\boldsymbol{X}}_{\text{new}}) + K_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s)V_s(\tilde{X}_s, \tilde{X}_s)^{-1}(\tilde{\boldsymbol{Y}}_s - m_s(\tilde{X}_s))$$

and

$$v_s^* = v_s(\tilde{\boldsymbol{X}}_{\text{new}}) - K_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s)V_s(\tilde{X}_s, \tilde{X}_s)^{-1}K_s(\tilde{X}_s, \tilde{\boldsymbol{X}}_{\text{new}}).$$

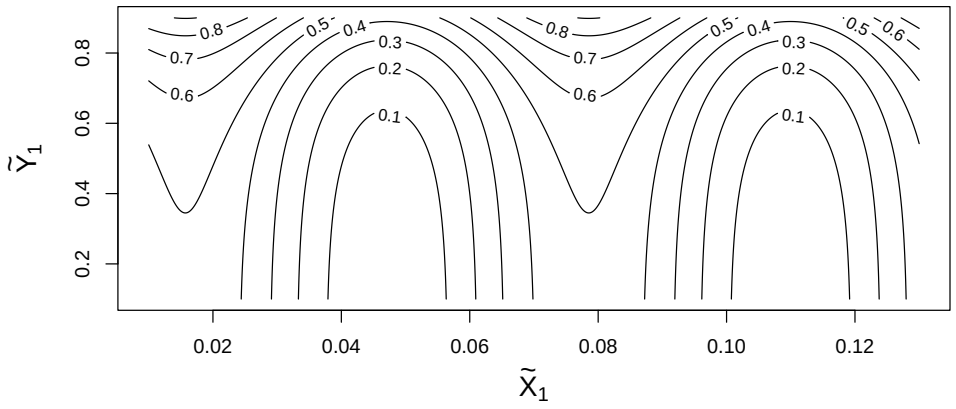

Figure 4.40: A contour plot representing the relationship between the response variable and the two input variables corresponding to that defined in Equation 4.3.4.1. The contours represent values of the response variable, the horizontal axis represents values of the first input variable, and the vertical axis represents values of the second input variable.

The subscript s refers to joint random sample s from the joint posterior distribution $p(\boldsymbol{\theta}|Y, X_1, X_2)$, where the matrices $K_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s)$ and $V_s(\tilde{X}_s, \tilde{X}_s)$ are produced by the covariance kernel function $k_{\text{ARD},s}$ given the joint random sample s for its parameters. Similarly, the covariance of the vector of new true values, $v_s(\tilde{\boldsymbol{X}}_{\text{new}})$ is given by $\sigma_{\epsilon,s}^2 + \sigma_{k,s}^2$. Moreover, the vector $\tilde{\boldsymbol{Y}}_s$ is joint random sample s of the true values for the response variable, and the matrix $\tilde{X}_s \in \mathbb{R}^{n_g \times 2}$ is the joint random sample s of the true values for the two input variables. So, for each posterior sample s, a distribution summarises the prediction of the response, and one sample of the prediction is

taken from each EIV GP posterior to provide the complete distribution of predictions $p(\tilde{Y}_{\text{new}}|\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2, \boldsymbol{\phi}, \tilde{\boldsymbol{X}}_{\text{new}})$.

An example of fitting the multi-input EIV GP to simulated data is now considered. This simulation is, in a sense, a combination of two of the simulations considered in the single-input EIV GP case, from Sections 4.3.1 and 4.3.3. The chosen true values for the first input variable are $\tilde{X}_{1,i} = i$ for $i = 1, \ldots, 13$, and the second input variable has chosen true values of $\tilde{\boldsymbol{X}}_2 = (1, 3, 4, 5, 7, 8, 7, 4, 3, 2, 5, 6, 9)$ (both before scaling). Their observed data $(X_{1,i,k}, X_{2,i,k})'$ is jointly simulated from a normal distribution with mean $(\tilde{X}_{1,i}, \tilde{X}_{2,i})'$ and covariance matrix built from marginal standard deviations of 0.01 and correlation 0.35. The true values of the response variable, $\tilde{Y}_i$, are then simulated from the equation

$$\tilde{Y}_i = 0.3(80 + 80\sin(\tilde{X}_{1,i}) + \exp(0.55\tilde{X}_{2,i})) + \epsilon_i, \tag{4.3.4.1}$$

where the chosen true values are given for $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$, then the model error term ϵ_i is simulated from a normal distribution with mean 0 and standard deviation 0.1. This relationship (after scaling), defined in Equation 4.3.4.1, is plotted as a contour plot in Figure 4.40.

Finally, the replicate observations $Y_{i,j}$ of the response variable are simulated from the normal distribution with mean $\tilde{Y}_i$ and standard deviation 0.1.

Issues are discovered in this example relating to the joint posterior densities of l_1 and σ_ϵ, as were previously observed in the single-input-variable case from Section 4.3.1. A suitable resolution, which was also investigated in the EIV BR example with two input variables and a full quadratic linear predictor (see Section 4.2.5), was to adjust the prior distributions for the between-materials precision and the measurement-error precision based on the data being scaled onto [0,1]. An adjustment to the prior distribution of the GP mean, α, was also implemented given the same justification. Further investigation of these adjustments to the prior specification is given in Appendix B.2, with the model fit corresponding to the chosen prior specification examined here.

The prior distributions for τ_η, T_δ and $T_{\tilde{X}}$ are therefore adjusted to $\tau_\eta \sim \text{Gamma}(1, 6.57 \times 10^{-4})$, $T_\delta \sim \text{Wishart}(S_\delta, 2)$ and $T_{\tilde{X}} \sim \text{Wishart}(S_{\tilde{X}}, 2)$, with $S_\delta = 1522.07 I_2$ and $S_{\tilde{X}} = 15.3664 I_2$. These derivations are discussed in Section 4.2.5. This leads to an appreciable reduction in the prevalence of the σ_ϵ and l_1 issue, where this model in previous attempts (as well as in the single-input EIV GP case) produced posterior density at larger values of σ_ϵ, which coincided with more variance in the posterior samples of l_1.

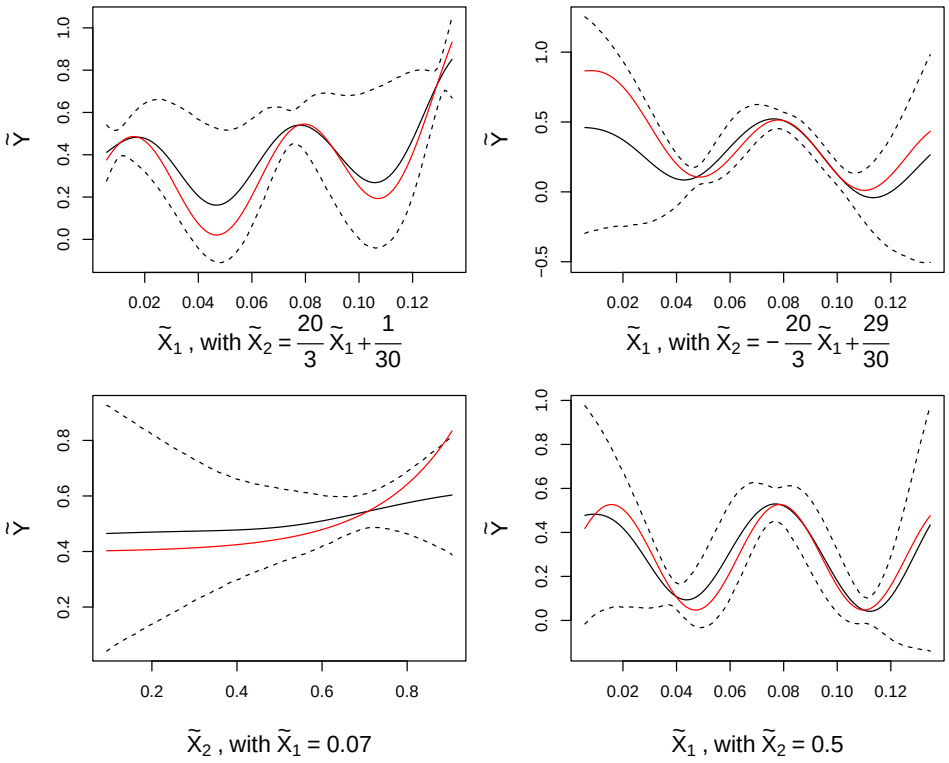

Figure 4.41: Four plots of the GP posterior for the two-input EIV GP simulation, where each plot corresponds to predictions of the GP posterior in different lines of the 2D input space. The top-left plot corresponds to the line $X_2 = \frac{20}{3}X_1 + \frac{1}{30}$ with $X_1 \in [0.01, 0.13]$; the top-right plot corresponds to the line $X_2 = -\frac{20}{3}X_1 + \frac{29}{30}$ with $X_1 \in [0.01, 0.13]$; the bottom-left plot corresponds to the line $X_1 = 0.07$ with $X_2 \in [0.1, 0.9]$; the bottom-right plot corresponds to the line $X_2 = 0.5$ with $X_1 \in [0.01, 0.13]$. In each plot, the solid black line represents the mean of the GP posterior, the dotted black line represents a 95% prediction interval for the GP posterior, and the red line represents the values of the response variable from the underlying simulation.

While this issue is not completely eliminated, it is noted that 0.164% of the posterior samples of σ_ϵ are greater than 0.075, which has been deemed here to be negligible. With sufficient mixing and convergence to the posterior both observed, the model fit is considered to be appropriate.

Four plots of the fitted model for this two-input EIV GP example are provided in Figure 4.41, where the predictions of the output variable (vertical axis) are demonstrated for different lines within the 2D plane of possible input variables. The top-left plot corresponds to the line $X_2 = \frac{20}{3}X_1 + \frac{1}{30}$ with $X_1 \in [0.01, 0.13]$ (and X_1 represented on the horizontal axis); the top-right plot corresponds to the line $X_2 = -\frac{20}{3}X_1 + \frac{29}{30}$ with $X_1 \in [0.01, 0.13]$ (and X_1 represented on the horizontal axis); the bottom-left plot corresponds to the line $X_1 = 0.07$ with $X_2 \in [0.1, 0.9]$ (and X_2 represented on the horizontal axis); the bottom-right plot corresponds to the line $X_2 = 0.5$ with $X_1 \in [0.01, 0.13]$ (and X_1 represented on the horizontal axis). In each plot, the solid black line represents the mean of the GP posterior, the dotted black line represents a 95% prediction interval for the GP posterior, and the red line represents the values of the response variable from the underlying simulation. A plot of these lines with the joint marginal posterior densities (represented by 95% credible ellipses) for the true values of the input variables for each group i is provided in Figure 4.42, to give an indication as to the changes in variability in the GP posterior.

The GP posterior predicts the response variable appropriately given the information supplied to the model (i.e., the true values of the response variable and input variables). For example, in the top-left plot of Figure 4.41, the GP posterior shows confidence in its predictions of the response variable for values close to $X_1 = 0.01$ (and therefore $X_2 = 0.1$) and $X_1 = 0.13$ (and therefore $X_2 = 0.9$), which is logical given the corresponding line intersects the true values of Groups 1 and 13 in the plot in Figure 4.41. Conversely, there is significant uncertainty in the predictions of the response variable in the top-right plot at smaller values of X_1, such as $X_1 = 0.01$, with the corresponding value of X_2 being 0.9 – evidently from Figure 4.41, there are no groups in this vicinity, hence the GP posterior has not learned about the behaviour of the function in this region.

The following section consider the extension of the single-output Gaussian process to the multi-output Gaussian process.

4.3.5 Multi-output EIV GP

The multi-input EIV GP is extended to multiple output variables, providing the multi-output EIV GP (often referred to here as EIV MOGP). The theory

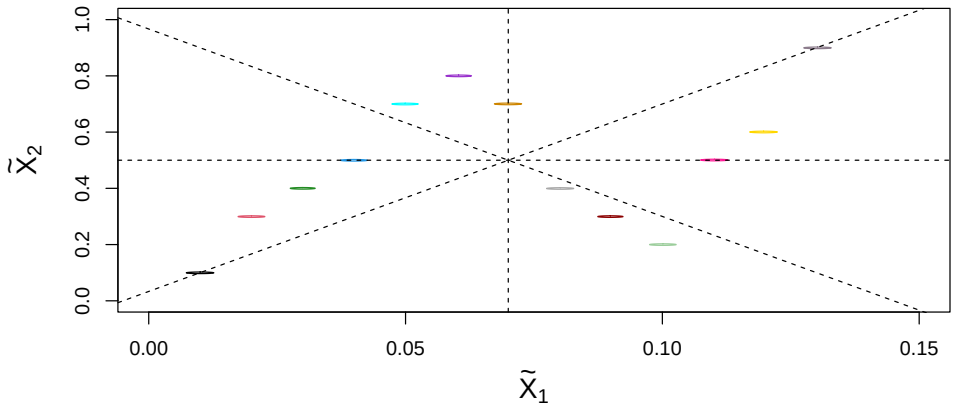

Figure 4.42: A plot showing the joint marginal posterior densities, for each group, of the true values for the pair of input variables in the two-input EIV GP example, for comparison with the four lines which are used for comparing predictions of the response variable with the underlying simulation.

of this regression model is covered in Section 3.5.5, and a summary of the content is covered here as a reminder, with two output variables and two input variables.

Naturally, consider the true values of groups $i = 1, \ldots, n_g$ for the two output variables $\tilde{Y}_{1,i}$ and $\tilde{Y}_{2,i}$, and for the two input variables $\tilde{X}_{1,i}$ and $\tilde{X}_{2,i}$. Relating the true values for the response variables to those of the input variables, while accounting for any correlation between the true values $\tilde{Y}_{1,i}$ and $\tilde{Y}_{2,i}$, is carried out in this work using a linear model of coregionalisation multi-output GP. Defining the matrix of true values for the output variables as $\tilde{Y} \in \mathbb{R}^{n_g \times 2}$, the vectorisation of these output variables stacks the true values for the first output variable on top of the true values for the second output variable, giving the vector $\text{vec}(\tilde{Y}) = (\tilde{Y}_{1,1}, \ldots, \tilde{Y}_{1,n_g}, \tilde{Y}_{2,1}, \ldots, \tilde{Y}_{2,n_g})'$. This vector of true values is then assumed to be multivariate normal, with mean $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'$, and covariance matrix

$$V_{\text{MOGP},\tilde{X}} = V_{K,1} \otimes K_1(\tilde{X}, \tilde{X}) + V_{K,2} \otimes K_2(\tilde{X}, \tilde{X}) + \begin{pmatrix} \sigma^2_{\epsilon_1} I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma^2_{\epsilon_2} I_{n_g} \end{pmatrix},$$

that is,

$$\text{vec}(\tilde{Y}) \sim \text{N}_{2n_g}\left(\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}, V_{\text{MOGP},\tilde{X}} \right). \tag{4.3.5.1}$$

A summary of the key features of the EIV MOGP is provided below:

- The mean of the distribution is set up to assume that the mean of the first output variable (for each true value) is not necessarily the mean of the second output variable (for each true value), hence the introduction of the vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^{n_g}$

- The Kronecker product $A \otimes B$ between two matrices A and B is informally defined to be a matrix whose dimensions are the products of the dimensions of A and B (number of rows of $A \otimes B$ given by the product of the number of rows of A and the number of rows of B, for example), and each 'block' matrix of $A \otimes B$ is the product between each individual element of the first matrix and the entirety of the second matrix, which produces a block matrix with the same dimensions of B. The Kronecker product is defined more formally in Section 3.5.5.

- the matrices $V_{K,1}$ and $V_{K,2}$ are defined to be

$$V_{K,1} = \begin{pmatrix} \sigma_{k_1}^2 & \sigma_{k_1}\sigma_{k_2}\rho_{V_K} \\ \sigma_{k_1}\sigma_{k_2}\rho_{V_K} & \sigma_{k_2}^2 \end{pmatrix} \tag{4.3.5.2}$$

and

$$V_{K,2} = \begin{pmatrix} \lambda_1^2\sigma_{k_1}^2 & \lambda_1\sigma_{k_1}\lambda_2\sigma_{k_2}\rho_{V_K} \\ \lambda_1\sigma_{k_1}\lambda_2\sigma_{k_2}\rho_{V_K} & \lambda_2^2\sigma_{k_2}^2 \end{pmatrix}, \tag{4.3.5.3}$$

respectively. That is, these represent signal covariance matrices for the two output variables. This firstly means that, the signal variance term that has been included in the covariance kernel function for the single-output EIV GPs, is moved into these matrices the EIV MOGP, and secondly, that there are two covariance kernel matrices considered in this model. Note also that it is assumed that there is a relationship between the elements of $V_{K,1}$ and $V_{K,2}$.

- The covariance kernel matrix $K_1(\tilde{X}, \tilde{X})$ is defined to be

$$K_1(\tilde{X}, \tilde{X}) = \begin{pmatrix} k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,1}, \tilde{\boldsymbol{X}}_{,n_g}) \\ \vdots & \ddots & \vdots \\ k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,1}) & \cdots & k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,n_g}, \tilde{\boldsymbol{X}}_{,n_g}) \end{pmatrix},$$

where

$$k_{1,\text{ARD}}(\tilde{\boldsymbol{X}}_{,i}, \tilde{\boldsymbol{X}}_{,i'}) = \exp\left\{ -\frac{1}{2}\sum_{d=1}^{2} \frac{(\tilde{X}_{d,i} - \tilde{X}_{d,i'})^2}{l_{1,d}^2} \right\}.$$

The covariance kernel matrix $K_2(\tilde{X}, \tilde{X})$ is similarly defined, with $l_{2,d}$ replacing $l_{1,d}$. Note again that the signal standard deviation term has been dropped from the function $k_{1\text{ARD}}$, and is moved across to the matrix $V_{K,1}$ (and similarly with $k_{2,\text{ARD}}$ and $V_{K,2}$).

- Separate model error standard deviation terms are assumed for each response variable, leading to σ_{ϵ_1} and σ_{ϵ_2}.

The above outlines the EIV MOGP prior distribution and its hyperparameters. After estimating the hyperparameters, the EIV MOGP posterior distribution can be considered for the purpose of jointly estimating the output variables for a given input vector of true values $\tilde{\boldsymbol{X}}_{\text{new}}$. Given joint posterior sample s of the hyperparameters $\boldsymbol{\phi}$, the prediction of the response variables at the input vector $\tilde{\boldsymbol{X}}_{\text{new}}$, $p(\tilde{\boldsymbol{Y}}_{\text{new},s}|\tilde{Y}_s, \tilde{X}_s, \boldsymbol{\phi}_s, \tilde{\boldsymbol{X}}_{\text{new}})$, is bivariate normally distributed with mean $\boldsymbol{m}_s^*$ and covariance matrix V_s^*, i.e.,

$$\tilde{\boldsymbol{Y}}_{\text{new},s}|\tilde{Y}_s, \tilde{X}_s, \boldsymbol{\phi}_s, \tilde{\boldsymbol{X}}_{\text{new}} \sim \text{N}_2(\boldsymbol{m}_s^*, V_s^*) \tag{4.3.5.4}$$

where

$$\boldsymbol{m}_s^* = (\alpha_{1,s}, \alpha_{2,s})' + \boldsymbol{K}_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s)V_{\text{MOGP},\tilde{X},s}^{-1}(\text{vec}(\tilde{Y}_s) - (\boldsymbol{\alpha}_{1,s}, \boldsymbol{\alpha}_{2,s})'), \tag{4.3.5.5}$$

and

$$V_s^* = S_{\text{MOGP},\tilde{\boldsymbol{X}}_{\text{new}},s} - \boldsymbol{K}_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s)V_{\text{MOGP},\tilde{X},s}^{-1}\boldsymbol{K}_s(\tilde{X}_s, \tilde{\boldsymbol{X}}_{\text{new}}). \tag{4.3.5.6}$$

As in previous cases, the subscript s refers to joint posterior sample s from the hyperparameter posterior distribution of $\boldsymbol{\phi}$, and any function which depends on the hyperparameters (or true values) is evaluated at the joint posterior sample s of the relevant parameters. The covariance matrix $V_{\text{MOGP},\tilde{\boldsymbol{X}}_{\text{new}},s}$ is analogous to that of $V_{\text{MOGP},\tilde{X},s}$, replacing the joint posterior sample s of the true values of the input variables with the new true vector $\tilde{\boldsymbol{X}}_{\text{new}}$. The matrix $\boldsymbol{K}_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s)$ is defined to be

$$\boldsymbol{K}_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s) = V_{K,1,s} \otimes K_{1,s}(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s) + V_{K,2,s} \otimes K_{2,s}(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s),$$

and finally, $\boldsymbol{K}_s(\tilde{X}_s, \tilde{\boldsymbol{X}}_{\text{new}}) = \boldsymbol{K}_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s)^T$. The distributions of the predictions of the response variable for each joint posterior sample s are summarised to produce the overall distribution of the predictions of the response variable $p(\tilde{\boldsymbol{Y}}_{\text{new}}|\tilde{Y}, \tilde{X}, \boldsymbol{\phi}, \tilde{\boldsymbol{X}}_{\text{new}})$ by sampling once from the EIV GP posterior for each $s = 1, \ldots, S$, and calculating the mean over those samples and 95% prediction intervals over those samples.

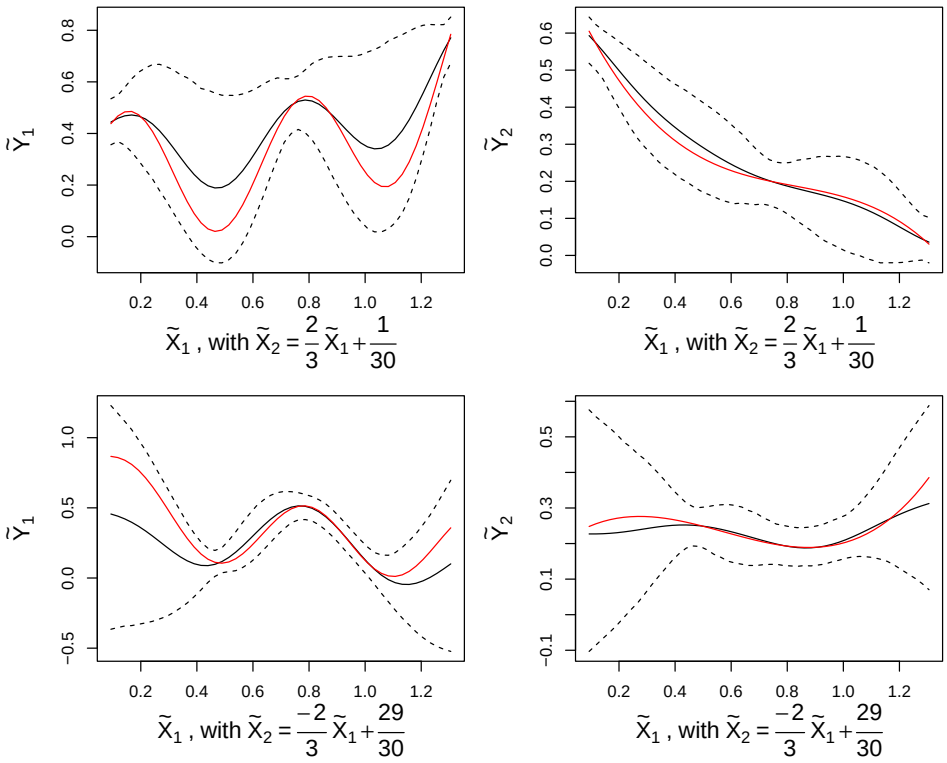

Figure 4.43: Joint predictions of the output variables given the lines $X_2 = \frac{2}{3}X_1 + \frac{1}{30}$ (first row) and $X_2 = -\frac{2}{3}X_1 + \frac{29}{30}$ (second row). The predictions of the first output variable are given in column 1, and for the second output variable in column 2. The solid black lines provides the mean of the predictions, the dotted black lines provide 95% prediction intervals, and the red lines provide the underlying values of the output variables provided by the simulation function.

The prior distributions for the hyperparameters are considered here. Most notably, the reparameterisation that has been considered in previous subsections is not considered here. This is due to the fact that two model error standard deviation parameters now exist, and so the relationship between the distance-scaling parameters $l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2}$ and the standard deviations σ_{ϵ_1} and σ_{ϵ_2} is not as clear. Further investigation into possible relationships could be performed by considering more examples. Instead, each distance-scaling parameter takes the prior distribution Gamma$(2, \frac{2}{0.25})$. The use of the reparameterisation $\sigma_{k_1} = k_{\sigma_{k_1}}\sigma_{\epsilon_1}$ (and similarly for the second output variable) could be justified, but is not considered here; instead, σ_{k_1} and σ_{k_2} take the prior distribution Gamma$(3, \frac{3}{0.4})$. Logically, the means α_1 and α_2 take the weakly informed prior distribution used for α in the single-output case. For the precision parameters τ_{η_1}, τ_{η_2}, T_δ and $T_{\tilde{X}}$, the weakly informed priors that were developed previously are also assumed here. Furthermore, the correlation parameter ρ_{V_K} in the signal covariance matrices $V_{K,1}$ and $V_{K,2}$ takes a uniform prior over the range [-1,1] and the scalars λ_1 and λ_2 of the signal standard deviations σ_{k_1} and σ_{k_2} respectively take uniform priors over the range [0,5].

The simulated data that is used here to fit the EIV MOGP is built using the multi-input EIV GP simulation example covered in Sections 4.3.4 for the first output variable, and a second function to describe the relationship between the second output variable and the two input variables is detailed below. That is, the true values chosen for the matrix $\tilde{X} \in \mathbb{R}^{n_g \times 2}$ and the respective observed data are carried over to this simulation from Section 4.3.4. Furthermore, the true values for the response variables, given by $\tilde{Y}_{1,i}$ and $\tilde{Y}_{2,i}$ for $i = 1, \ldots, n_g$ are simulated from the equations

$$\tilde{Y}_{1,i} = 0.3(80 + 80\sin(\tilde{X}_{1,i}) + \exp(0.55\tilde{X}_{2,i})) + \epsilon_{1,i}, \tag{4.3.5.7}$$

and

$$\tilde{Y}_{2,i} = 3(60 + (\tilde{X}_{1,i} - 10)^2 - 0.9(\tilde{X}_{2,i} - 5)^3) + \epsilon_{2,i}, \tag{4.3.5.8}$$

with the first function taken from Section 4.3.4. Similarly, the observed data for both variables are simulated as in Section 4.3.4, with standard deviation 0.1.

As in Sections 4.3.1 and 4.3.4, there are issues with the posterior density of the model error standard deviation, in this case with σ_{ϵ_1}, for which an appropriate solution is difficult to find. A discussion of the issue in more detail is given in Appendix B.3. An improvement in the model fit was observed with a rescaling of the observed data for first input variable; that is, in previous sections, the observed data that are simulated from the 'chosen'

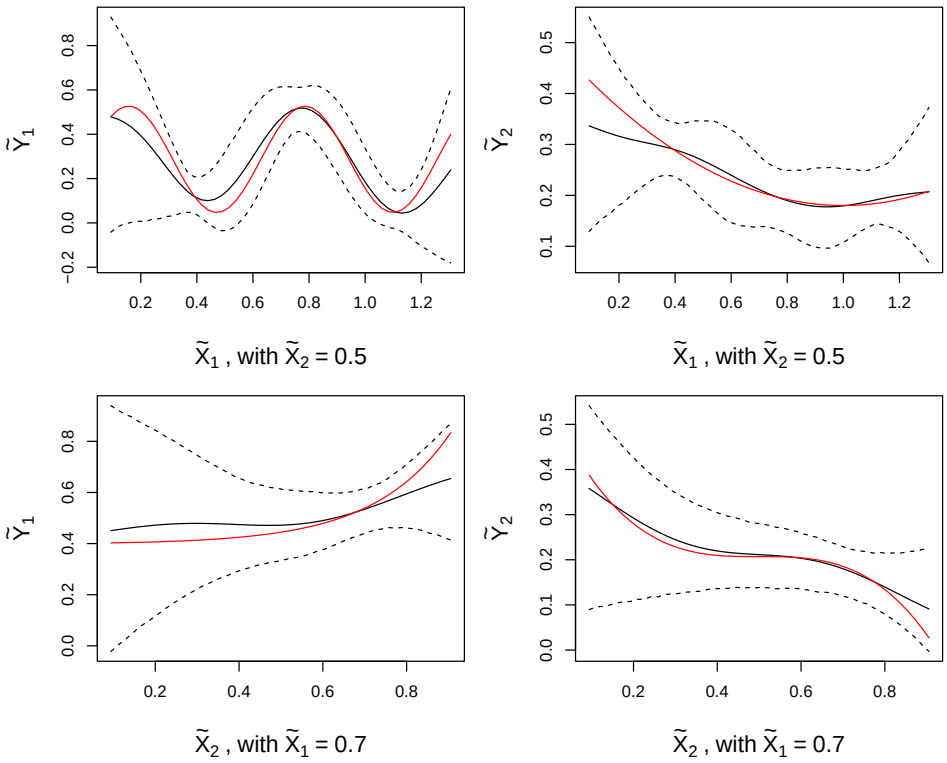

Figure 4.44: Joint predictions of the output variables given the lines $X_2 = 0.5$ (first row) and $X_1 = 0.7$ (second row). The predictions of the first output variable are given in column 1, and for the second output variable in column 2. The solid black lines provides the mean of the predictions, the dotted black lines provide 95% prediction intervals, and the red lines provide the underlying values of the output variables provided by the simulation function.

true values, given by

$$\tilde{\boldsymbol{X}}_1 = (1, 2, \ldots, 13),$$

were divided by 100, so that all the observed data lied in the interval [0,1]. Adjustments had been made in previous sections to prior distributions in the model, in order to accommodate for this awkward scaling. In this case, the adjustment was made to divide the observed data points by 10 instead of 100. Ultimately, this means that the required values of the distance-scaling parameter for the first input variable are now 10 times larger, which leads to samples that are more accessible in the prior distribution Gamma$(2, \frac{2}{0.25})$. Despite being able to overcome the difficulties in previous cases, the issues with this extreme case, where such small values of the distance-scaling parameter are required in the posterior distribution, persist in the EIV MOGP. What this example demonstrates is that the data-scaling process applied in this work is possibly suboptimal for the EIV GP, and a more preferable method should be sought. While the manual override of the data-scaling process was applied here, and provides a suitable estimate of the two functional relationships from the underlying simulation, a data-scaling process which scales the data onto the range [0,1] more appropriately would be beneficial.

A demonstration of the fitted model for the EIV MOGP is provided in Figures 4.43 and 4.44, and shows that the overall modelling process can capably estimate two functional relationships simultaneously. As with the single-output, multi-input EIV GP example from Section 4.3.4, the reason for the variation in the predictions for certain input vectors relative to other input vectors is due to the the relative distance between the 'new' input vector and the existing input vectors. Since this simulation and the simulation from Section 4.3.4 consider the same underlying input vectors (which are both estimated to an almost identical level), the variance in the predictions of the response variables for given input vectors can be related to the plot of the posterior true input values from Figure 4.42, where the dotted lines represent the input vectors that are used to predict the response variables here.

4.4 Powder flow data

This section now considers real-life data, relating to powder flow and powder bed deposition. The data relates to the variables introduced in Chapter 2, with specific references to Section 2.1.2 for tapped density (TD), Section 2.1.3 for angle of repose (AoR), and Section 2.1.4.

A reminder of how the data were collected is provided here. For each variable, data has been collected for seven powders. For each powder, three replicate measurements were taken for all variables except for angle of repose (five replicate measurements), with a different subsample of powder used for each replicate measurement. In the case of three of the explanatory variables, being conditioned bulk density (CBD), specific energy (SE), and basic flowability energy (BFE), which are all variables measured from the FT4 powder rheometer, replicate measurement k for each variable has been measured on the same subsample of powder. There are two response variables and seven explanatory variables in total.

4.4.1 EIV Bayesian regression, uninformed priors

The work carried out in this section focuses on the application of EIV BR models to the real data, with the use of uninformed prior distributions on the parameters in the model. The reason for considering the uninformed prior distributions here is to provide some comparison between the cases in the following sections that consider the informed prior distributions, which will highlight the benefits of considering a more complete version of Bayesian modelling.

With uninformed prior distributions, the posterior distribution of the model parameters is heavily influenced by the available data. In this sense, the posterior distributions of the models in this section provide results in the case where only data is considered as a source of information for the model. It is true in any case that fitting a model to a small sample of data should be approached with care, and conclusions that are drawn from the model inference should highlight the issue of a small sample size.

In a typical statistical modelling process, a visual inspection of the data is carried out initially as an exploration of the data set. This is not advised here, as there is no one-to-one correspondence between the observations of the response variable(s) and the observations of the explanatory variable(s). As an alternative, the simple linear EIV BR model is fitted with each response variable and each explanatory variable in turn, to provide some initial insight into potential relationships between the variables, since the true values for each powder for both the response and explanatory variable can be plotted in terms of 95% credible ellipses over the posterior distribution. Simultaneously, the model coefficients are estimated, and so these models can also be compared using approximate LOO-CV-IC. Figure 4.45 displays seven plots of seven fitted simple linear models, with TD as the response variable in each case, and the seven explanatory variables taken in turn as

the sole explanatory variable in the model.

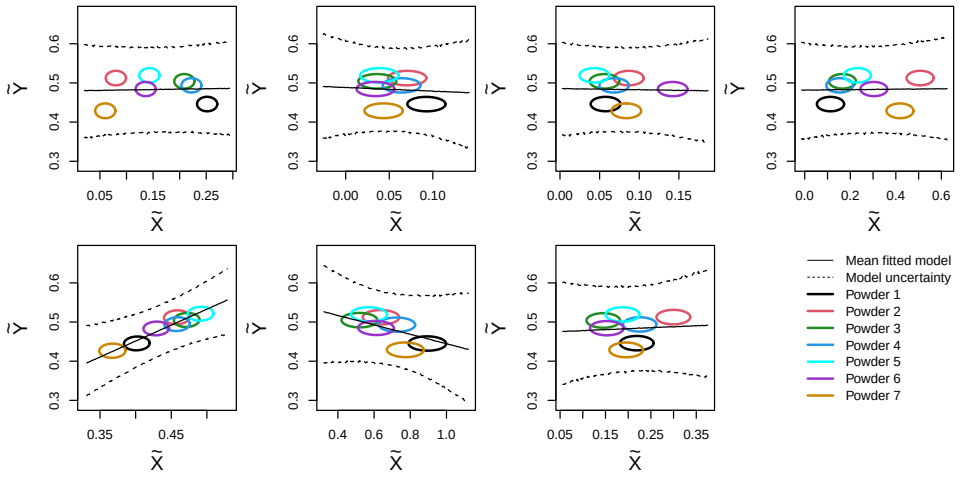

Figure 4.45: EIV BR fitted models with TD as the response variable, and (working row-wise) (1) pressure drop, (2) tapped consolidation, (3) compressibility %, (4) aeration variable, (5) CBD, (6) SE and (7) BFE as the explanatory variable. The ellipses represent 95% credible ellipses for $(\tilde{Y}_i, \tilde{X}_i)$ for each powder, the solid black line represents the posterior mean model fit, and the black dotted lines represent 95% prediction intervals for the model fit. The models here are fitted with uninformed prior distributions.

The seven ellipses in each plot correspond to 95% credible ellipses of the true values $\tilde{Y}_i$ and $\tilde{X}_i$ for each powder $i = 1, \ldots, 7$ (i.e., the marginal joint distributions of each $(\tilde{Y}_i, \tilde{X}_i)$ for each powder i), the solid black line represents the posterior mean model fit (i.e., the fitted values of the response variable within the range of possible true values for the input variable), and the dotted black lines represent 95% prediction intervals for the model fit. The most notable plot is that in the first column, second row, which corresponds to the model with CBD as the explanatory variable. This clearly indicates a strong positive relationship is quite likely between TD and CBD. There is some possibility of a weak negative relationship between TD and SE (second column, second row)—outside of these two plots, it is clear there is not much influence from the explanatory variables on the response variable. This is supported further by Table 4.5, which indicates that the model with the best predictive performance with TD as the response variable is with CBD as the explanatory variable.

The initial investigation of TD from above is also carried out here with

Explanatory variable	PSIS-LOO-CV-IC
Pressure drop	-22.1
Tapped consolidation	-23.1
Compressibility %	-23.1
Aeration variable	-22.0
CBD	-32.9
SE	-26.1
BFE	-22.9

Table 4.5: PSIS-LOO-CV-IC (mean estimates) for the models with TD as the response variable, and each explanatory variable considered in turn in an EIV simple linear model. The best-fitting model, highlighted in green, is given by the model with CBD as the input variable. The models here are fitted with uninformed prior distributions.

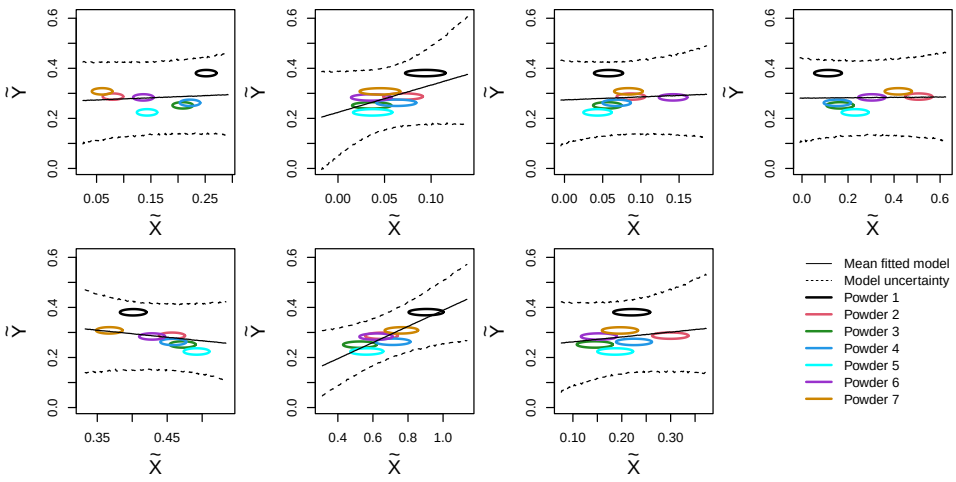

Figure 4.46: The analogous plot to Figure 4.45 with AoR as the response variable.

AoR as the response variable. The analogous plots to those in Figure 4.45 are given for AoR in Figure 4.46, and the analogous table to Table 4.5 is given for AoR in Table 4.6. From the plots in Figure 4.46, there appears to be two explanatory variables that could have some relationship with AoR, being tapped consolidation (row 1, column 2), and SE (row 2, column 2)—with tapped consolidation, there is much more uncertainty in the model fit, relative to the uncertainty with SE as the input variable. While there is significant uncertainty in the prediction intervals, there could possibly be a strong negative relationship between CBD and AoR. There is sufficient evidence from Table 4.6 that the model with SE as the explanatory variable is the best-fitting model of those with a single input variable.

The investigation of the forward modelling with the uninformed prior distributions is taken a step further, by comparing the two best-fitting models for each response variable in turn (i.e., the univariate regression models considered above) with all possible two-covariate models in each case. That is, the model with TD as the response variable is fitted against the possible two-covariate models, where CBD will always be chosen as one of the two covariates. For TD as the response variable, this shows that the inclusion of any further explanatory variables does not improve the predictive performance of the model compared with the model with just CBD, and so a forwards variable selection with the uninformed priors suggests a model with just CBD as the explanatory variable. Similarly, the inclusion of any explanatory variable alongside SE for the model with AoR as the response variable does not improve the predictive performance of the model over the model with just SE as the input variable, and so a forwards variable selection suggests a model with SE as the input variable.

Considering each response variable in turn with univariate regressions provides some information although a more accurate representation of the modelling situation is given by considering the bivariate regression models with TD and AoR as the two response variables. Given the information provided above from the univariate cases, there is a suggestion that the best-fitting model for the bivariate regression would be with CBD and SE as the two inputs variables. This is investigated briefly by considering a backwards variable selection, which is also carried out in the case of fitting with informed prior distributions in Section 4.4.3.

As described in Section 3.4, the backwards variable selection takes the full model to be that with all seven possible explanatory variables included, whose model fit is compared with the seven models that can be fitted where each explanatory variable is removed in turn. The best-fitting model of these eight models, based on the smallest value of approximate LOO-CV-IC, is

Explanatory variable	PSIS-LOO-CV-IC
Pressure drop	-16.7
Tapped consolidation	-19.7
Compressibility %	-16.9
Aeration variable	-16.5
CBD	-18.8
SE	-25.0
BFE	-17.1

Table 4.6: PSIS-LOO-CV-IC (mean estimates) for the models with AoR as the response variable, and each explanatory variable considered in turn in an EIV simple linear model. The best-fitting model, highlighted in green, is given by the model with SE as the input variable. The models here are fitted with uninformed prior distributions.

then selected as the 'full' model—if the full model is the model with all seven explanatory variables, then the backwards variable selection would be suggesting that this is the best-fitting model and the variable selection process is complete. If one of the seven models with six explanatory variables is the best-fitting model, then this model is compared with the six models that are fitted where each of the remaining six explanatory variables is removed in turn. Note that this is not the definitive investigation of the forward modelling analysis in this work—this is used as a template, with the results being compared with those found with the informed prior distributions.

The outcome of the first round of the backwards variable selection is provided in Table 4.7, which compares the first eight models and recommends that removing the explanatory variable 'compressibility %' provides the largest improvement in predictive performance, and so this model is selected as the best-fitting model for the next round of comparisons.

The backwards variable selection continues until there is no improvement observed by removing an explanatory variable. This occurs when considering the three-covariate EIV BR model with tapped consolidation, CBD and SE as the three explanatory variables. This provides some surprise, given what was observed in Tables 4.5 and 4.6 in the univariate regression cases. The approximate LOO-CV-IC value for this best-fitting model is -60.89,

compared with the value of -59.53 with CBD and SE as the two input variables. If this were the definitive case of forward modelling analysis of this data, this would be investigated further to see whether the inclusion of tapped consolidation really provides a worthwhile increase of predictive performance by considering some fitted values plots. This is not carried out here.

Explanatory variables	PSIS-LOO-CV-IC
All seven	-55.7
All but pressure drop	-57.2
All but tapped consolidation	-56.9
All but compressibility %	-57.9
All but aeration variable	-57.2
All but CBD	-49.6
All but SE	-55.0
All but BFE	-56.9

Table 4.7: PSIS-LOO-CV-IC (mean estimates) for the full model with seven explanatory variables and for the seven six-explanatory-variable models, considering uninformed prior distributions. The best-fitting model, highlighted in green, is given by the model with all explanatory variables excluding compressibility %.

On a sidenote, given that the data is in the form of replicate measurements for 7 groups, the relationships estimated here are effectively found based on 7 groups—this requires care and attention when choosing a suitable forward model, due to possibilities of overfitting. This is discussed further in Section 4.4.2.

4.4.2 EIV Gaussian processes, uninformed priors

Having considered the EIV BR model fitted to the real powder data using uninformed prior distributions, the EIV GP models are now fitted to the real powder data using uninformed prior distributions. The main reason for considering some EIV GP model fits here is to be able to compare with the model fits found in Section 4.4.4 to show how including informed prior

information from experts can change the fit of a model.

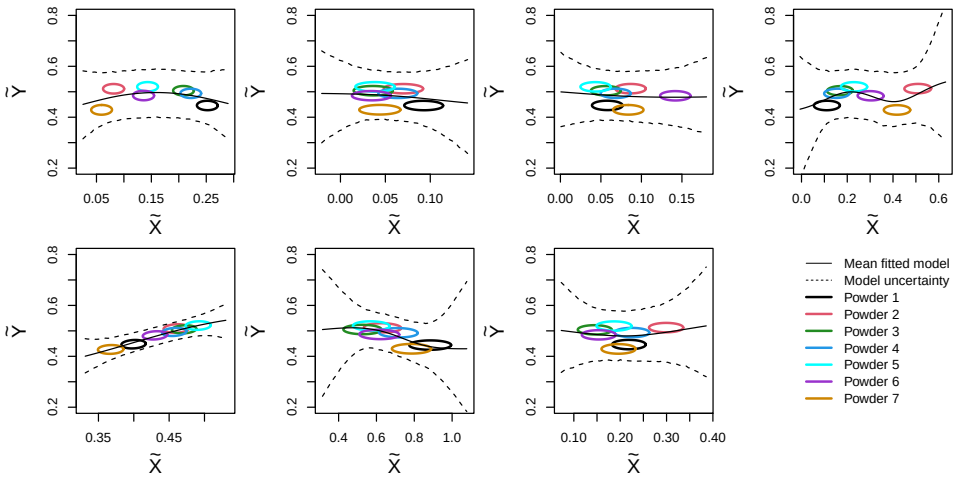

Figure 4.47: Seven fitted models with TD as the response variable, and each input variable taken in turn as the sole input variable. The models are fitted using EIV GPs. These plots are analogous to those in Figure 4.45, with the fits here found using EIV GPs with uninformed prior distributions.

In the previous section, the estimated true values for the variables were found using the EIV BR simple linear model to be able to compare the response variable with each input variable in turn, simultaneously estimating the model fit and carrying out a visual exploration of the data set. This is also carried out here, and will also demonstrate some differences in model fits between the EIV BR and the EIV GP (with uninformed priors in both cases). The plot of single-input single-output EIV GP model fits for the powder flow data with TD as the response variable is provided in Figure 4.47. Similarly to the case with EIV BR, for uninformed prior distributions and TD as the response variable (see plots in Figure 4.45), the model with the input variable as CBD (row 2, column 1) appears to provide the best-fitting relationship with TD, with the prediction intervals being most narrow in this case compared with the other six model fits. The notable difference between the model fits for the EIV GP and the EIV BR (comparing Figure 4.47 with Figure 4.45) is the increased uncertainty in the predictions of the response at the extreme points of each input variable (beyond their maximum and minimum values), which is due to the nature of the nonparametric modelling. The EIV GP model fit with pressure drop as the input variable (row 1, column 1, Figure 4.47) appears to be the only case where the EIV GP fit

Explanatory variable	Mixed LOO-CV-IC
Pressure drop	-25.1
Tapped consolidation	-25.0
Compressibility %	-24.2
Aeration variable	-27.1
CBD	-47.8
SE	-32.1
BFE	-24.2

Table 4.8: Mixed LOO-CV-IC (mean estimates) for the models with TD as the response variable, and each explanatory variable considered in turn in a single-input single-output EIV GP. The best-fitting model, highlighted in green, is given by the model with CBD as the input variable. The models here are fitted with uninformed prior distributions.

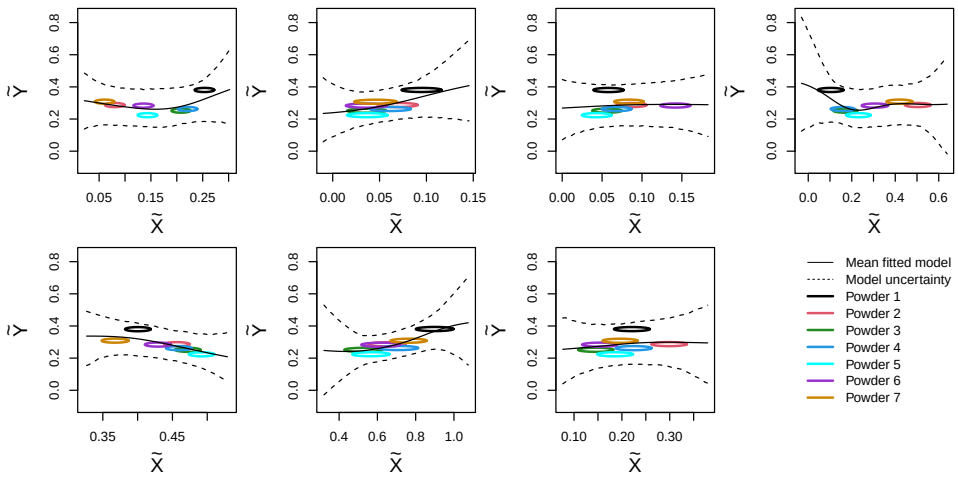

Figure 4.48: Seven fitted models with AoR as the response variable, and each input variable taken in turn as the sole input variable. These plots are analogous to those in Figure 4.46, with the fits here found using EIV GPs with uninformed prior distributions.

looks more appropriate than the EIV BR model fit. This is unsurprising given that a quadratic model fit could be justified between TD and pressure drop, meaning the simple linear model may not be the most appropriate parametric regression in this case.

As in Section 4.4.1, the single-input single-output EIV GP model fits with TD as the response variable and each of the possible input variables taken to be the sole input variable are compared using an approximate form of LOO-CV-IC. For the EIV GP models, this approximate form is the mixed LOO-CV-IC, which is detailed in Section 3.5.4. This entails fitting the EIV GP to provide the posterior distribution for the hyperparameters and true values of the model using the entire data set, then estimating the probability density function $p(\tilde{Y}_i|\boldsymbol{\phi}, \tilde{\boldsymbol{Y}}_{-i}, \tilde{X}_i)$, whose log-likelihood provides information about the out-of-sample predictive performance of the model. The mixed LOO-CV-IC values for each of the seven single-input EIV GP models with TD as the output variable are provided in Table 4.8. The result here that the best-fitting single-input EIV GP model is with CBD as the input variable matches the result found in the EIV BR when comparing EIV simple linear models, and corresponds to what is observed in the model fits in Figure 4.47.

Explanatory variable	Mixed LOO-CV-IC
Pressure drop	-22.1
Tapped consolidation	-24.5
Compressibility %	-19.6
Aeration variable	-24.7
CBD	-24.5
SE	-28.2
BFE	-19.7

Table 4.9: Mixed LOO-CV-IC (mean estimates) for the models with AoR as the response variable, and each explanatory variable considered in turn in a single-input single-output EIV GP. The best-fitting model, highlighted in green, is given by the model with SE as the input variable. The models here are fitted with uninformed prior distributions.

The analogous plot of single-input EIV GP model fits, with AoR as the response variable, is given in Figure 4.48. As with the EIV BR models

with uninformed priors (see plots in Figure 4.46), the majority of the input variables provides model fits with significant uncertainty in the predictions of AoR, with the exception of SE (row 2, column 2).

Furthermore, the single-input EIV GP model fits with AoR as the output variable, and each input variable considered in turn, are compared in Table 4.9, which again supports SE as the input variable that provides the model with the best predictive performance (when considering a single input variable).

4.4.3 EIV Bayesian regression, informed priors

In this section, the powder data is fitted using EIV BR models with *informed* prior distributions to discover which combination of explanatory variables produces the best-fitting model, that is, the model which predicts the response variables best, determined by approximate LOO-CV-IC.

Investigating the modelling using informed prior distributions provides the most complete version of Bayesian modelling. With uninformed priors, the sole source of information provided is the available data. This can be sufficient given a large amount of available data, but this is not necessarily the case, and evidently is not the case with the data considered here. An additional source of information is provided in the form of expert information, which is incorporated into the prior distributions. This expert information is extracted using *elicitation* (see O'Hagan & Oakley (2004), O'Hagan (2005), Garthwaite et al. (2005), O'Hagan (2019)). The details of which parameters are of interest in the elicitation and what expert information has been provided are described below.

Since they exist in both the EIV BR and the EIV GP, the main parameters of interest in the elicitation are the precision parameters that describe the *variability in the measurement error* $\boldsymbol{\delta}_{i,k}$ and $\boldsymbol{\eta}_{i,j}$, and the *variability between materials* for each of the variables, corresponding to the precision parameter $T_{\tilde{X}}$ (or $\tau_{\tilde{X}}$ in the case of one explanatory variable). That is, firstly, for a given powder, the replicate measurements of this powder do not produce the true, unobservable value for that powder (and the given variable), instead they produce values about the true value, with some level of variance. This level of variance is of interest here. Secondly, for a given variable, the true values of each powder exist within some range of values—identifying this range allows for an estimate of the between-materials variability to be extracted. These pieces of information were ascertained by asking for an interval in which 95% of the data lie; for the measurement error, this corresponds to the subsamples varying about the true value, and

for the between-materials variability, this corresponds directly to the range of true values. That is:

1. For a given powder i, of interest is the lower bound l_1 and upper bound u_1 such that 95% of the replicate measurements lie within the range $[l_1, u_1]$ (this is asked about each of the variables in the data). The equivalent probability statement, for example for a given explanatory variable, is $P(l_1 \leq X_{i,k} \leq u_1) = 0.95$.

2. For a given variable, of interest is also the lower bound l_2 and upper bound u_2 such that 95% of the powder true values lie within the range $[l_2, u_2]$. The equivalent probability statement, for a given explanatory variable, is $P(l_2 \leq \tilde{X}_i \leq u_2) = 0.95$.

These questions were posed to Dr Candice Majewski, a supervisor of this PhD research, who provided information that she collated from the literature. This elicited information is provided for the variables tapped density, angle of repose, CBD, SE and BFE in Table 4.10, having scaled the elicited information appropriately (i.e., using the same scaling as for the observed data in the model). Information about the remaining variables was not provided. In one case, for the explanatory variable BFE, the elicited information for the range of true values for the material was 700mJ to 1500mJ. This was in large contrast to the observed data for this variable, whose range (over all powders) is given by 141.5mJ to 322.4mJ, to 1 decimal place. The conclusion was made to set the range of the true values in the modelling to $0.9 \times 141.532 = 127.3788$mJ as the lower bound and $1.1 \times 1500mJ = 1650$mJ as the upper bound, to compensate for the stark contrast between the elicitation and the observed data. For the lower bound, 10% of the value is subtracted from the minimum of the observed data to incorporate further uncertainty, and similarly for the upper bound. Due to the upper bound of the range of true values, the observed data is scaled by dividing by 10000, so that the maximum true value is within the range [0,1]. In conclusion, the initial probability statement of $P(0.07 \leq \tilde{X}_i \leq 0.15) = 0.95$ is adjusted to $P(0.0127 \leq \tilde{X}_i \leq 0.165) = 0.95$, which is reflected in Table 4.10.

Having established the probability statements that correspond to the provided expert information, these now must be converted into informed prior distributions. An established process for carrying this out is provided in Section 4.2.5 under the 'Weakly informative priors for simulated data' heading. The application of this process for the informed prior distributions in the real data is described in detail in Appendix C.1. Those established prior distributions are summarised here in Tables 4.11 and Table 4.12.

	Powder sub-sample variation	Between-materials variation	Elicited probability statements
TD	$l_1 = 0.99\tilde{Y}_{1,i}$, $u_1 = 1.01\tilde{Y}_{1,i}$	$l_2 = 0.494$, $u_2 = 0.725$	$P(0.99\tilde{Y}_{1,i} \leq Y_{1,i,j_1} \leq 1.01\tilde{Y}_{1,i})$=0.95, $P(0.494 \leq \tilde{Y}_{1,i} \leq 0.725) = 0.95$
AoR	$l_1 = 0.95\tilde{Y}_{2,i}$, $u_1 = 1.05\tilde{Y}_{2,i}$	$l_2 = 0.25$, $u_2 = 0.528$	$P(0.95\tilde{Y}_{2,i} \leq Y_{2,i,j_2} \leq 1.05\tilde{Y}_{2,i})$=0.95, $P(0.494 \leq \tilde{Y}_{2,i} \leq 0.725) = 0.95$
CBD	$l_1 = 0.99\tilde{X}_i$, $u_1 = 1.01\tilde{X}_i$	$l_2 = 0.35$, $u_2 = 0.55$	$P(0.99\tilde{X}_i \leq X_{i,k} \leq 1.01\tilde{X}_i)$=0.95, $P(0.35 \leq \tilde{X}_i \leq 0.55) = 0.95$
SE	$l_1 = 0.95\tilde{X}_i$, $u_1 = 1.05\tilde{X}_i$	$l_2 = 0.5$, $u_2 = 0.8$	$P(0.95\tilde{X}_i \leq X_{i,k} \leq 1.05\tilde{X}_i)$=0.95, $P(0.5 \leq \tilde{X}_i \leq 0.8) = 0.95$
BFE	$l_1 = 0.915\tilde{X}_i$, $u_1 = 1.085\tilde{X}_i$	$l_2 = 0.0127$, $u_2 = 0.165$	$P(0.915\tilde{X}_i \leq X_{i,k} \leq 1.085\tilde{X}_i)$=0.95, $P(0.0127 \leq \tilde{X}_i \leq 0.165) = 0.95$

Table 4.10: Elicited information (with the discussed adjustments to BFE) and equivalent probability statements for some precision parameters in the forward modelling.

	Measurement error shape and rate parameters	Between-materials or model error shape and rate parameters	Diagonal elements for Wishart scale matrix
TD	$a_{\eta_1} = 0.6$, $b_{\eta_1} = \frac{0.6}{103410.5}$	$a_{\epsilon_1} = 0.2$, $b_{\epsilon_1} = 2 \times 10^{-6}$	$S_{\epsilon_{1,1}} = 10^5$
AoR	$a_{\eta_2} = 0.6$, $b_{\eta_2} = \frac{0.6}{10154.84}$	$a_{\epsilon_2} = 0.2$, $b_{\epsilon_2} = 3 \times 10^{-6}$	$S_{\epsilon_{2,2}} = \frac{2\times 10^5}{3}$
CBD	$a_\delta = 0.696$, $b_\delta = \frac{0.696}{189708.7}$	$a_{\tilde{X}} = 0.696$, $b_{\tilde{X}} = \frac{0.696}{384.16}$	$S_{\delta_{\text{CBD}}} = 189708.7$, $S_{\tilde{X}_{\text{CBD}}} = 384.16$
SE	$a_\delta = 1.73075$, $b_\delta = \frac{1.73075}{3637.019}$	$a_{\tilde{X}} = 1.73075$, $b_{\tilde{X}} = \frac{1.73075}{170.7378}$	$S_{\delta_{\text{SE}}} = 3637.019$, $S_{\tilde{X}_{\text{SE}}} = 170.7378$
BFE	$a_\delta = 0.993$, $b_\delta = \frac{0.993}{67324.68}$	$a_{\tilde{X}} = 0.993$, $b_{\tilde{X}} = \frac{0.993}{662.8092}$	$S_{\delta_{\text{BFE}}} = 67324.68$, $S_{\tilde{X}_{\text{BFE}}} = 662.8092$
X_{generic}	$a_\delta = 1$, $b_\delta = 0.000657$	$a_{\tilde{X}} = 1$, $b_{\tilde{X}} = \frac{1}{15.3664}$	$S_{\delta_{\text{generic}}} = \frac{1}{0.000657}$, $S_{\tilde{X}_{\text{generic}}} = 15.3664$

Table 4.11: Hyperparameters for informed gamma prior distributions on precision parameters for each of the variables that underwent elicitation.

	Prior with TD as response	Prior with AoR as response
β_0	$\mathrm{N}(0.6095, 0.0001)$	$\mathrm{N}(0.389, 0.0001)$
β_{CBD}	$\mathrm{N}(0, 2.028)$	$\mathrm{N}(0, 1.401)$
β_{SE}	$\mathrm{N}(0, 4.564)$	$\mathrm{N}(0, 3.151)$
β_{BFE}	$\mathrm{N}(0, 1.176)$	$\mathrm{N}(0, 0.812)$
β_{generic}	$\mathrm{N}(0, 1.171)$	$\mathrm{N}(0, 0.809)$

Table 4.12: Informed prior distributions for the model coefficients for CBD, SE and BFE (depending on the response variable), with weakly informed prior distributions for the model coefficients for the remaining explanatory variables.

For the simulated-data cases, initialisation of the prior distributions was not carried out. This refers to the choosing of overdispersed values, with respect to the posterior distribution, as starting points for each parallel chain, to check that the MCMC output converges to the same posterior distribution from a range of starting points. This is carried out for the model fits to real data, with the process discussed in Appendix C.2.

To demonstrate the effect of informed priors on the fit of EIV BR models, the example of the simple linear model using TD as the response variable and CBD as the input variable is fitted using informed prior distributions, and the model fit is visually compared with the model fit using uninformed priors identified in Section 4.4.1. The two model fits are presented in Figure 4.49, with the uninformed model fit on the left, and the informed model fit on the right. The use of expert information in the prior distributions results in increased certainty in the true values for each powder, with the 95% credible ellipses much smaller in the right plot than in the left plot. Furthermore, the uncertainty in the predictions of the response variable has been reduced as a result of there being less uncertainty in the true values and providing more information in the prior distributions of the model parameters β_0, β_1 and τ_ϵ.

The forward modelling is now investigated using a backwards variable selection, as outlined in Section 3.4. That is, the full model including all

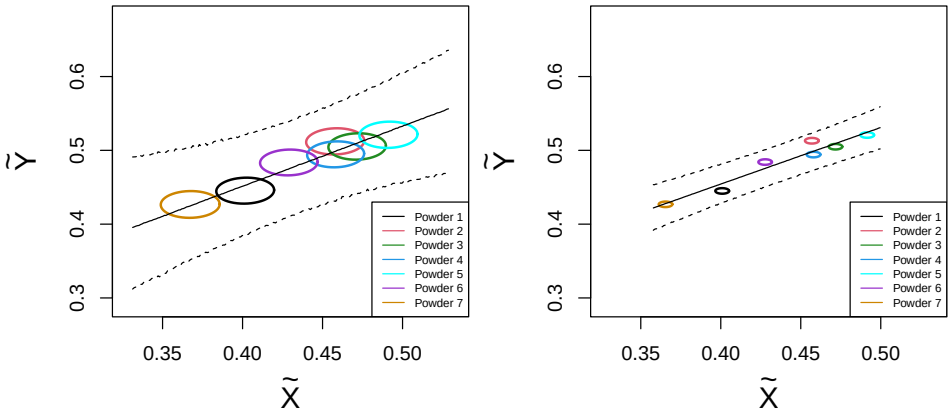

Figure 4.49: Two plots of EIV BR models fitted with TD as the response variable and CBD as the input variable, with the plot on the left corresponding to fitting with uninformed prior distributions, and on the right with informed prior distributions.

explanatory variables is the starting model, given by

$$\begin{pmatrix} \tilde{Y}_{1,i} \\ \tilde{Y}_{2,i} \end{pmatrix} = \begin{pmatrix} \beta_{0,1} & \beta_{1,1} & \beta_{2,1} & \beta_{3,1} & \beta_{4,1} & \beta_{5,1} & \beta_{6,1} & \beta_{7,1} \\ \beta_{0,2} & \beta_{1,2} & \beta_{2,2} & \beta_{3,2} & \beta_{4,2} & \beta_{5,2} & \beta_{6,2} & \beta_{7,2} \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{X}_{1,i} \\ \tilde{X}_{2,i} \\ \tilde{X}_{3,i} \\ \tilde{X}_{4,i} \\ \tilde{X}_{5,i} \\ \tilde{X}_{6,i} \\ \tilde{X}_{7,i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{pmatrix},$$

with the true values of the response variables $\tilde{Y}_{1,i}$, $\tilde{Y}_{2,i}$ corresponding to the true values of TD and AoR respectively, and the true values of the explanatory variables, given by $\tilde{X}_{1,i}$, $\tilde{X}_{2,i}$, $\tilde{X}_{3,i}$, $\tilde{X}_{4,i}$, $\tilde{X}_{5,i}$, $\tilde{X}_{6,i}$ and $\tilde{X}_{7,i}$, corresponding to those of the variables pressure drop, tapped consolidation, compressibility %, the log of total energy at 2mm/s on the 4mm/s aeration test, CBD, SE and BFE respectively. Note again that the subscript i varies from 1 to $n_g = 7$. The observed data for the two response variables are of the form

$$Y_{1,i,j_1} = \tilde{Y}_{1,i} + \eta_{1,i,j_1},$$

$$Y_{2,i,j_2} = \tilde{Y}_{2,i} + \eta_{2,i,j_2},$$

with $\eta_{1,i,j_1} \sim \mathrm{N}(0, \tau_{\eta_1})$ and $\eta_{2,i,j_2} \sim \mathrm{N}(0, \tau_{\eta_2})$, for $j_1 = 1, 2, 3$ and $j_2 = 1, \ldots, 5$. Furthermore, the observed data for the seven explanatory variables are of the form

$$\begin{aligned} X_{1,i,k_1} &= \tilde{X}_{1,i} + \delta_{1,i,k_1}, \\ X_{2,i,k_2} &= \tilde{X}_{2,i} + \delta_{2,i,k_2}, \\ X_{3,i,k_3} &= \tilde{X}_{3,i} + \delta_{3,i,k_3}, \\ X_{4,i,k_4} &= \tilde{X}_{4,i} + \delta_{4,i,k_4}, \\ \begin{pmatrix} X_{5,i,k} \\ X_{6,i,k} \\ X_{7,i,k} \end{pmatrix} &= \begin{pmatrix} \tilde{X}_{5,i} \\ \tilde{X}_{6,i} \\ \tilde{X}_{7,i} \end{pmatrix} + \begin{pmatrix} \delta_{5,i,k} \\ \delta_{6,i,k} \\ \delta_{7,i,k} \end{pmatrix}, \end{aligned}$$

with $\delta_{1,i,k_1} \sim \mathrm{N}(0, \tau_{\delta_1})$, $\delta_{2,i,k_2} \sim \mathrm{N}(0, \tau_{\delta_2})$, $\delta_{3,i,k_3} \sim \mathrm{N}(0, \tau_{\delta_3})$ and $\delta_{4,i,k_4} \sim \mathrm{N}(0, \tau_{\delta_4})$, for $k_1, k_2, k_3, k_4 = 1, 2, 3$ and

$$\begin{pmatrix} \delta_{5,i,k} \\ \delta_{6,i,k} \\ \delta_{7,i,k} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, T_\delta \right),$$

for $k = 1, 2, 3$.

The model is fitted with the informed prior distributions noted in Tables 4.11 and 4.12; for the clarity of the precision matrices T_ϵ, $T_{\tilde{X}}$ and T_δ, these each take the Wishart distributions, with

$$\begin{aligned} T_\epsilon &\sim \text{Wishart}(S_\epsilon, 2), \\ T_{\tilde{X}} &\sim \text{Wishart}(S_{\tilde{X}}, 7), \\ T_\delta &\sim \text{Wishart}(S_\delta, 3), \end{aligned}$$

where the scale matrix for the model error precision matrix Wishart prior is given by

$$S_\epsilon = \begin{pmatrix} 10^5 & 0 \\ 0 & \frac{2}{3} \times 10^5 \end{pmatrix},$$

the scale matrix for the between-materials precision matrix Wishart prior is given by

$$S_{\tilde{X}} = \begin{pmatrix} 15.3664 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 15.3664 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 15.3664 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 15.3664 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 384.16 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 170.7378 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 662.8092 \end{pmatrix},$$

and the scale matrix for the (explanatory variable) measurement error precision matrix Wishart prior is given by

$$S_\delta = \begin{pmatrix} 189708.7 & 0 & 0 \\ 0 & 3637.019 & 0 \\ 0 & 0 & 67324.68 \end{pmatrix}.$$

With the initialisation implemented as described in Appendix C.2, the remaining MCMC tuning parameters for the first attempt at fitting the model are given by an adaptation phase of length 1000 samples, a burn-in phase of length 25000 samples, 200000 posterior draws sampled from the MCMC output and storing every 10^{th} sample to give 20000 posterior samples. This leads to poor levels of mixing and convergence in the MCMC output, so the tuning parameters are adjusted until convergence is found and mixing is sufficient. This occurs for an adaptation phase of length 5000 samples, a burn-in phase of length 1.75×10^6, 6.25×10^6 posterior draws sampled from the output, taking every 25^{th} sample of these to give 250000 posterior samples. Appropriate model output is checked to ensure that the model is working as intended, for example checking the posteriors of true values for some of the response and explanatory variables, as well as the fitted values of the response variables against the marginal posteriors of the true values of the responses. These plots are omitted here and can be found in Appendix D.2.

The backwards variable selection continues with the fitting of seven models that are nested within the full model discussed above. Each nested model corresponds to removing one of the seven explanatory variables from the full model, whose model fit is then compared with the model fits of the seven nested models. Whichever of the seven nested model fits provides the largest improvement in predictive performance according to approximate LOO-CV-IC will be selected as the 'new' full model, to be compared with its nested models.

The seven nested models of the full model are fitted, and the approximate LOO-CV-IC for each (as well as for the full model) is provided in Table 4.13. The approximate LOO-CV-IC indicates that the best-fitting of the eight models, including the full model, is the model with six explanatory variables, i.e. all explanatory variables excluding compressibility %. This appears to be a narrow improvement on the predictive performance of the model relative to the full model with all seven explanatory variables. There are also two other models, one with all explanatory variables excluding pressure drop and one with all explanatory variables excluding tapped consolidation, which provide a marginal increase in predictive performance over the full model,

Explanatory variables	PSIS-LOO-CV-IC
All seven	-86.9
All but pressure drop	-87.4
All but tapped consolidation	-87.2
All but compressibility %	-87.7
All but aeration variable	-86.2
All but CBD	-76.4
All but SE	-86.7
All but BFE	-85.1

Table 4.13: PSIS-LOO-CV-IC (mean estimates) for the full model with seven explanatory variable and for the seven six-explanatory-variable models. The best-fitting model, highlighted in green, is given by the model with all explanatory variables excluding compressibility %.

and are marginally worse in predictive performance than the best-fitting model. Another notable outcome of this first round of model comparison using backwards variable selection is the decrease in predictive performance of the model with the variable CBD excluded. This suggests that there is a relatively high probability of the best-fitting model including CBD.

Based on Table 4.13, the conclusion from the first round of backwards variable selection is that the best-fitting model so far is with all explanatory variables excluding compressibility %. This model is now chosen to be the 'full' model, which is to be compared with six nested models, each of which have one of the six remaining explanatory variables removed. These models are fitted (with MCMC tuning parameters that provide convergence of the MCMC output to the posterior distribution and sufficient levels of mixing within each chain), and the approximate LOO-CV-IC estimates are compared in Table 4.14. The conclusion from the second round of backwards variable selection is that the best-fitting model is the full model, with all explanatory variables excluding compressibility %. This concludes the backwards variable selection, as no improvement in predictive performance is found by removing any of the variables from the full model, and so the backwards variable selection recommends to model with all explanatory variables except for compressibility %.

Explanatory variables	PSIS-LOO-CV-IC
All but compressibility	-87.7
All but compressibility % and pressure drop	-84.9
All but compressibility % and tapped consolidation	-87.3
All but compressibility % and aeration variable	-83.3
All but compressibility % and CBD	-79.0
All but compressibility % and SE	-84.3
All but compressibility % and BFE	-86.8

Table 4.14: PSIS-LOO-CV-IC (mean estimates) for the full model with all explanatory variables excluding compressibility % and for the six five-explanatory-variable models. The best-fitting model, highlighted in green, is given by the model with all explanatory variables excluding compressibility %.

This is a surprising outcome. Given the modelling based on the uninformed priors, there was a strong indication that CBD and SE would be the best-fitting model again. The reality here is that there is a major concern with recommending that a six-explanatory-variable model is the best-fitting model, given the model fit is based on seven true values for each variable. It is quite likely that this is an issue of *overfitting*, which is described here. It is known that, including more terms in a linear model, whether it be from other explanatory variables or from including polynomial terms or interaction terms from the existing explanatory variables, will improve the predictive performance of the model. In principle, this sounds like a good idea, to try to improve the model fit as much as possible; in practice, particularly with little data available, the model can become overfitted to the available data. This means that the overfitted model could then become inappropriate if new data were to become available. Consider again the example from Chapter 1, which compared classical estimates in backward modelling. The main focus here is on the observed data points and overfitting. Suppose a small subset of observed data points are taken from this example, and two models are fitted: a simple linear model, and a complicated polynomial model. The left plot in Figure 4.50 provides the fitted

lines for both the simple linear model (solid black line) and the complicated polynomial model (dotted black line)—it is clear from the plot that the complicated polynomial model provides the model fit that minimises the model error for the available data. This data is simulated from a straight line equation (with some error), and simulating more data points from this function provides the plot on the right, with the red points representing the new data points. It is clear that the straight line model fit from the already existing data points is much more suitable for the entire data set than the model fit from the complicated polynomial model. This is a clear case of overfitting, where there is not enough data to suggest that more complicated model fits are appropriate—the model becomes too heavily adjusted to the existing data.

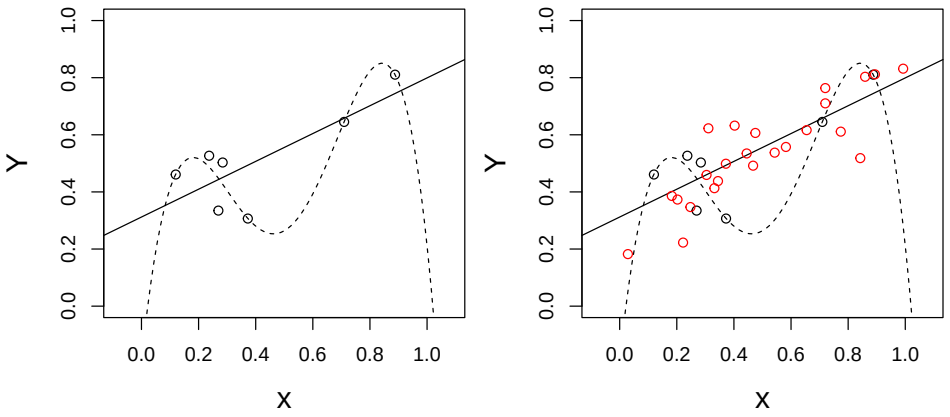

Figure 4.50: Two plots demonstrating the model fit from two models that are fitted to the black data points (the existing data from an underlying simulation). The left plot suggests that the black dotted line, the model fit of a complicated polynomial model, better minimises the model error than the simple linear model. The right plot demonstrates that having more information from the underlying simulation definitively shows that the complicated polynomial model is 'overfitted', and does not represent the underlying simulation, whereas the simple linaer model looks appropriate.

Reverting to the real data and the above model fits, there is a high probability that the 'best-fitting' six-explanatory-variable model is overfitted to the seven available powders. Given the lack of available data, it would be preferable to choose a simpler model, even if this is not suggested by the approximate LOO-CV-IC. It is also possible that the adjustment for effective number of parameters in approximate LOO-CV-IC is not strong

enough. For this reason, an alternative model is desired, which provides a better chance of being appropriate for a larger data set.

The above evidence suggests that the explanatory variable CBD is highly likely to be included in a 'best-fitting' model; from both Table 4.13 and Table 4.14, there is a clear drop-off in predictive performance with CBD excluded from the model. Based on this, it is assumed here that best-fitting model will include CBD as one of its explanatory variables. In order to avoid overfitting, the best-fitting model will be selected from six two-covariate models as well as the one-input model with CBD as the explanatory variable. In effect, a quasi-forwards variable selection is carried out, with the assumption that CBD is the first explanatory variable, and with the restriction of considering models with two explanatory variables. As above, these models are fitted with MCMC tuning parameters that produce convergence and suitable levels of mixing, and their predictive performances are compared with approximate LOO-CV-IC in Table 4.15. The approximate LOO-CV-IC suggests that CBD and SE as the two explanatory variables provides the model with the best predictive performance, which is closely followed by the combination of CBD and tapped consolidation. Comparing with the model with just CBD, there is clearly an improvement in the model when including SE. The model with CBD and BFE appears to be worse than the model with just CBD, whereas the other combinations of CBD and an explanatory variable appear to provide some improvement. Based on the discussion of overfitting above, it is concluded here that, given the informed prior distributions, the best-fitting model is that with CBD and SE as the two explanatory variables.

At this point, it is noted that, given the approximate method of LOO-CV-IC 'PSIS-LOO-CV-IC' (Pareto-smoothed importance sampling leave-one-out cross-validation information criterion) is applied here for measuring predictive performance, diagnostics of the approximate method are considered. The Pareto smoothing is applied to the importance ratios for estimating the LOO log-likelihoods, for which the generalised Pareto distribution is required, and its shape parameter k is estimated. The LOO likelihoods are found for each $i = 1, \ldots, n_g$ (as each group i is 'left-out' and used as a test group), and so estimates of the shape parameter k are given for each group. The minimum estimate is given by 0.861, and the maximum is given by 1.033. The paper Vehtari et al. (2015) notes that, for shape parameter estimates larger than 0.7, the required size of the posterior samples for convergence of the Pareto smoothing sampler grows 'infeasibly large', with values of 0.8 roughly requiring 10^{10} posterior samples. Without convergence of the sampler, the importance weights are unreliable, and so the PSIS-LOO-CV-IC may not be a good representation of true predictive performance. While

Explanatory variables	PSIS-LOO-CV-IC
CBD	-53.6
CBD + pressure drop	-60.9
CBD + tapped consolidation	-76.9
CBD + compressibility %	-54.9
CBD + aeration variable	-63.2
CBD + SE	-77.8
CBD + BFE	-51.4

Table 4.15: PSIS-LOO-CV-IC (mean estimates) for the model with just CBD as the explanatory variable and for the six two-explanatory-variable models for each combination of CBD and an explanatory variable. The best-fitting model, highlighted in green, is given by the model with CBD and SE.

the relative values of PSIS-LOO-CV-IC in this case may be unreliable, there is reasonable evidence from the fitted models given in Figures 4.45 and 4.46 that CBD predicts TD well, and SE predicts AoR well (albeit with an uninformed prior specification). This work will continue using PSIS-LOO-CV-IC as one of the measures of predictive performance of EIV BR models, and so it is assumed that the two-covariate model with CBD and SE as the input variables and an additive linear predictor is the best-fitting model. Suggestions on how to navigate this issue with PSIS-LOO-CV-IC are provided in Chapter 6.

A visual demonstration of the predictive performance of the model is provided by the fitted values plot of each response variable given in Figure 4.51 for TD and in Figure 4.52 for AoR. Comparing these fitted plots with those of the full model with all explanatory variables, given in Appendix D (Figure D.4), it is noted that there is a drop-off in the level of predictive performance, which is a sacrifice that had to be made given the high probabilities of overfitting the model to the data with too many explanatory variables. Interestingly, the uncertainty in the fitted values is much more unique than the uncertainty in the model with all explanatory variables. Consider firstly the fitted values plot for tapped density from the first plot of Figure D.4, which provides a consistent level of variation in the fitted val-

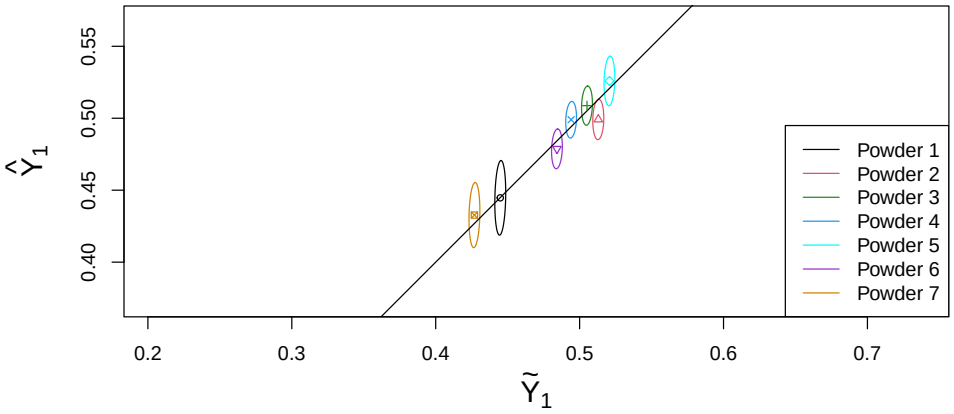

Figure 4.51: Joint 95% credible ellipses for the joint distribution of fitted values and true values of the response variable TD. Output is based on the fitted model with CBD and SE as the explanatory variables.

ues for each of the powders. Comparatively, the fitted values for TD from Figure 4.51 contain different levels of variation in the fitted values for some of the powders—there appears to be a clear split, with Powders 1 and 7 having larger variation in the fitted values than those of Powders 2 through to 6. It also appears that the variation for the fitted values of tapped density for Powders 2 through to 6, in the case of the model with just CBD and SE, is smaller than that for the fitted values of for Powders 2 through to 6 in the model with all explanatory variables. This is explained by the fact that the two-explanatory-variable model has more information about the relationship between TD and CBD and SE around Powders 2 to 6 relative to Powders 1 and 7, whereas this relative difference is less pronounced in the model with more explanatory variables. The level of uncertainty is larger around these powders in the model with more explanatory variables as there is simply more possible contributions to the uncertainty, given there are more explanatory variables in the model.

The posterior densities of the model intercepts and the slope terms for each response variable are provided in Figure 4.53 for the model with CBD and SE as the explanatory variables. This provides some breakdown as to how influential each explanatory variable is in the predictions of the response variables. It is somewhat surprising to see a lack of smoothness in the estimated posterior densities of some of the model coefficients. It is clear that there is a positive relationship between TD and CBD (first row, middle plot), with all the posterior density at values greater than 0. There

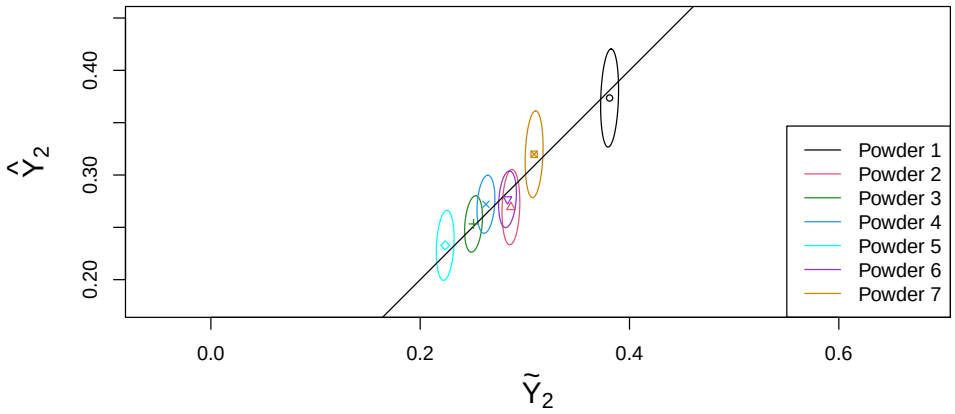

Figure 4.52: Joint 95% credible ellipses for the joint distribution of fitted values and true values of the response variable AoR. Output is based on the fitted model with CBD and SE as the explanatory variables.

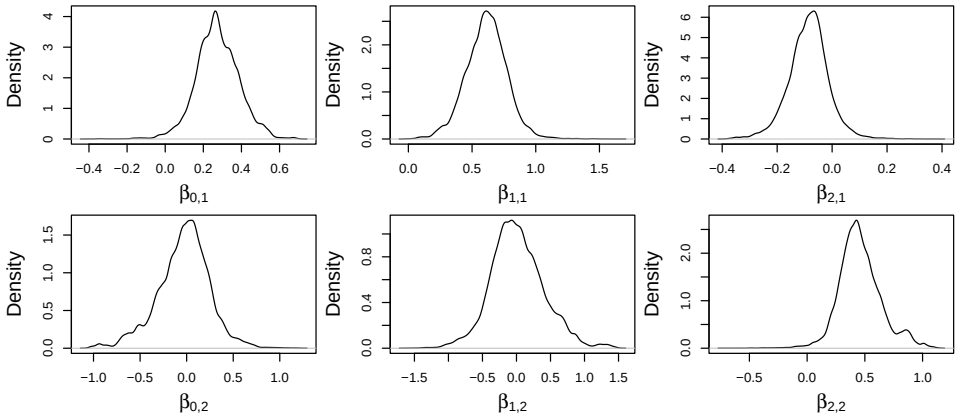

Figure 4.53: Posterior densities of the model coefficients for the model fitted with the two explanatory variables CBD and SE. The top row of plots corresponds to the model coefficients for TD, and the bottom row for AoR. The first column of plots represents the intercept terms, the second column the slope terms for CBD, and the third column the slope terms for SE.

is also a relatively small negative contribution from SE to the predictions of TD. For AoR, it appears that there is more uncertainty in the posterior densities—the mean of CBD looks to be close to 0, but with a large range of values relative to the other posterior densities. In the case of AoR, the contribution of SE looks to be more significant, with a positive relationship similar to that between TD and CBD.

The EIV BR forward modelling with informed prior distributions has determined that the best-fitting model, with manual input regarding the threat of overfitting, is the model with the two explanatory variables CBD and SE. These two explanatory variables are investigated further in the following section by fitting the EIV MOGP with TD and AoR as the two output variables and CBD and SE as the two input variables, whose model fit will contribute to the comparison of the two methods at the end of this chapter.

4.4.4 EIV Gaussian processes, informed priors

In this section, the EIV MOGP is fitted with the real data, with TD and AoR as the output variables and CBD and SE as the input variables, for the purposes of comparing the model fits of the two statistical models considered in this work. The process of variable selection is not carried out here. The model fit here is estimated assuming the informed prior distributions described in the previous section. That is, for the between-materials precision and measurement error precision $T_{\tilde{X}}$ and T_{δ} respectively, their respective informed prior distributions are given by

$$T_{\tilde{X}} \sim \text{Wishart}(S_{\tilde{X}}, 2)$$

and

$$T_{\delta} \sim \text{Wishart}(S_{\delta}, 2),$$

where

$$S_{\tilde{X}} = \begin{pmatrix} 384.16 & 0 \\ 0 & 170.7378 \end{pmatrix}$$

and

$$S_{\delta} = \begin{pmatrix} 189708.7 & 0 \\ 0 & 3637.019 \end{pmatrix}.$$

Moreover, the measurement error precision parameters for TD and AoR take the prior distributions

$$\tau_{\eta_1} \sim \text{Gamma}\left(0.6, \frac{0.6}{103410.5}\right)$$

and

$$\tau_{\eta_2} \sim \text{Gamma}\left(0.6, \frac{0.6}{10154.84}\right)$$

respectively. The model error standard deviations are treated marginally in the EIV MOGP. The model error prior information has been formulated in terms of a precision prior, which is also used here, and is then transformed into a standard deviation parameter in the MCMC. That is, the model error standard deviation terms $\sigma^2_{\epsilon_1}$, $\sigma^2_{\epsilon_2}$ are given by

$$\sigma^2_{\epsilon_1} = \sqrt{\frac{1}{\tau_{\epsilon_1}}}$$

and

$$\sigma^2_{\epsilon_2} = \sqrt{\frac{1}{\tau_{\epsilon_2}}},$$

where

$$\tau_{\epsilon_1} \sim \text{Gamma}(0.2, 2 \times 10^{-6})$$

and

$$\tau_{\epsilon_2} \sim \text{Gamma}(0.2, 3 \times 10^{-6}).$$

The EIV MOGP hyperparameters are considered further here. The prior distributions of the correlation parameter ρ_{V_K}, and the scalars λ_1, λ_2 remain unchanged; that is, $\rho_{S_K} \sim \text{U}(-1, 1)$, and $\lambda_1, \lambda_2 \sim \text{U}(0, 5)$. The hyperprior distributions for the distance-scaling parameters $\boldsymbol{l}$ (or in the single-input, single-output case, l), the signal standard deviations $\boldsymbol{\sigma}_k$ (or in the single-output case, σ_k), and GP prior means α_1 and α_2 are adjusted given the elicitation.

The distance-scaling parameters are adjusted in the cases of CBD, SE and BFE, where the explanatory variable true value ranges are provided in the elicitation. These ranges influence the distance-scaling parameter as the value of the covariance kernel function is influenced by both the distance-scaling parameter and the range of the input variable—see the discussion of this in Section 3.5.1. The informed prior distribution for the distance-scaling parameter corresponding to each of these input variables is chosen using the following method:

1. Values of the squared exponential covariance kernel, given by

$$k(\tilde{X}, \tilde{X}') = \exp\left\{-\frac{(\tilde{X} - \tilde{X}')^2}{2l^2}\right\}$$

are produced for any true values $\tilde{X}, \tilde{X}'$ in the elicited range for the input variable, and for different values of l.

2. The value of l is identified for which the relationship between k and $\tilde{X} - \tilde{X}'$ is as close to a straight line as possible, with k decreasing as $\tilde{X} - \tilde{X}'$ increases; this results in, a priori, assuming that each distance between two values for the input variable produces a different value of k. This is visualised in Figure 4.54.

3. Take the value of l identified above to be the median value of the gamma prior distribution for l, assuming a shape of 2—this provides an estimate of the rate parameter (in other words, the rate parameter is adjusted until the value of l is the prior median).

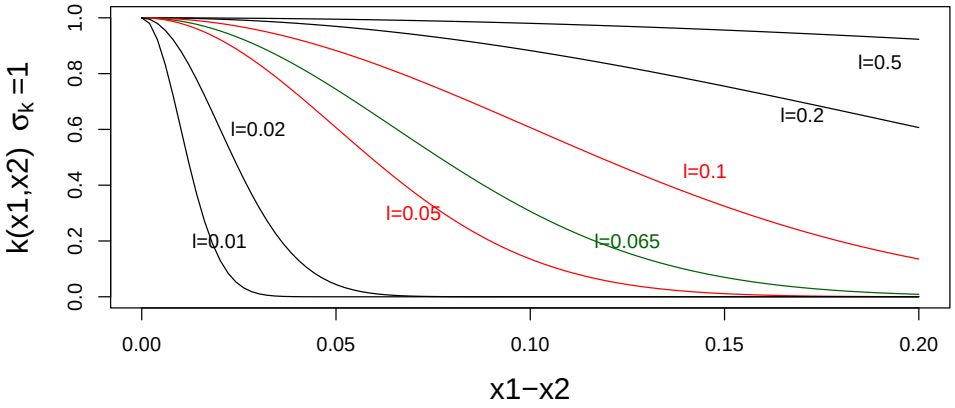

Figure 4.54: A plot of the values of the squared exponential function k against distances between two input values, for input values within the range of true CBD, i.e., 0.35 to 0.55 (so distances vary from 0 to 0.2). The lines represent these functions for different values of the distance-scaling parameter l, demonstrating that the value of $l = 0.065$ corresponds to assuming (roughly) different values of the squared exponential function for different input-value distances.

Assuming a shape parameter of 2, the rate parameter which results in $l = 0.065$ being the prior median is 25.8, to 3 decimal places. That is, the prior for the distance-scaling parameter with CBD as the corresponding input variable is $l_{\text{CBD}} \sim \text{Gamma}(2, 25.8)$. The median is chosen so that it is equally likely for values of l to be greater than 0.065, corresponding to

larger values of the squared exponential function, and less than 0.065, corresponding to smaller values of the squared exponential function. The same method is carried out for choosing the prior distribution for the distance-scaling parameters of SE and BFE, leading to $l_{\text{SE}} \sim \text{Gamma}(2, 16.67)$ and $l_{\text{BFE}} \sim \text{Gamma}(2, 33.33)$.

An informed prior distribution for the signal standard deviation σ_k is also found, given the ranges elicited for TD and AoR. The justification for the prior of σ_k in the simulation examples (i.e., assuming the range of the response variable was [0,1], given the scaling) was developed with the help of Figure 4.26. This plot indicated that a σ_k value of 0.25 appeared to appropriately scale samples of the GP prior on the range [0,1]. Because of this, a prior distribution whose density close to 0 was small and whose mode was close to 0.25 was desired, and for this reasons the prior of $\text{Gamma}(3, \frac{3}{0.4})$ has been used. Given the width of the ranges of the response variables is $0.725 - 0.494 = 0.231$ and $0.528 - 0.25 = 0.278$ for TD and AoR respectively, the 'ideal' value of $\sigma_k = 0.25$ from the simulation examples is divided by 4 (given the intervals have width roughly a quarter of the width of [0,1]), and so a gamma prior with shape 3 and mean 0.1 (0.4 divided by 4) is chosen, i.e., $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.1})$.

Finally, the GP prior means of α_1 and α_2 take the informed prior distributions $\alpha_1 \sim \text{N}(0.6095, 202.8466)$ (for TD) and $\alpha_2 \sim \text{N}(0.389, 198.8303)$ (for AoR), with the means given by the midpoint of the ranges from the elicitation, and the precisions estimated from equating the ranges of each response variable to $3.29\sigma_\alpha$, i.e. assuming an estimate of the standard deviation using a 90% confidence interval for the range of the data.

Moreover, the initialisation of the prior distributions is carried out analogously to that in the previous section, with a Latin hypercube sampling produced to cover the space of possible values of the joint prior distribution for the model parameters.

As in Section 4.4.2, an example of fitting the EIV GP model to the real data is investigated for the purpose of comparing the fit with uninformed prior distributions and with informed prior distributions. This has been carried out with TD as the response variable and CBD as the explanatory variable, and values of TD are predicted for CBD values in the range [0.35,0.55]. The two model fits are presented in Figure 4.55, with the uninformed model fit on the left, and the informed model fit on the right. Similarly to with EIV BR, the 95% credible ellipses of the true values for each powder are much smaller in the right plot than in the left plot, showing more confidence in the true values. In this case with the EIV GP, the widths of the 95% prediction intervals of the response variable are very similar in both cases, unlike in the

EIV BR case, where the prediction interval narrowed significantly with the uninformed priors. This is confirmed by the the summary statistics of the marginal posterior distribution of σ_ϵ in both cases, with the informed-priors case quoted first (and uninformed-priors case second) in the following list: median values of 0.00978 and 0.0107, 2.5% quantiles of 0.00517 and 0.00199 and 97.5% quantiles of 0.0242 and 0.0345 (all to 3 significant figures). The most likely explanation for this is the prior distribution for the model error for the EIV GP with uninformed priors is still somewhat 'informed', unlike in the case of the EIV BR. For the EIV GP, the 'uninformed' prior $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$ is preferred in this work, whose density is much more concentrated compared with the uninformed $\tau_\epsilon \sim \text{Gamma}(0.001, 0.001)$ for the model error precision in the EIV BR. Finally, the effect of the informed prior distribution for the distance-scaling parameter l is evident, with the model fit in the informed-priors case having more curvature than that in the uninformed-priors case (posterior mean in informed-priors case of 0.111, compared with 0.335 in the uninformed-priors case). This is most notable for the predictions as CBD approaches its upper bound of 0.55, with the informed-priors case showing that predictions of TD are more quickly approaching the mean of the response than in the uninformed-priors case.

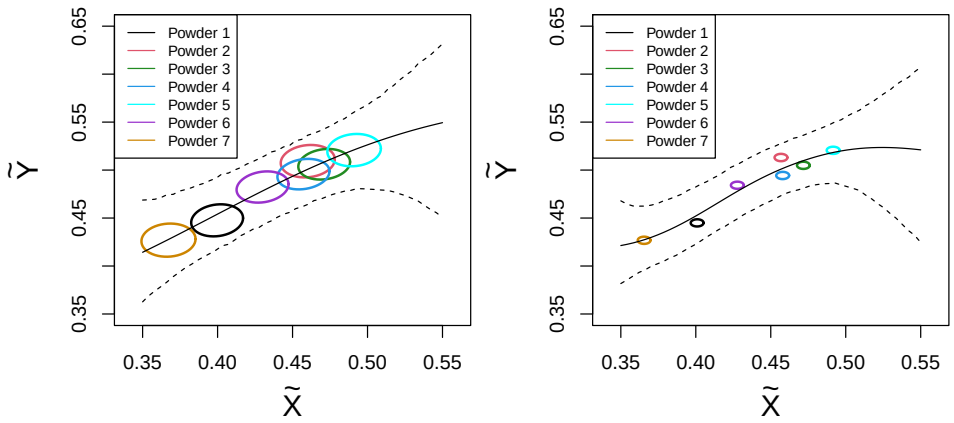

Figure 4.55: Two plots of EIV GP models fitted with TD as the response variable and CBD as the input variable, with the plot on the left corresponding to fitting with uninformed prior distributions, and on the right with informed prior distributions.

Of interest now is the best-fitting model identified in the EIV BR with informed priors, that is, fitting an EIV MOGP with TD and AoR as the output variables and CBD and SE as the input variables, assuming the in-

formed priors noted earlier in this section. The model is fitted with four parallel chains, with the first attempted fit using the MCMC tuning parameters of an adaptation length of 1000 samples, a burn-in length of 25000 samples, and 20000 posterior samples stored having taken every 10^{th} sample of 200000 posterior draws. In this case, the adaptation must be run for longer to achieve optimal sampling behaviour in the algorithm, then the other MCMC tuning parameters are adjusted to ensure that the MCMC output is sufficiently mixed and converges to the posterior distribution. This is achieved for an adaptation of length 3000, a burn-in of length 500000 samples, and taking every 20^{th} sample from 1.5×10^6 posterior samples (giving 75000 posterior samples for each parallel chain).

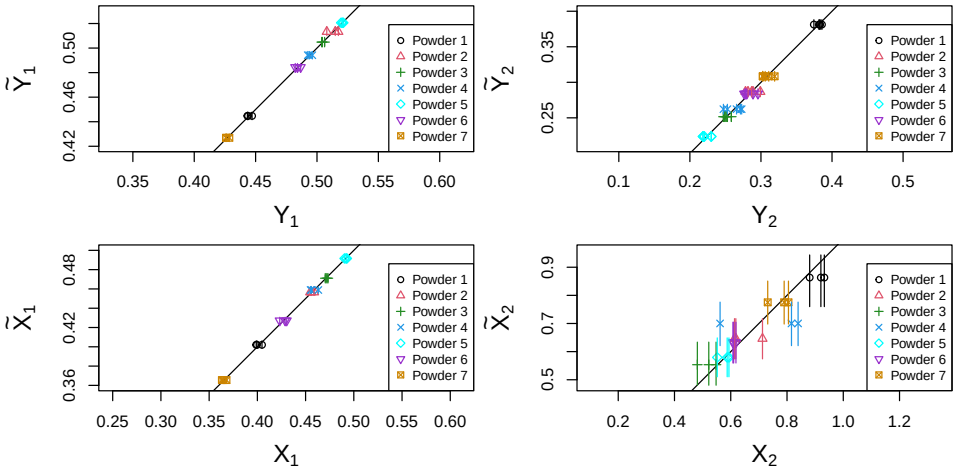

Figure 4.56: Posterior densities of the true values for each powder for the response variable tapped density (top left), the response variable angle of repose (top right), the explanatory variable CBD (bottom left), and the explanatory variable SE (bottom right). These are represented by the 95% credible intervals for the true value (vertical lines), which are plotted for each of the observed data points for the corresponding powder.

The posterior densities of the true values of each variable are firstly checked. These plots are given in Figure 4.56, with TD in the top left, AoR in the top right, CBD in the bottom left, and SE in the bottom right. The true values for TD, AoR and CBD have been estimated well by the model, given the narrow credible intervals and the posterior means lying on the line $\tilde{Y} = Y$ and $\tilde{X} = X$. The true values for SE appear to have a lot of uncertainty. This was also true of the previous section when fitting the data

with the EIV BR models. The posterior means do not align with the line $\tilde{X} = X$ in the cases of Powders 1 and 3—this is not essential, but it is highly likely that the posterior mean for a powder lies within the range of the observed data.

To get an idea of the predictions of the response variable, plots of fitted values for given lines of the input plane are provided in Figures 4.57 and Figure 4.58. That is, the GP posterior for given values of the input variables is summarised in each plot. For example, the first row of Figure 4.57 is a plot of the marginal density of predictions of TD (left plot) and AoR (right plot) for values of CBD, $\tilde{X}_{\text{new},1}$, in the range (0.35,0.55) and values of SE, $\tilde{X}_{\text{new},2}$, equal to

$$\tilde{X}_{\text{new},2} = 1.5 \times \tilde{X}_{\text{new},1} - 0.025.$$

This equation corresponds to the line which connects the minimum values of CBD and SE to the maximum values of CBD and SE. The GP posterior for predictions of the response variables at each of these values of $(\tilde{X}_{\text{new},1}, \tilde{X}_{\text{new},2})$, for a given posterior sample of the hyperparameters and true values s, is given by

$$\tilde{\boldsymbol{Y}}_{\text{new},s} | \boldsymbol{\phi}_s, \tilde{\boldsymbol{X}}_{\text{new}}, \tilde{Y}_s, \tilde{X}_s \sim \text{N}_2(\boldsymbol{m}_s^*, V_s^*) \tag{4.4.4.1}$$

where

$$\boldsymbol{m}_s^* = (\alpha_{1,s}, \alpha_{2,s})' + K_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s) V^{-1}_{\text{MOGP},\tilde{X},s} (\text{vec}(\tilde{Y}_s) - (\boldsymbol{\alpha}_{1,s}, \boldsymbol{\alpha}_{2,s})'), \tag{4.4.4.2}$$

and

$$V_s^* = V_{\text{MOGP},\tilde{\boldsymbol{X}}_{\text{new}},s} - K_s(\tilde{\boldsymbol{X}}_{\text{new}}, \tilde{X}_s) V^{-1}_{\text{MOGP},\tilde{X},s} K_s(\tilde{X}_s, \tilde{\boldsymbol{X}}_{\text{new}}). \tag{4.4.4.3}$$

The solid black line in each plot corresponds to the (marginal) mean of the GP posterior means over the posterior samples, and the dotted lines correspond to 95% prediction intervals at each value of the response variable. The plots in Figures 4.57 and 4.58 therefore provide information as to what the model predicts the behaviour of the response variable to be for given vectors of the inputs—for example, the top-left plot of Figure 4.57 shows predictions of the first response variable given $\tilde{X}_{\text{new},1} \in [0.35, 0.55]$ and $\tilde{X}_{\text{new},2} = 1.5\tilde{X}_{\text{new},1} - 0.025$, where it is clear that for $(\tilde{X}_{\text{new},1}, \tilde{X}_{\text{new},2}) = (0.36, 0.515)$ the predictions of TD are quite uncertain, with a large range of possible values in the 95% prediction interval. In comparison, considering the same plot, the prediction of TD at the input vector $(\tilde{X}_{\text{new},1}, \tilde{X}_{\text{new},2}) = (0.44, 0.635)$ has relatively much less uncertainty, and the GP posterior is

confident that TD lies close to its mean prediction at this input vector. The reason why the GP posterior is more confident in its prediction for some input vectors is that the GP posterior has learnt more about the functional relationship at values close to those input vectors, which ties in with how the covariance kernel is built for the GP—if two input vectors of CBD and SE are close in terms of Euclidean distance, then it is expected that the corresponding vectors of TD and AoR are also relatively close. So, notably from the bottom row of plots of Figure 4.57, corresponding to predictions of TD and AoR for increasing CBD values and decreasing SE values, there appears to be confidence in the predictions of the responses for these input vectors.

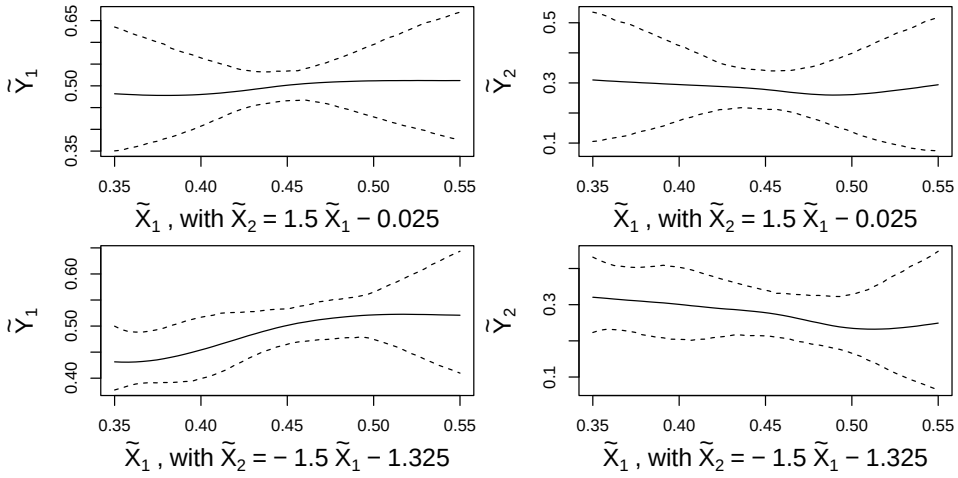

Figure 4.57: Plots of the GP posterior, i.e. predictions of the response variable for some vectors of the inputs, conditioned on the posterior true values of TD, AoR, CBD and SE. The first row corresponds to predictions of the response variables for $\tilde{X}_{\text{new},1} \in [0.35, 0.55]$ and $\tilde{X}_{\text{new},2} = 1.5 \times \tilde{X}_{\text{new},1} - 0.025$, the second row corresponds to predictions of the response variables for $\tilde{X}_{\text{new},1} \in [0.35, 0.55]$ and $\tilde{X}_{\text{new},2} = -1.5 \times \tilde{X}_{\text{new},1} - 1.325$, the first column are the predictions for TD, and the second column are the predictions for the AoR.

It would appear from both Figures 4.57 and 4.58 that the model is performing as intended—the model is providing confidence in the predictions of the response variables where expected, i.e. where there are true values for the input variables. Moreover, at input vectors that are far away from the estimated true values, there is much more uncertainty in the predictions.

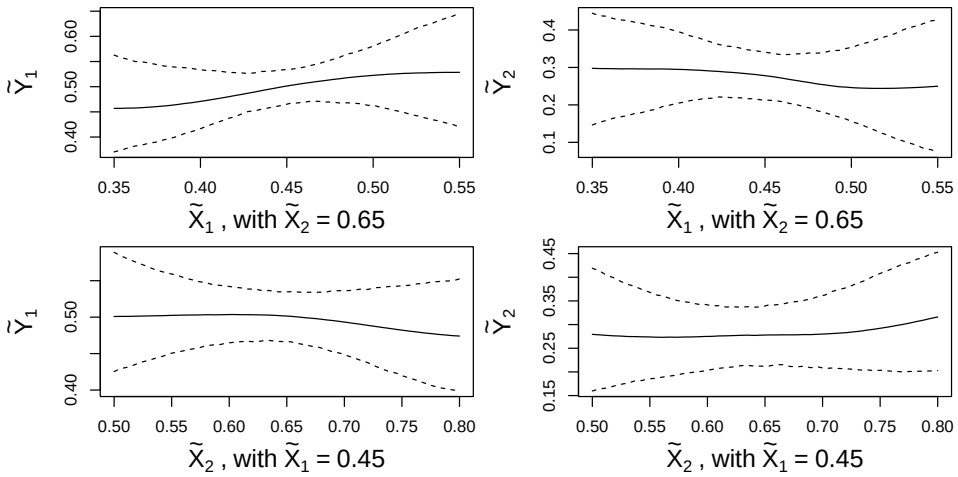

Figure 4.58: Plots of the GP posterior, i.e. predictions of the response variable for some vectors of the inputs, conditioned on the posterior true values of TD, AoR, CBD and SE. The first row corresponds to predictions of the response variables for $\tilde{X}_{\text{new},1} \in [0.35, 0.55]$ and $\tilde{X}_{\text{new},2}$ fixed at 0.65 (the midpoint of the elicited SE range), the second row corresponds to predictions of the response variables for $\tilde{X}_{\text{new},2} \in [0.5, 0.8]$ and $\tilde{X}_{\text{new},1}$ fixed at 0.45 (the midpoint of the elicited CBD range), the first column are the predictions for TD, and the second column are the predictions for AoR.

4.5 Method comparison using real data

The inference from each of the models, having been fitted with tapped density (TD) and angle of repose (AoR) as the (multivariate) response variables and conditioned bulk density (CBD) and specific energy (SE) as the input variables, is now contrasted. This section firstly considers a comparison of the posterior distribution for given parameters which exist in both models (looking mainly at the measurement error and model error standard deviations), followed by comparing the predictions of the response variables for each model for given subspaces of the input space.

	σ_{η_1} (for TD)	σ_{η_2} (for AoR)	σ_{δ_1} (for CBD)	σ_{δ_2} (for SE)
EIV BR	[0.00185, 0.00386]	[0.00599, 0.0102]	[0.00218, 0.00483]	[0.0576, 0.118]
EIV GP	[0.00184, 0.00386]	[0.00599, 0.0101]	[0.00213, 0.00461]	[0.0491, 0.113]

Table 4.16: Comparison of the measurement error standard deviation posteriors from the EIV BR and the EIV GP model fits to the real data, with TD and AoR as the output variables and CBD and SE as the input variables. Each cell gives the centred 95% credible interval for the corresponding measurement error standard deviation and model fit.

As suggested, the posterior standard deviation for the measurement error of the given variables can be compared straightforwardly, as they are dealt with in the same manner for both models. The posterior standard deviation samples for the response variable measurement errors are found by taking the square root of the reciprocal of each posterior precision sample for the response variable measurement errors. That is, for each $s = 1, \ldots, S$, (with S being a large number of posterior samples) the measurement error standard deviations samples $\sigma_{\eta_1,s}$ and $\sigma_{\eta_2,s}$ are given by

$$\sigma_{\eta_1,s} = \sqrt{\frac{1}{\tau_{\eta_1,s}}} \text{ and } \sigma_{\eta_2,s} = \sqrt{\frac{1}{\tau_{\eta_2,s}}}.$$

Furthermore, the posterior standard deviation samples for the explanatory variable measurement errors are found by inverting the precision matrix samples of the joint measurement error for CBD and SE, then taking the square root of each diagonal element (with the first diagonal element of each

matrix corresponding to the posterior variance samples for the measurement error of CBD, and the second element analogously for SE). That is, for each $s = 1, \ldots, S$, the measurement error covariance matrix samples Σ_{δ_s} are given by

$$\Sigma_{\delta_s} = T_{\delta,s}^{-1},$$

where the square roots are taken of its diagonal elements $\sigma_{\delta_1,s}^2$ and $\sigma_{\delta_2,s}^2$ to give $\sigma_{\delta_1,s}$ and $\sigma_{\delta_2,s}$ respectively.

Each measurement error standard deviation is compared between the two models using the centred 95% credible interval over the posterior samples, with these values given in Table 4.16. For the first response variables TD and AoR (whose measurement error standard deviations are σ_{η_1} and σ_{η_2} respectively), both models provide very similar posterior densities.

	σ_{ϵ_1} (for TD)	σ_{ϵ_2} (for AoR)
EIV BR	[0.00386, 0.0202]	[0.00249, 0.0420]
EIV GP	[0.00108, 0.0271]	[0.00142, 0.0525]

Table 4.17: Comparison of the model error standard deviation posteriors from the EIV BR and the EIV GP model fits to the real data, with TD and AoR as the output variables and CBD and SE as the input variables. Each cell gives the centred 95% credible interval for the corresponding model error standard deviation and model fit.

The posterior quantiles of the model error standard deviations σ_{ϵ_1} (corresponding to TD) and σ_{ϵ_2} (corresponding to AoR) are compared in Table 4.17, using the centred 95% credible interval for each parameter. Notably, the EIV GP finds wider credible intervals for the model error standard deviations, with both a smaller lower bound and larger upper bound than the EIV BR case. The nonparametric modelling method tends to provide a more uncertain fit relative to an EIV BR model with an additive linear predictor because of the obvious lack of a predetermined relationship, so this outcome is unsurprising.

The predictive performances of both the EIV BR model fit and the EIV GP model fit are assessed using approximations of exact LOO-CV-IC; for EIV BR, this approximation is the PSIS-LOO-CV-IC, and for EIV GP, this is the mixed LOO-CV-IC. While the approximate methods are different, they clearly still attempt to estimate the same quantity in an attempt to

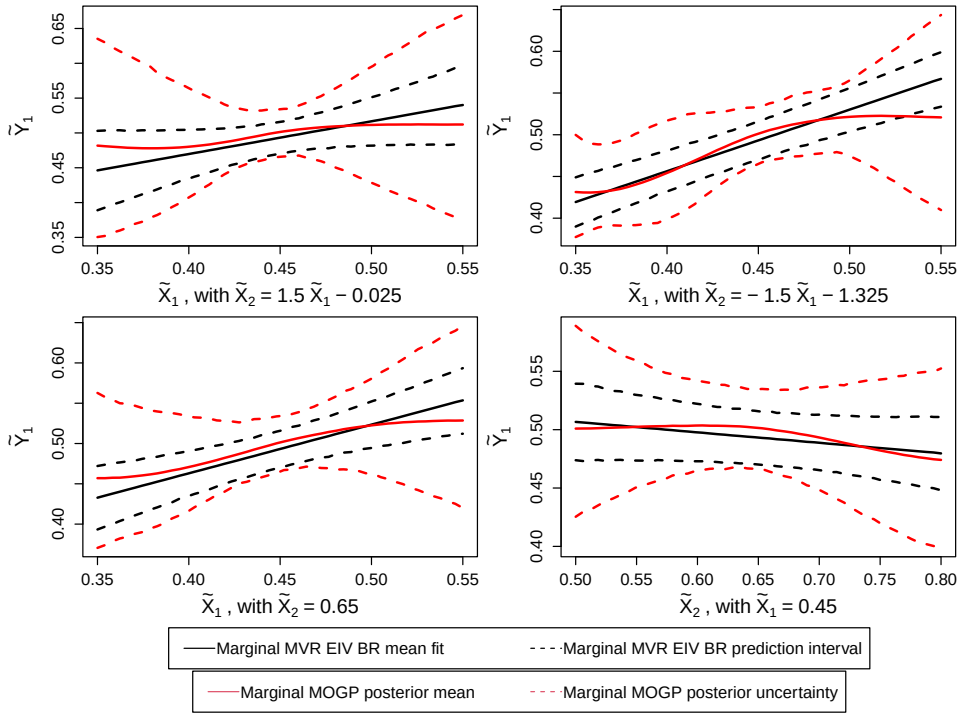

Figure 4.59: Predictions of the response variable TD given lines of the 2D input plane, for comparing the EIV BR and EIV GP model fits. The solid lines represent the mean prediction of the response variable given the model, and the dotted lines represent 95% prediction intervals. Moreover, the red lines correspond to the EIV GP, and the black lines correspond to the EIV BR. Finally, the top-left plot represents predictions of TD for $X_2 = 1.5X_1 - 0.025$, with $X_1 \in [0.35, 0.55]$; the top-right plot for $X_2 = -1.5X_1 + 1.325$ with $X_1 \in [0.35, 0.55]$; the bottom-left plot for $X_2 = 0.65$ with $X_1 \in [0.35, 0.55]$; the bottom-right plot for $X_1 = 0.45$ with $X_2 \in [0.5, 0.8]$. For the bottom-right plot, the variable SE is represented on the horizontal axis; for the other three plots, the variable CBD is represented on the horizontal axis.

convey how likely it is that true value of the response variable for some powder could be produced by the model. As provided in Table 4.15 in Section 4.4.3, the mean estimate of the approximate LOO-CV-IC for the EIV BR model fitted with TD and AoR as the response variables and CBD and SE as the input variables is -77.8. Conversely, the mean estimate of mixed LOO-CV-IC for the EIV GP with the same variables is -67.9, with a direct comparison between the two values indicating that the EIV BR provides a better model fit based on out-of-sample predictive performance. A possible reason for this has been noted earlier in this section, that the nature of the nonparametric modelling to small data sets can lead to overfitting, which can be highlighted by this measure of out-of-sample predictive performance. On the other hand, it is more likely due to the larger estimates of the model error standard deviation in the EIV GP than in the EIV BR.

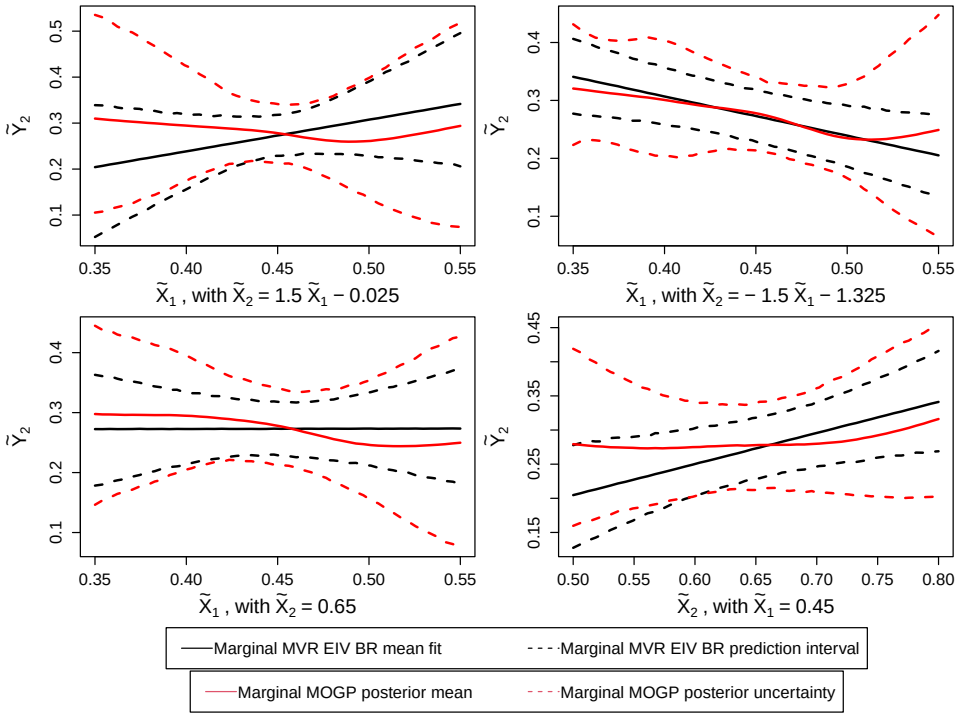

Figure 4.60: The analogous plot to Figure 4.59 for the predictions of the response variable AoR.

Two visual comparisons between the two model fits (specifically their

forward models) are provided in Figures 4.59 and 4.60, with Figure 4.59 providing predictions of TD and Figure 4.60 providing predictions of AoR. In each figure, the four plots correspond to predicting the response variable for different lines within the 2D input plane of possible input vectors. The top-left plot corresponds to predictions of the response for $X_2 = 1.5X_1 - 0.025$, with $X_1 \in [0.35, 0.55]$; the top-right plot for $X_2 = -1.5X_1 + 1.325$ with $X_1 \in [0.35, 0.55]$; the bottom-left plot for $X_2 = 0.65$ with $X_1 \in [0.35, 0.55]$; the bottom-right plot for $X_1 = 0.45$ with $X_2 \in [0.5, 0.8]$. For all plots and for both models, the predictions of the response variables are carried out jointly, with the plots presenting the summaries of the predictions for each variable in turn. As expected with the EIV GP (whose mean prediction is given by the solid red line, and 95% prediction interval is given by the dotted red lines), there is more uncertainty in the predictions of both response variables for more extreme values of the input variable(s) than with the EIV BR (black lines). There is more certainty in the predictions at values closer to the mean (or midpoint) of the explanatory variables, given the narrower prediction intervals towards the centre of each plot. It still appears that the EIV BR has more confidence in the predictions at these points, given the narrower intervals compared with the EIV GP. For the predictions of TD in Figure 4.59, it is noted that the mean predictions of the two methods for each plot are much more similar than the analogous mean predictions for AoR in Figure 4.60. The cause of this is likely due to CBD being such a strong predictor of TD, particularly in comparison with the strength of SE as a predictor of AoR (see the out-of-sample predictive performance assessments from Table 4.5 compared with Table 4.6 and Table 4.8 compared with Table 4.9, albeit with uninformed prior distributions). This strength means that both the EIV GP and EIV BR are more in alignment with predicting TD compared with AoR.

The four straight lines along which the response variables are predicted for the plots in Figures 4.59 and 4.60 are plotted in Figure 4.61, along with 95% credible ellipses for the true values of each group for the two explanatory variables, from both fitted models. This plot gives some indication as to why there is more certainty in the predictions in some areas than in others, which has already been suggested above, but is more clear from the plot. In particular, this is relevant for the EIV GP model fit, which is primarily dependent on predicting the response variable, with higher confidence, at input vectors close to the existing true input vectors provided in Figure 4.61. An example of this is in the top-right plots of Figures 4.59 and 4.60, where the response variables are predicted given $X_2 = -1.5X_1 + 1.325$ and $X_1 \in [0.35, 0.55]$, i.e., the straight black line with negative gradient in Figure

4.61 – the true vector of Powder 7 lies on this line, with its true CBD roughly between 0.36 and 0.37, and the prediction interval for both response variables between CBD values of 0.36 and 0.37 is slightly narrower than values immediately either side of this interval. It is also noted from Figure 4.61 that the major radii for 95% credible ellipses of multiple powders is appreciably smaller for the EIV BR than for the EIV GP. More specifically, for powders 2 to 6, the 95% credible intervals for the true values of SE are narrower for the EIV BR than for the EIV GP. In comparison, the estimated true values of CBD are almost identical for both models. It is possible that the predetermined straight line relationship of the EIV BR imposes more restriction on the true values of SE than the nonparametric nature of the EIV GP, which only occurs for SE due to the higher levels of prior uncertainty in its measurement error precision. The between-materials correlation between SE and CBD is clear; it is also noted that within each group, there is a negative correlation between the true value of SE and the true value of CBD, with the non-vertical alignment of each ellipse.

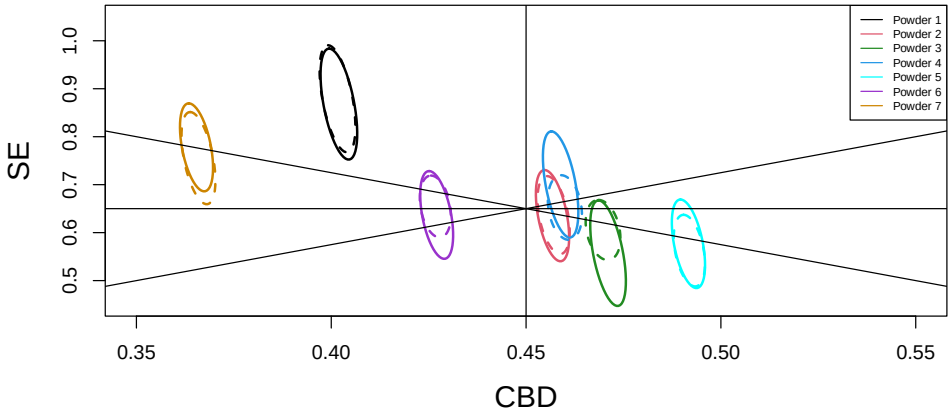

Figure 4.61: Plot presenting the four lines in which the response variables are predicted in Figures 4.59 and 4.60, along with 95% credible ellipses for the true values of each group for the two explanatory variables; these are given for both model fits, with those for EIV GP represented by the solid lines, and those for the EIV BR represented by the dotted lines.

The forward modelling process is concluded here. Having developed the modelling process for EIV BR and EIV GP based on several simulations, the case study on powder data was investigated using both methods. The choice of the best-fitting model was determined using the EIV BR modelling, and it was identified that CBD and SE, with an additive relationship in both

variables, provides the best-fitting model with TD and AoR as the output variables. The backward modelling process is now similarly investigated, with consideration of simulations to develop the modelling process, then the backward modelling is applied to the powder data to determine optimal values of CBD and SE to produce desired values for the output variables TD and AoR.

Chapter 5

Producing desired powder flow

In this chapter, the ultimate aim of this PhD thesis is considered, which is to optimise the values of the explanatory variables (identified from the forward modelling) in order to produce some desired values of the response(s). This backward modelling is investigated in two ways, similarly to Chapter 4, where the effectiveness of the modelling is demonstrated by consideration of simulated data, then the backward modelling process is applied to the real data. The two methods that have been discussed in previous chapters, errors-in-variables Bayesian regression and errors-in-variables Gaussian processes, are investigated here. In particular for the simulated data, many possible linear predictors (for the Bayesian regression) and many possible desired responses are fitted with the backward modelling process.

While Chapter 4 considers the forward modelling process, looking to estimate the relationship between the response variables and the explanatory variables, the backward modelling described here utilises the posterior distribution estimated in the forward modelling to optimise the input variables corresponding to a desired response. This process is discussed in detail in Chapter 3.6, and is summarised briefly later in this chapter.

The first section of this chapter considers fitting the backward model given a forward model fitted to simulated data, where many linear predictors are considered, and given a 'chosen' desired response value. The purpose of this is to ensure that the model performs as intended, and if required, for making adjustments to the modelling process to deal with any issues that occur. The simulations considered here are those from Chapter 4. In some cases, the backward model may be fitted for multiple desired response(s),

that is, the backward modelling process may be carried out multiple times for the same posterior distribution from the forward modelling, and in each case different values may be chosen for the desired response. Some desired response(s) are chosen with the intention of being 'easier' for the backward model to identify optimal input values (i.e., the model has already learned about the relationship between the variables close to these values). In other cases, the backward model is investigated in more challenging scenarios, where more extreme desired responses are considered.

5.1 Errors-in-variables Bayesian regression on simulated data

Of interest here is applying the backward modelling process to each of the Bayesian regression models that have been fitted to simulated data in Section 4.2. That is, the desire here is to 'invert' the modelling process, with the interest now in the behaviour of the input variable(s) given values of the output variable(s). Each Bayesian regression model that is considered (differing in linear predictor and number of response variables) has its own subsection, and in each subsection, multiple values (or vectors in the case of multivariate Bayesian regression) may be considered.

5.1.1 Simple linear model

In this subsection, the backward modelling process is described again in some detail, to reiterate how the process is carried out in this work. This is firstly demonstrated for the simplest parametric model that has been fitted, assuming a straight line relates the response variable to the explanatory variable.

In Section 4.2.1, a simple linear errors-in-variables Bayesian regression model was fitted, where the true values for the response variable were simulated from the equation (before scaling)

$$\widetilde{Y}_i = 3 + 5\widetilde{X}_i + \epsilon_i,$$

where $i = 1, \ldots, 13$. After scaling, the chosen true values for the intercept and slope were $\beta_0 = 0.03$ and $\beta_1 = 5$ respectively, which were effectively recovered in the posterior distribution in the forward modelling.

The task in the backward modelling process is to then optimise the corresponding input variable $X^*|\boldsymbol{\theta}, Y^* = y^*$ to be able to produce some desired value for the response variable $Y^* = y^*$ (note that $\boldsymbol{\theta}$ refers to the

vector of parameters required to fit the backward model). That is, a value x^* is suggested for X^* which is considered the best candidate value for producing $Y^* = y^*$. For each backward modelling process, a large number of posterior samples S_{back} are taken from $\boldsymbol{\theta}$, so that a candidate value can be found for X^* for each of those samples, which allows a posterior distribution to be built for X^* given the uncertainty in the forward modelling.

With the example of the simple linear model, the backward modelling process is carried out by fitting the model

$$Y^* = \beta_{0,b} + \beta_{1,b} X_b^* + \epsilon_b^*, \tag{5.1.1.1}$$

where the variable Y^* is fixed at some desired response value y^*, and X_b^* is the corresponding input variable, which is assumed to be random. Moreover, $\beta_{0,b}$ and $\beta_{1,b}$ are the model coefficients of the b^{th} posterior sample from the forward model, and the random error term ϵ_b^* assumes a normal distribution with mean 0 and precision $\tau_{\epsilon,b}$, the model error precision from the b^{th} posterior sample of the forward model. That is, the joint behaviour of the model coefficients and the model error precision is preserved by choosing the same posterior sample from the forward model (in other words, they are sampled jointly from the posterior). In this case, of interest is the distribution $X^*|\boldsymbol{\theta}, Y^* = y^*$, where $\boldsymbol{\theta} = (\beta_0, \beta_1, \tau_\epsilon)$. This distribution is built based on samples from the posterior distributions $X_b^*|\boldsymbol{\theta}_b, Y^* = y^*$ for $b = 1, \ldots, S_{\text{back}}$, with S_{back} being some subset of the S_{back} random samples stored from the forward model posterior distribution. That is, for each posterior sample from the forward model $b = 1, \ldots, S_{\text{back}}$, a posterior distribution is found for the corresponding variable X^*, given by X_b^*, and a random sample can be drawn from each X_b^* and collected together to build the complete posterior distribution for X^*. The distribution $X^*|\boldsymbol{\theta}, Y^* = y^*$ is often shortened here to the distribution X^* (and similarly when the subscript b is considered).

A uniform prior distribution is placed on each X_b^*. Given the data are scaled onto the range [0,1], these are appropriate bounds for an initial fitting of the backward model. One should consider the range of the data after scaling, and adjust appropriately. In this case, the range of the input variable is [0.01,0.13] (roughly speaking), and so a range of [0,1] is a relatively notable extrapolation of the data. These endpoints could also be adjusted given prior information.

To carry out the backward modelling process, $S_{\text{back}} = 2000$ posterior samples are randomly selected from the forward model posterior. The first 'chosen' desired response value is $Y^* = 0.38$, which is calculated by rounding the mean of the 'chosen' true values from the forward model (after scaling)

to 2 decimal places. That is, the underlying true values of the response variable for this simulation were simulated from the equation

$$\tilde{Y}_i = 3 + 5\tilde{X}_i + \epsilon_i,$$

with ϵ_i simulated from a normal distribution with mean 0, precision 4. Note further that $\tilde{\boldsymbol{X}} = (1, 2, \ldots, 13)$. This provides the 13 true values of the response variable from the underlying simulation, where the true values are scaled onto the range [0,1], and the mean of these values (rounded to 2 d.p.) is chosen to be the desired response. This 'chosen' desired response provides a straightforward test for the backward modelling process, since the model should have the best understanding of the relationship between the output and the input towards the mean of the output variable.

For each $b = 1, \ldots, 2000$, the model

$$0.38 = \beta_{0,b} + \beta_{1,b} X_b^* + \epsilon_b^*$$

is fitted with $X_b^* \sim \mathrm{U}(0, 1)$ and $\epsilon_b^* \sim \mathrm{N}(0, \tau_{\epsilon,b})$, in order to find a posterior distribution for X_b^*. The model is run with an adaptation phase of length 500, a burn-in of length 15000, then 2000 posterior samples are stored from 20000 posterior samples, taking every 10^{th} sample. Each backward model is initialised with different values for the random variable X_b^* (randomly sampled from its prior distribution), to demonstrate convergence for each backward model run with overdispersed starting values. In other words each X_b^* is assigned a starting point for the respective MCMC run, chosen randomly from a sample of 2000 values of the distribution U(0,1). Each backward model is also run with two parallel chains, so that the PSRF can be calculated. For these simulated-data cases, the same initial values are assigned to X_b^* for each chain, which is an error and is rectified for the backward modelling process on the real data. Before combining posterior samples from X_b^* for $b = 1, \ldots, 2000$ to build the posterior distribution for X^*, the metrics PSRF (potential scale reduction factor) and effective sample size are estimated to test that the MCMC output is working effectively, for each $b = 1, \ldots, 2000$. An initial check is to find the maximum value of the upper bounds on the 95% confidence interval for PSRF, which in this case gives 1.020. This is below the value of 1.1, which is typically used as the cutoff for appropriate values of PSRF (recall from Section 3.1.4). Given this is the maximum over all backward models, it is fair to say that each backward model has converged to its posterior distribution. The second check is to find the minimum value of effective sample size of X_b^* from each parallel chain of each backward model. This is given by 1284 (rounded to 0

decimal places), which suggests that each chain, and each backward model, has mixed effectively (comfortably larger than the cutoff of $5m$, with m being twice the number of parallel chains). For the subsequent backward modelling processes carried out in this chapter, the PSRF and effective sample sizes are only discussed when the criteria for convergence and mixing are not met (i.e., only where there are issues with the convergence and mixing in the backward modelling).

The complete posterior distribution for X^* can now be compiled, which is performed by taking 10 random samples from each posterior distribution X_b^* for $b = 1, \ldots, 2000$, giving 20000 posterior samples of X^*, whose density is now examined. Given the linear predictor chosen for this simulation, it is known that a value of 0.07 (after scaling) would produce the desired response $Y^* = 0.38$, by solving

$$0.38 = 0.03 + 5x^*.$$

In essence, with the backward models carried out on simulated data, one of the main aims is to try to recover this value of x^* as a high-density value in the posterior distribution of $X^*|Y^* = 0.38$. A plot of the posterior density of $X^*|Y^* = 0.38$ is provided in Figure 5.1, along with two vertical lines: the red value indicates the true corresponding value of $x^* = 0.07$, and the green line indicates the posterior mode of $X^*|Y^* = 0.38$, given by 0.0701.

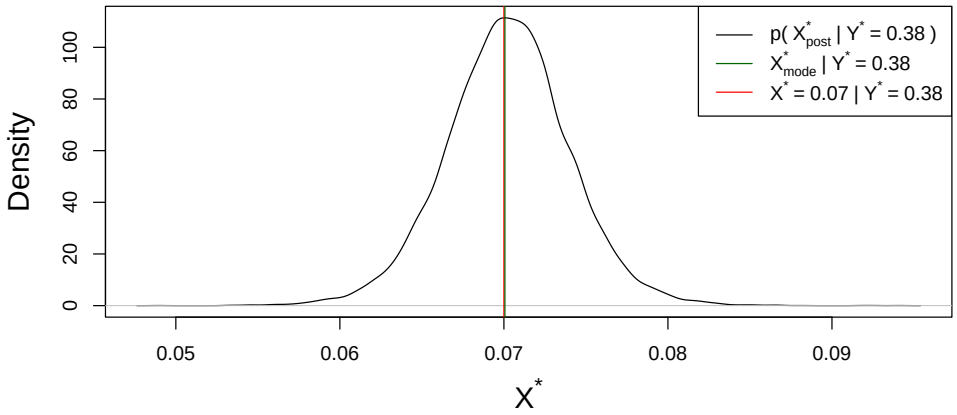

Figure 5.1: The posterior density of X^* given the desired response value of $Y^* = 0.38$, along with two vertical lines: the red indicating the real value of $x^* = 0.07$ and the green indicating the posterior mode of X^* of 0.0701.

Clearly the backward model is working as intended, with the posterior mode of X^*, given by the green vertical line, lying very close to the real value

$x^* = 0.07$. Some level of uncertainty is expected here given the setup of the backward model, acknowledging the uncertainty in the model for the model coefficients and model error term.

The candidate point that is suggested is the posterior mode, which maximises the posterior density of X^*. The posterior mode is given by $X^*_{\text{mode}} = 0.0701$ to 3 significant figures. This point can be tested to see how well it produces the desired response $Y^* = 0.38$ by providing a distribution of predictions of the response variable at the posterior mode, with each value given by

$$Y_{\text{pred},s} = \beta_{0,s} + \beta_{1,s} X^*_{\text{mode}} + \epsilon_s,$$

where $\epsilon_s \sim \text{N}(0, \tau_{\epsilon,s})$. Note the use of the subscript s, where $s = 1, \ldots, S$ with $S = 20000$, i.e., using the entire samples drawn from the forward model posterior distribution to provide predictions of the response given X^*_{mode}. These values are summarised here with a plot of the density of $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode}}$, provided in Figure 5.2, along with two vertical lines representing the desired response $Y^* = 0.38$ (red) and the mode of $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode}}$ (green), given by 0.379 (to 3 significant figures). In this case, the posterior mode is an effective point for being able to produce the desired response.

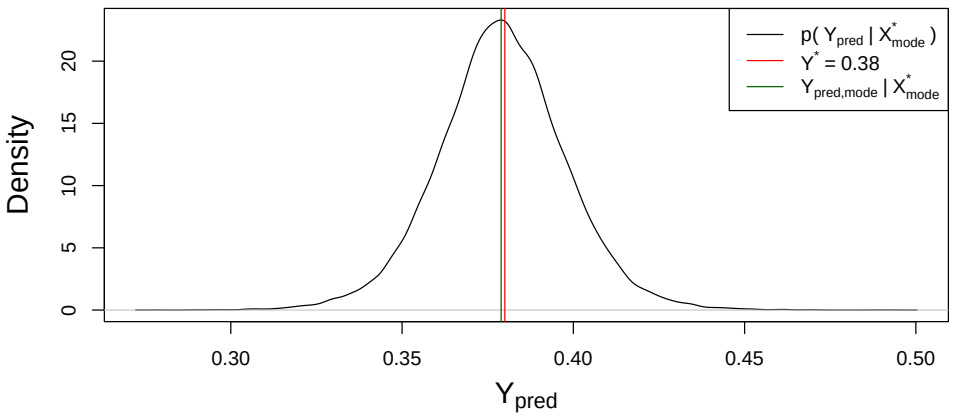

Figure 5.2: Probability density of $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode}}$, where $X^*_{\text{mode}} = 0.0701$ is the posterior mode of X^* given the desired response value of $Y^* = 0.38$. The red line represents this desired response value, and the green line represents the value 0.379, the mode of the distribution $Y^*|\boldsymbol{\theta}, X^*_{\text{mode}}$.

A second backward modelling attempt is now performed with the desired response value $Y^* = 0.85$, which is found by taking the maximum of the existing true values of the response variable (the underlying values, as opposed

to the estimated posterior values), multiplied by 1.25 (to ensure this is an extreme point that has not yet been observed), and rounded to 2 decimal places. The demand of this example is that there is no data provided in the forward modelling that exists close to this desired response value, and so the backward model is forced (at least in this case—it is not necessarily true that an 'extreme' value of the response variable requires an extreme value of the explanatory variable in order to produce it) to search in a new region of the input variable space (given the forward model posterior distribution) to find appropriate candidate values. An identical process to what is described above is now carried out.

Given the scaled chosen value of the linear predictor, the value of the explanatory variable that produces the desired response of $Y^* = 0.85$ is given by $x^* = \frac{0.85-0.03}{5} = 0.164$. The posterior density of $X^*|Y^* = 0.85$ is given in Figure 5.3, along with two vertical lines: the red value indicates the true corresponding value of $x^* = 0.164$, and the green line indicates the posterior mode of $X^*|Y^* = 0.85$.

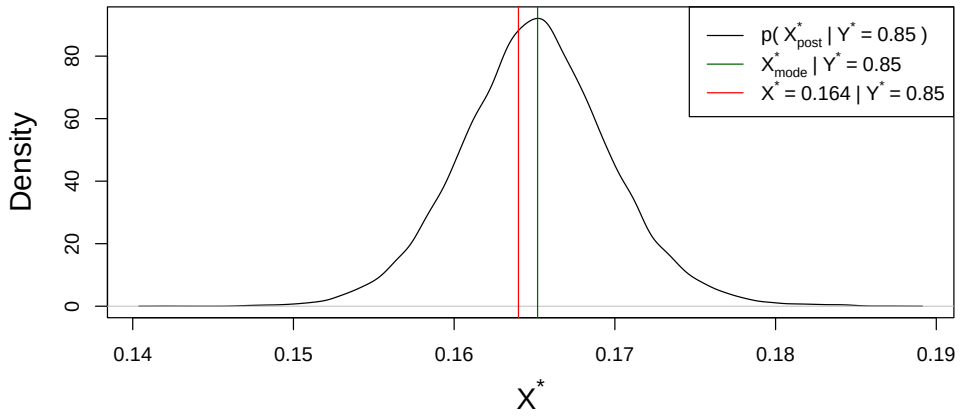

Figure 5.3: The posterior density of X^* given the desired response value of $Y^* = 0.85$, along with two vertical lines: the red indicating the real value of $x^* = 0.164$ and the green indicating the posterior mode of X^* of 0.165.

The posterior distribution for X^* in this case of $Y^* = 0.85$ is adequate—the real value of $x^* = 0.164$ is a high-density value in the posterior. The further distance between the posterior mode and the real value of x^*, compared with that found in Figure 5.1, is likely due to the model having to extrapolate, having to find values of the explanatory variable that have not been seen in the forward model in order to produce a response value that has also not been seen in the forward model (neither in the simulated obser-

vations nor in the estimates of the true values). This posterior mode is given by $X^*_{\text{mode}} = 0.165$ to 3 significant figures, and is tested in the forward model to see what values of the response variable it produces. A plot of the probability density of prediction values of $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode}}$ with $X^*_{\text{mode}} = 0.165$, is given in Figure 5.4. A pair of vertical lines are superimposed, showing the desired response $Y^* = 0.85$ (red) and the mode of $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode}}$ (green), which is equal to 0.852 to 3 significant figures.

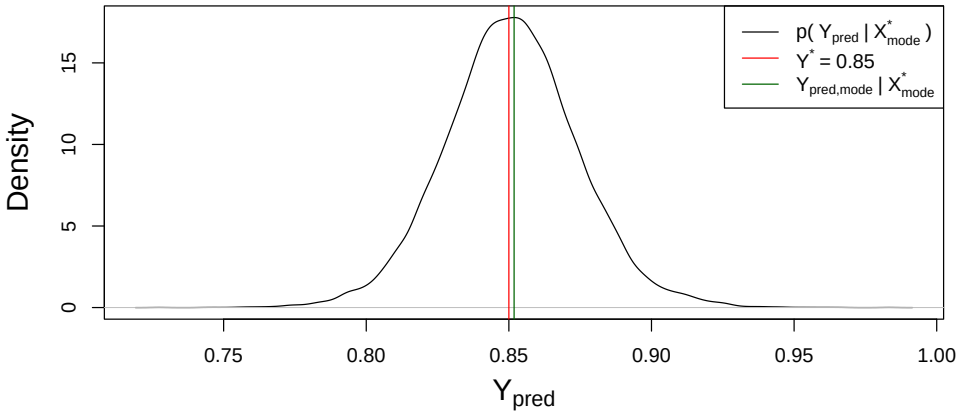

Figure 5.4: Probability density of $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode}}$, where $X^*_{\text{mode}} = 0.165$ is the posterior mode of X^* given the desired response value of $Y^* = 0.85$. The red line represents this desired response value, and the green line represents the value 0.852, the mode of the distribution $Y^*|\boldsymbol{\theta}, X^*_{\text{mode}}$.

As with the previous case of $Y^* = 0.38$, the case of $Y^* = 0.85$ provides evidence that the backward model is able to find suitable candidate values for X^* to be able to produce desired response values, given that $Y^* = 0.85$ is a high-density value in the distribution of the predictions given the posterior mode $X^*_{\text{mode}} = 0.165$.

5.1.2 Cubic model

The backward model corresponding to the cubic EIV BR model fitted in Section 4.2.2 is now considered. This introduces another considerable difficulty for the backward model; consider the bottom-left plot of Figure 4.6, the fitted model found for the cubic simulation, which shows that there are multiple values for the input variable that can produce a single response value.

This is tested by considering the desired response value $Y^* = 0.404$ (after

scaling). Rough estimates for input values that could produce this desired response for the cubic simulation (found by trial and error) are given by $x^* = 0.0512$ and $x^* = 0.122$ (both rounded to 3 significant figures). For this cubic simulation, the backward model of interest is

$$Y^* = \beta_{0,b} + \beta_{1,b}X_b^* + \beta_{11,b}(X_b^*)^2 + \beta_{111,b}(X_b^*)^3 + \epsilon_b^*,$$

with $\epsilon^* \sim \mathrm{N}(0, \tau_{\epsilon,b})$, and $\beta_{0,b}$, $\beta_{1,b}$, $\beta_{11,b}$, $\beta_{111,b}$ and $\tau_{\epsilon,b}$ being jointly sampled from the forward model posterior distribution. This is fitted for each joint posterior sample $b = 1, \ldots, S_{\text{back}}$, with the prior distribution $X_b^* \sim \mathrm{U}(0, 1)$ for each backward model fit. For each $b = 1, \ldots, S_{\text{back}}$, the MCMC output of X_b^* is checked for mixing and convergence, and a small posterior sample is randomly selected from each X_b^* and combined to build a complete posterior distribution for X^*.

There are issues here with the mixing and convergence of each of the backward models for this example. The major issue lies with the convergence of the MCMC output to the posterior distribution, which is assessed initially using the PSRF. From all S_{back} backward models, the largest upper bound for the 95% confidence interval of PSRF for X_b^* is 287.3, which would initially suggest that the X_b^* is not close to converging. Looking further into these upper bounds for each backward model, there appear to be several backward models for which the upper bound is small—looking at the summary statistics for the upper bound of PSRF for each backward model, the median upper bound is 1.0086, to 4 decimal places, suggesting that at least half of the backward models have converged, according to PSRF. On the other hand, the 75% quantile is 144.6, so there are clearly a significant percentage of the backward models that are not converging. Deeming a backward model to not have converged if the upper bound of the 95% confidence interval for the PSRF estimate is greater than 1.1, there are 958 out of 4000 (note that for each $b = 1, \ldots, 2000$, 2 parallel chains are run) backward models that have not converged, meaning 3042 backward models of the 4000 appear to have converged. It is noteworthy that the smallest upper bound for the PSRF of those backward models that have not converged is 56.0 to 3 significant figures, which is significantly larger than the criterion for convergence. This would suggest there is a substantial issue with the backward models. This is investigated further in Appendix E, to show that genuine convergence to the posterior distribution occurs in 1 of the $S_{\text{back}} = 2000$ backward model runs.

As such, a solution for this issue is not investigated further in this work, and the investigation of backward models fitted to different simulation examples continues, despite this issue of a lack of convergence. While the

relative densities surrounding posterior modes in the complete backward model posterior for X^* cannot be compared, there is sufficient evidence that the modes that are recovered are appropriate for the given simulation and desired response.

The focus is re-centred onto the complete posterior distribution of X^*, which is plotted in Figure 5.5. Despite the issues discussed above with the convergence of each backward model run, it is clear that the backward model algorithm is still able to identify both possible candidate values with a high level of accuracy. There are two clear peaks of density, and both are centred very close to the real values identified earlier.

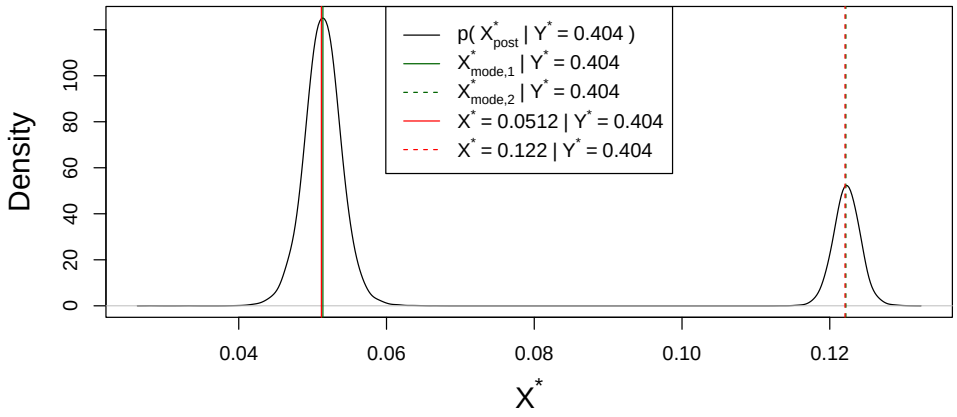

Figure 5.5: The posterior density of X^* for the cubic model simulation, given the desired response value of $Y^* = 0.404$. This posterior density is multimodal, with the 'first' mode (i.e., the mode with the highest density) indicated by the solid green line, and the 'second' mode indicated by the dotted green line. The two 'real' values are also indicated on the plot, with $x^* = 0.0512$ given by the solid red line (solid as it is relatively close to the first mode), and $x^* = 0.122$ given by the dotted red line.

The posterior modes recovered in the backward modelling are now compared by running these values through the forward model, given the posterior distribution estimated for the model parameters β_0, β_1, β_{11}, β_{111} and τ_ϵ. This provides a distribution of predictions of the response variable given each posterior mode, whose densities are directly compared in Figure 5.6. The mode of the distribution of predictions of the response is slightly closer to the desired response value in the case of the second mode. In contrast, the variance of the distribution of predictions of the response variable is noticeably larger given the second mode. For these reasons, the question of

which posterior mode is the better candidate value is not straightforward to answer. As discussed in the following example, this can depend on the question being asked—it appears that the second mode is more likely to produce the desired response, but it appears that the first mode is more likely to produce values within, say, 5% of the desired response value. If values of the response variable within 5% of the desired response are equally valuable, then the first mode is preferable (63.5% of the distribution of predictions given the first mode are within 5% of the desired response, whereas this percentage drops 51.0% for the second mode).

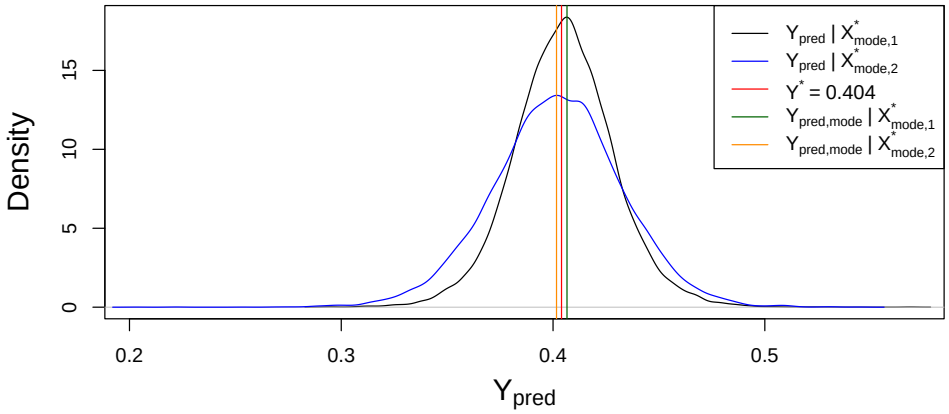

Figure 5.6: Probability densities of $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode},1}$ (black line) and $Y_{\text{pred}}|\boldsymbol{\theta}, X^*_{\text{mode},2}$ (blue line), where $X^*_{\text{mode},1} = 0.0514$ is the 'first' posterior mode of X^* given the desired response value of $Y^* = 0.404$ and $X^*_{\text{mode},1} = 0.122$ is the 'second' posterior mode of X^* given the desired response value of $Y^* = 0.404$ (red line). The green line represents the value 0.407 and the orange line represents the value 0.402, the modes of the distributions $Y^*|\boldsymbol{\theta}, X^*_{\text{mode},1}$ and $Y^*|\boldsymbol{\theta}, X^*_{\text{mode},2}$ respectively.

5.1.3 Two explanatory variables

In this section, EIV BR models with two explanatory variables are considered in the backward modelling. That is, the corresponding input variable X^* to the desired response value $Y^* = y^*$ becomes 2-dimensional. The multidimensional corresponding input variable is denoted as $\boldsymbol{X}^* = (X^*_1, X^*_2)$, whose prior distribution can take a variety of 2-dimensional shapes. As discussed in Section 3.6.5, there are several possible prior specifications for the corresponding input vector $\boldsymbol{X}^*$. If, a priori, there are no known assumptions

for the possible range of values of the input variables as well as for the relationship between the input variables, then placing a uniform box over $\boldsymbol{X}^*$ (with lower bound 0 and upper bound 1) is the most logical prior, and this is carried out for this example. Moreover, as hinted in the previous section, it may be desirable to produce values of the response within a certain range of the desired response value $Y^* = y^*$; for example, values within 1% of the desired response value $Y^* = y^*$ may be just as desirable as $Y^* = y^*$, in which case we want to optimise $\boldsymbol{X}^*$ in order to produce desired response values within the range $(0.99y^*, 1.01y^*)$. This is also discussed in Section 3.6.6.

Returning to the example from the forward modelling with one response variable and two explanatory variables, consider the plot in Figure 5.7, which provides a heatmap of the simulation on which the example is based (having scaled the data), i.e.,

$$\tilde{Y} = 0.3 + 8\tilde{X}_1 - 0.5\tilde{X}_2.$$

Here, the horizontal axis corresponds to the first input variable, the vertical axis corresponds to the second input variable, and the coloured heatmap indicates the value of the response variable corresponding to the vector of inputs $(\tilde{X}_1, \tilde{X}_2)$. Furthermore, a red region is indicated on the plot, which corresponds to vectors of the input variables which are able to produce values of the response variable within the range $(0.99 \times 0.34, 1.01 \times 0.34)$, that is, values within 1% of some desired response value $Y^* = 0.34$ (found by taking the mean of the observed data for the response variable and rounding to 2 decimal places). In theory, the backward model should be able to identify these vectors as suitable candidates for the corresponding input vector $\boldsymbol{X}^*$. The backward model of interest is therefore

$$0.34 = \beta_{0,b} + \beta_{1,b}X_1^* + \beta_{2,b}X_2^* + \epsilon_b^*,$$

where $\epsilon_b^* \sim \mathrm{N}(0, \tau_{\epsilon,b})$, with the additional assumption that, for some variable Y',

$$Y' = Y^* + Z,$$

with $Z \sim \mathrm{U}(-0.01 \times 0.34, 0.01 \times 0.34)$. Thus Z implies a range of values 1% either side of the desired response value $y^* = 0.34$. For the initial vector of values used for $\boldsymbol{X}_b^*$ in each model, Latin hypercube sampling is carried out with two variables and $S_{\text{back}} = 2000$ samples, and for each backward model b, a random sample of the Latin hypercube sampling is chosen as the initial value for $\boldsymbol{X}_b^*$ in the MCMC. This is carried out to ensure that the backward modelling is able to converge to the same posterior distribution from all possible starting vectors for $\boldsymbol{X}^*$.

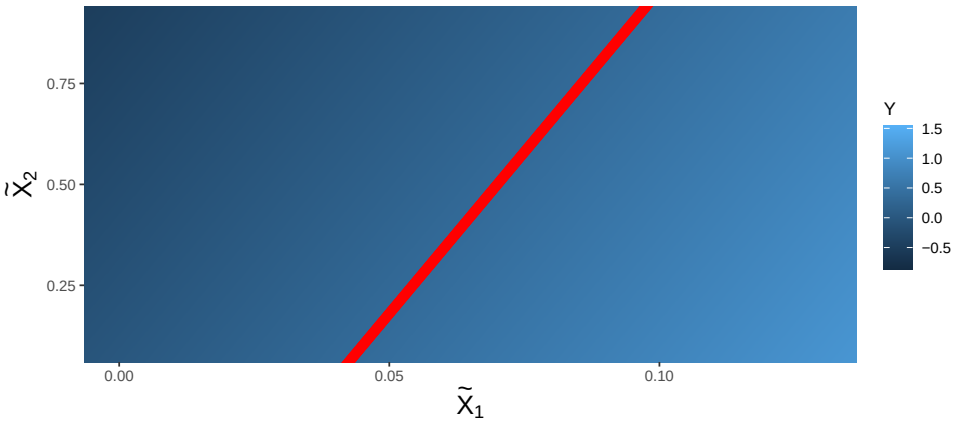

Figure 5.7: A heatmap of the simulation with two inputs and one response variable (additive relationship between inputs), with the points highlighted in the plot that satisfy the inequalities $0.99 \times 0.34 \leq 3 + 8\tilde{X}_1 - 5\tilde{X}_2 \leq 1.01 \times 0.34$.

The backward model algorithm as described directly above is then fitted, and the PSRF is used to check the convergence of each model. With the corresponding input now being a vector, the PSRF is estimated for both $X^*_{1,b}$ and $X^*_{2,b}$ for each backward model iteration in $b = 1, \ldots, S_{\text{back}}$, and so the maximum of the upper bound on the 95% confidence interval of these estimates is calculated. In an initial run, the PSRF is 2.18 to 3 significant figures. This is significantly larger than the cutoff of 1.1 for convergence, suggesting an issue with at least one of the backward models. A 95% confidence interval of these upper bounds (i.e., the upper bounds of PSRF for $X^*_{1,b}$ and $X^*_{2,b}$ for each $b = 1, \ldots, S_{\text{back}}$ is given by [1.00,1.16] to 2 decimal places, indicating that a significant majority of the backward models are close to convergence. Inspecting further shows that the 92.75% quantile of these values is 1.099 to 3 decimal places, indicating that over 92% of the backward models are converging. In this case, it is fair to suggest that some adjustment to the MCMC parameters is suitable, with increasing the number of posterior samples drawn from each backward model (that is for the purpose of assessing the convergence of each $X^*_{1,b}$ and $X^*_{2,b}$, as opposed to drawing further samples of these distributions to build the complete posterior of $\boldsymbol{X}^*$) being a logical step. Before continuing, the effective sample size estimates are also calculated to check the levels of autocorrelation in the MCMC are sufficiently small. The smallest estimate of effective sample size for either parameter in any of the backward models is 21.1 to 3 significant

figures, and a 95% credible interval for effective sample size estimates over all backward models and each random variable within the backward model is given by [35.8,269] to 3 significant figures. It would seem that the mixing is only a significant issue for a small percentage of the backward models.

With a burn-in length of 20000 samples, and 15000 posterior samples stored (taking every 10^{th} sample from 150000 samples), it is noted that 1999 out of 2000 backward models achieve convergence (with the remaining backward model having a PSRF upper bound of 1.11), and sufficient mixing is observed. This is deemed to be sufficient, and the complete posterior distribution of $\boldsymbol{X}^*$ is built, whose posterior density is now considered. Note that the marginal posterior densities of X_1^* and X_2^* do not tell much of the story. It is clear from the plot of possible vectors of $\boldsymbol{X}^*$ that could produce the response variable within the desired range of $[0.99\times0.34, 1.01\times0.34]$ that there is some relationship between these two variables: a given value of one of these variables gives a small range of possible values for the other variable. A joint posterior density of $\boldsymbol{X}^*|Y^* \in [0.99 \times 0.34, 1.01 \times 0.34]$ is provided in Figure 5.8, with the red region of possible input vectors identified in Figure 5.7 superimposed. The joint posterior density of $\boldsymbol{X}^*$ is clearly performing well, identifying the entire region of input vectors from Figure 5.7 as high-density vectors in the posterior density.

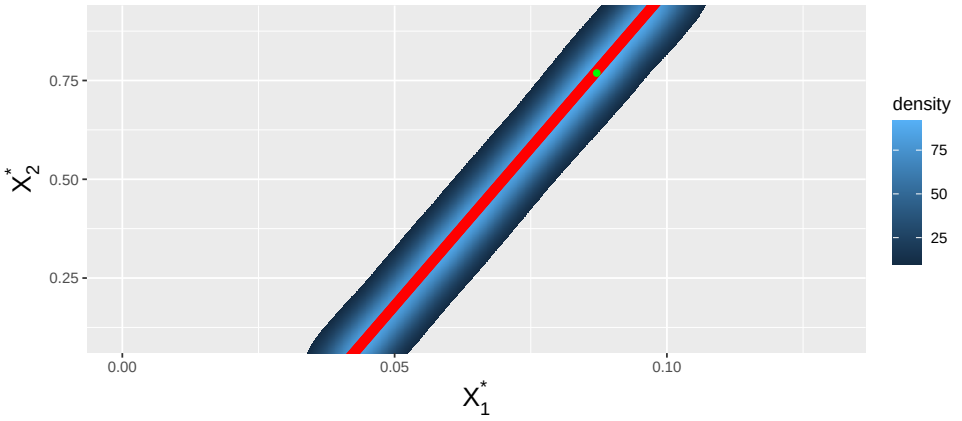

Figure 5.8: Joint posterior density of $\boldsymbol{X}^*$ given a desired response range of $[0.99 \times 0.34, 1.01 \times 0.34]$ (blue region), with the red region of input vectors identified in Figure 5.7 as vectors that produce the desired response to within the specified range superimposed. The green point indicates the joint posterior mode of $\boldsymbol{X}^*$, given by $\boldsymbol{X}^*_{\text{mode}} = (X^*_{1,\text{mode}} = 0.0871, X^*_{2,\text{mode}} = 0.769)$.

The joint posterior mode of $\boldsymbol{X}^*$ is identified as $\boldsymbol{X}^*_{\text{mode}} = (X^*_{1,\text{mode}} =$

$0.0871, X^*_{2,\text{mode}} = 0.769$), rounded to 3 significant figures. This posterior mode is now tested using the posterior distribution from the forward model to provide a distribution of predictions of the response variable. These predictions of the response variable are given by

$$Y_{\text{pred},s} = \beta_{0,s} + \beta_{1,s} X^*_{1,\text{mode}} + \beta_{2,s} X^*_{2,\text{mode}} + \epsilon_s,$$

where $\epsilon_s \sim \text{N}(0, \tau_{\epsilon,s})$, and the density of these predictions are provided in Figure 5.9, with the vertical green line representing the mode of $Y_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode}}$, 0.337, the solid red line representing the desired response value of $Y^* = 0.34$, and the dotted red lines representing the desired response range $[0.99 \times 0.34, 1.01 \times 0.34]$. The mode of the distribution of predictions overlaps the lower bound of the desired response range; these values are equal to 4 decimal places. It is satisfactory to see that the desired response range is practically centred in the distribution of predictions. We observe in this case that the backward modelling algorithm is working effectively.

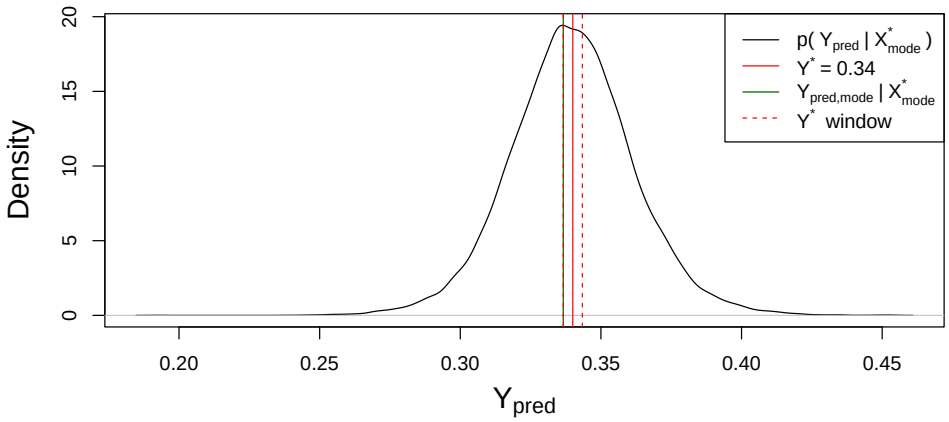

Figure 5.9: Probability density of $Y_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode}}$ (black line), where $\boldsymbol{X}^*_{\text{mode}} = (0.0871, 0.769)$ is the posterior mode of $\boldsymbol{X}^*$ given the desired response range of $[0.99 \times 0.34, 1.01 \times 0.34]$. The green line represents the value 0.337, the mode of the distribution $Y^*|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode}}$, the solid red line corresponds to the desired response value $Y^* = 0.34$, and the dotted red lines are the desired response range.

5.1.4 Two explanatory variables with interaction term

The backward modelling investigation with simulated data continues with the example considered in Section 4.2.5 of one response variable, two ex-

planatory variables and a full quadratic linear predictor. In this backward modelling example, the model of interest is therefore

$$Y^* = \beta_{0,b} + \beta_{1,b} X^*_{1,b} + \beta_{2,b} X^*_{2,b} + \beta_{11,b} (X^*_{1,b})^2 + \beta_{12,b} X^*_{1,b} X^*_{2,b} + \beta_{22,b} (X^*_{2,b})^2 + \epsilon^*_b,$$

where $\epsilon^*_b \sim \mathrm{N}(0, \tau_{\epsilon,b})$. As in the previous section, the initial values of $(X^*_{1,b}, X^*_{2,b})$ are randomly selected from a Latin hypercube sampling with two variables and $S_{\text{back}} = 2000$ samples, with the uniform 'box' prior distribution over $[0,1] \times [0,1]$.

The desired response value for Y^* is selected to be $Y^* = 0.205$, which is calculated by taking the 40% quantile from the 'chosen' true values in the forward modelling. The relationship between the response variable and the two explanatory variables, given the chosen values of the model parameters, is provided in Figure 5.10 as a contour plot, along with a red curve, which indicates vectors of the two explanatory variables that are able to produce the desired response value $Y^* = 0.205$.

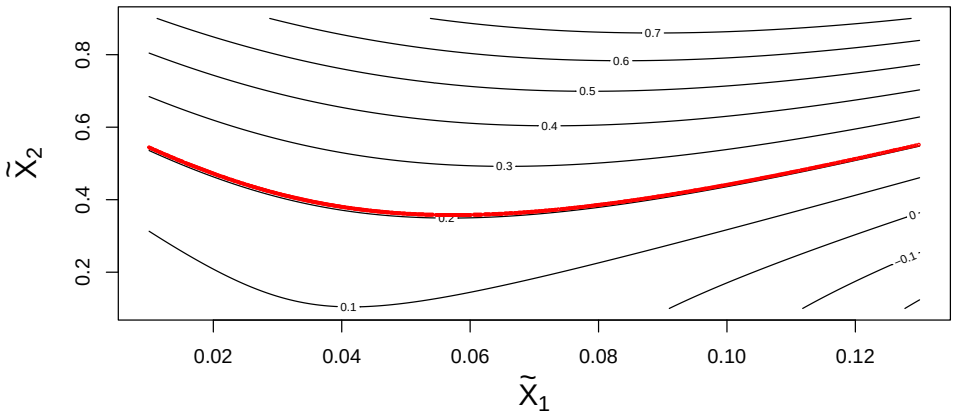

Figure 5.10: A contour plot of the simulation with two inputs and one response variable (full quadratic relationship between inputs), with the points highlighted in the plot that satisfy the equation $36 + 28\tilde{X}_1 - 11\tilde{X}_2 - 4\tilde{X}_1^2 + 6\tilde{X}_2^2 + 5\tilde{X}_1\tilde{X}_2 = 205$ (note this is before scaling, and the points in the plot have then been scaled accordingly).

Issues with the resulting MCMC output for each backward model run are observed again. The maximum upper bound of the 95% confidence interval of PSRF (for any parameter from any backward model) is 10.9 to 3 significant figures. Moreover, the median upper bound is 1.3, suggesting that a significant majority of the backward models are not converging. Appropriate adjustments are made for the MCMC tuning parameters, which

are significant, yet still lead to a lack of convergence. An adaptation length of 1500 samples, a burn-in length of 25000 and storing 20000 samples from each backward model chain, produce a maximum upper bound (for the 95% confidence intervals of PSRF) of 3.34, a median upper bound of 1.06, and a 75% quantile of 1.16. This indicates that a significant proportion of backward models are still not converging.

Given the main desire for recovering the possible input vectors highlighted in red in Figure 5.10, the prior distribution for $\boldsymbol{X}_b^*$ is adjusted from a uniform box over $[0,1]\times[0,1]$ to $[0.01, 0.13]\times[0.1, 0.9]$, corresponding to the existing range of true values for the input variables. The Latin hypercube sampling for the initialisation of $\boldsymbol{X}_b^*$ is also adjusted as above.

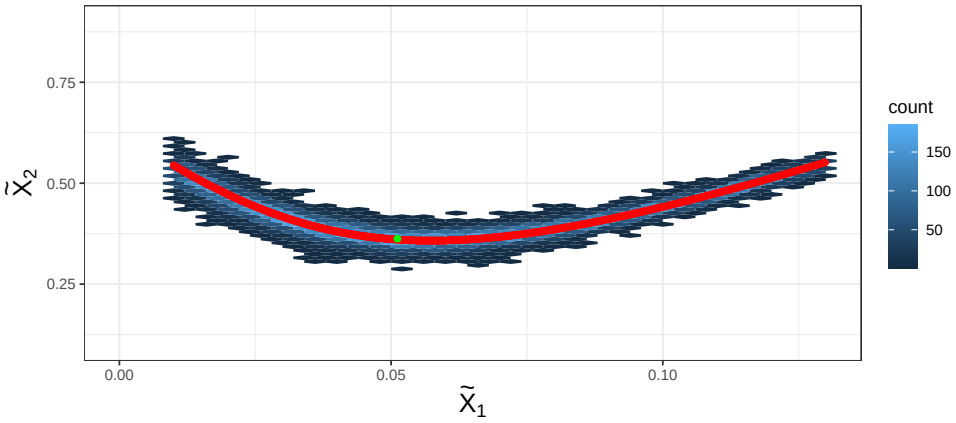

Figure 5.11: Joint posterior density of $\boldsymbol{X}^*$ given a desired response value of $Y^* = 0.205$ (blue region), with the red curve of input vectors identified in Figure 5.10 as vectors that produce the desired response superimposed. This corresponds to the simulation with one response variable and two explanatory variables with a full quadratic relationship between the explanatory variables, and the backward model algorithm has been fitted with the prior distributions $X_{1,b}^* \sim \mathrm{U}(0.01, 0.13)$ and $X_{2,b}^* \sim \mathrm{U}(0.1, 0.9)$, and the MCMC tuning parameters of an adaptation phase of length 500, a burn-in length of 25000 posterior samples and 5000 posterior samples stored (having taken every 10^{th} sample from 50000).

The backward model algorithm is fitted again with the tuning parameters (for each backward model iteration) initially used, i.e., an adaptation length of 500 samples, a burn-in length of 15000 samples, and 2000 posterior samples stored from 20000 samples, taking every 10^{th} sample. Comparing the convergence statistics with the analogous case for the uniform box prior

over $[0,1] \times [0,1]$ over $\boldsymbol{X}^*$ shows a clear improvement—the maximum upper bound for the 95% confidence interval of PSRF is 1.70, compared with 10.9 in the analogous case. The 75% quantile here is 1.03, indicating that a large majority of the backward models have converged. Further inspection of the upper bounds indicates that around 96.5% of the backward models have converged. The backward model algorithm is fitted with further increases in the MCMC tuning parameters, with the hope that 99% of the backward models converge.

Increasing the burn-in period to 25000 samples, and storing 5000 posterior samples from 50000 (taking every 10^{th} sample) and fitting the backward model algorithm produces a maximum upper bound of 1.14, a 75% quantile of 1.01 and at least 99% of the backward models converging to the posterior distribution (99% quantile of 1.08). While the effective sample size estimates have not been discussed for the previous model fits (with more of an emphasis on achieving convergence), the minimum effective sample size estimate from any backward model, any chain, and any parameter is 135.0, which is sufficiently large and indicates there is not an issue with autocorrelation between the posterior samples in each backward model.

The complete posterior distribution of $\boldsymbol{X}^*$ is compiled, taking 10 random posterior samples from each $\boldsymbol{X}_b^*$, and the joint posterior density of $\boldsymbol{X}^*$ is given in Figure 5.11 in the form of a hexibin chart. The highest-density region of values looks to surround the joint posterior mode of $\boldsymbol{X}^*_{\text{mode}} = (0.0512, 0.363)$, highlighted as the green point in Figure 5.11.

The posterior mode $\boldsymbol{X}^*_{\text{mode}} = (0.0512, 0.363)$ is now tested using the forward model posterior distribution, giving an indication of the values that could be expected for the response variable if the posterior mode of $\boldsymbol{X}^*_{\text{mode}}$ were to be observed as an input vector. This is presented by the density of predictions of the response variable given in Figure 5.12, with the lines of the desired response (red) and the mode of the predictions (green) also plotted—the proximity between these two values is very promising.

5.1.5 Multivariate regression models

The backward modelling is applied here to regression models with multiple response variables (multivariate regression), which introduces the challenge of having to optimise the input variables to be able to produce desired response values for multiple response variables simultaneously.

In many cases of backward modelling applied to multivariate regression models, there are subspaces in the space $\mathbb{R}^d$ (for d response variables) where there are no possible combinations of the explanatory variables that can

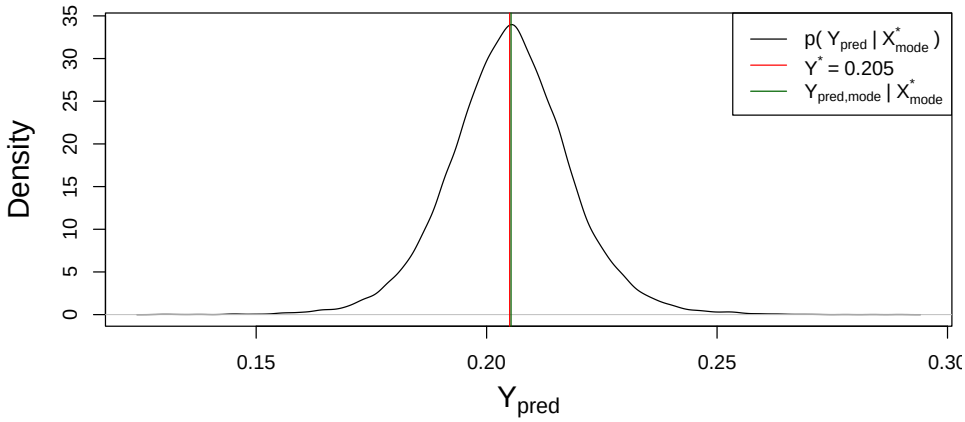

Figure 5.12: Probability density of $Y_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode}}$ (black line), where $\boldsymbol{X}^*_{\text{mode}} = (0.0512, 0.363)$ is the posterior mode of $\boldsymbol{X}^*$ given the desired response value of $Y^* = 0.205$. The green line represents the value 0.204, the mode of the distribution $Y^*|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode}}$ and the solid red line corresponds to the desired response value $Y^* = 0.205$.

produce some desired response value $\boldsymbol{Y}^* \in \mathbb{R}^d$. In other cases, it is possible to find firstly a combination of the explanatory variables which can produce some desired response value Y_1^* for say the first response variable, and secondly another combination of the explanatory variables which can produce some desired response value Y_2^* for say the second response variable, but these combinations for each desired response value may not coincide. This point reiterates the advantage of considering the joint behaviour between the response variables and the explanatory variables.

This section considers three possible desired response vectors, for the forward model estimated in Section 4.2.6, with two response variables and three input variables. These three desired response vectors are determined intentionally, having identified the subspace in $\mathbb{R}^2$ where possible vectors of the bivariate response can be produced by the underlying model. This subspace of possible response vectors is highlighted in Figure 5.13.

The first desired response vector considered is the point given by $\boldsymbol{Y}_1^* = (Y_{1,1}^*, Y_{2,1}^*) = (0.48, 0.24)$, located within a high-density region of the possible response vectors. In this case, the backward model should be able to identify a subspace of values for the three explanatory variables which is able to produce the desired response with a high level of accuracy. Given the underlying relationship in the model is known, this subspace can be determined before fitting the backward model, which provides a check as to

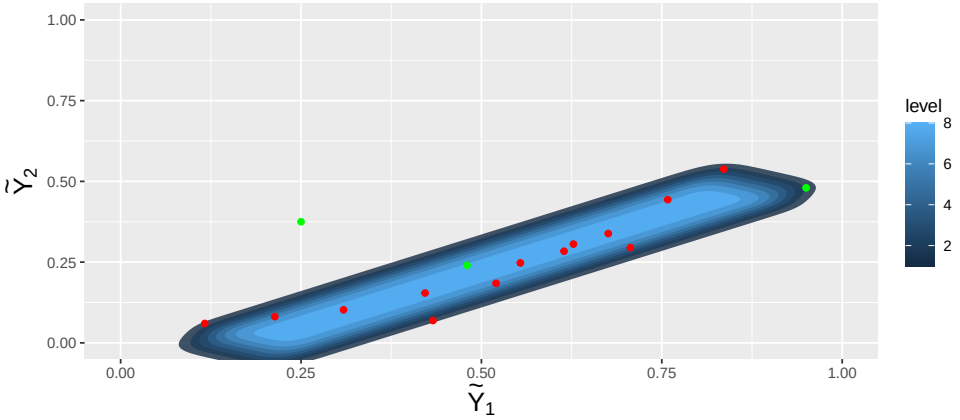

Figure 5.13: Joint density of possible response vectors that can be produced given the forward model simulation set up in Section 4.2.6. The red points correspond to the 'chosen' true values of the response variables in the forward modelling, and the green points located at (0.25,0.375), (0.48,0.24) and (0.95,0.48) are the possible desired response vectors to be explored.

.

whether the backward model is identifying appropriate vectors for the corresponding explanatory variable vector $\boldsymbol{X}^* = (X_1^*, X_2^*, X_3^*)$. This is carried out by producing a sequence of 50 values (evenly spaced) for each explanatory variable, ranging through their respective range of values (based on their minimum and maximum chosen true values)—then, for each combination of values from the three sequences, the multivariate response value is produced, giving $50^3 = 125000$ vectors of the bivariate response. Those response vectors that are within 1% of the desired response vector are then identified, and their corresponding combination of input vectors are recorded and produce the subspace of corresponding input vectors. This subspace is provided by the green points in the three plots in Figure 5.14, with the left plot corresponding to vectors for the first and second inputs, the middle plot corresponding to vectors for the first and third inputs, and the right plot corresponding to vectors for the second and third inputs. Possible input vectors that can produce the desired value for *first* response variable are given by the blue points, and those that can produce the desired value for the *second* response variable are given by the red points.

It is clear from Figure 5.14 that the third input variable has the most influence on this desired response vector (and any desired response, given the underlying model)—its possible value corresponding to the desired response

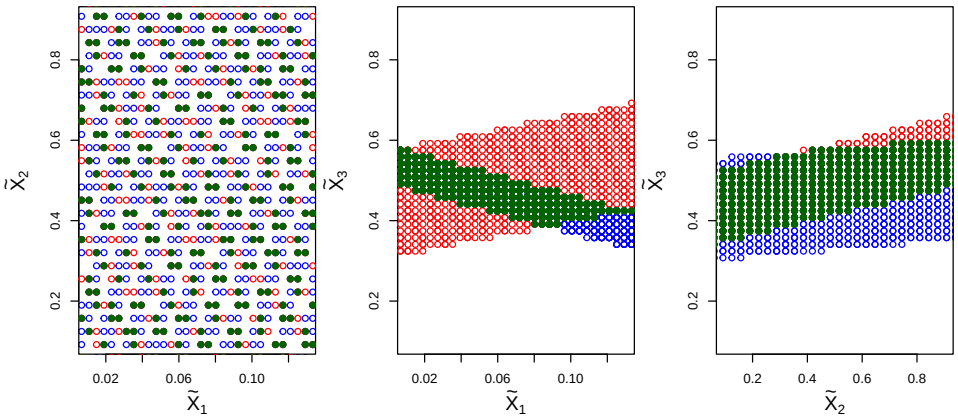

Figure 5.14: Plots of the three input variables demonstrating points which can either produce the desired response value $Y^*_{1,1} = 0.48$ (blue points), the desired response value $Y^*_{2,1} = 0.24$ (red points), or both desired response values $(0.48, 0.24)$ simultaneously (green points). The left plot provides vectors for (X^*_1, X^*_2), the middle plot for (X^*_1, X^*_3) and the right plot for (X^*_2, X^*_3).

vector $\boldsymbol{Y}^*_1$ looks to be restricted to between roughly 0.35 and 0.6. Any combination of values for the first two input variables look plausible here.

The first backward model to be fitted, with $\boldsymbol{Y}^* = (0.48, 0.24)$, is

$$\begin{pmatrix} 0.48 \\ 0.24 \end{pmatrix} = \begin{pmatrix} \beta_{0,1,b} & \beta_{1,1,b} & \beta_{2,1,b} & \beta_{3,1,b} \\ \beta_{0,2,b} & \beta_{1,2,b} & \beta_{2,2,b} & \beta_{3,2,b} \end{pmatrix} \begin{pmatrix} 1 \\ X^*_{1,b} \\ X^*_{2,b} \\ X^*_{3,b} \end{pmatrix} + \begin{pmatrix} \epsilon^*_{1,b} \\ \epsilon^*_{2,b} \end{pmatrix},$$

with the subscript b denoting joint posterior sample b of the parameters from the forward modelling posterior. The prior distribution for the corresponding input vector $\boldsymbol{X}^*$ is a uniform box over the space $[0, 1] \times [0, 1] \times [0, 1]$. A Latin hypercube sampling is carried out (with 3 variables and $S_{\text{back}} = 2000$ samples), to determine initial vectors for each of the backward model runs (a random sample (without replacement) from the Latin hypercube sampling is assigned to each backward model run). Having fitted the model, the MCMC output is investigated for convergence and mixing. The PSRF estimates give a maximum value for the upper bound of the 95% confidence interval of 7.18 (to 3 significant figures), indicating that at least one model has not converged, and is not close to be deemed to have converged. A look at quantile summary statistics over all upper bounds of PSRF shows that at least 75% of the models are deemed to have converged. It is clear from

previous simulations that the possible routes to explore in order to improve convergence are increasing the burn-in period for the MCMC for each backward model, collecting further posterior samples for the MCMC output, or restricting the prior distribution on $\boldsymbol{X}^*$. The restriction on the prior distribution is only significant for the first explanatory variable, as the range of its chosen true values is from 0.01 to 0.13, while the prior distribution on its corresponding input is $\mathrm{U}(0,1)$. While it is plausibly of interest in a backward modelling scenario to explore regions of the input space that are *not* already explored in the forward modelling, in this backward modelling case, the interest is to recover the values that have been highlighted in Figure 5.14, which are produced based on the range of chosen true values from the forward modelling. For this reason, the prior distribution for the first corresponding input X_1^* is altered, from U(0,1) to U(0.01,0.13). The prior distributions of X_2^* and X_3^* are similarly adjusted from U(0,1) to U(0.1,0.9) in both cases (due to their respective ranges of chosen true values existing in the forward modelling).

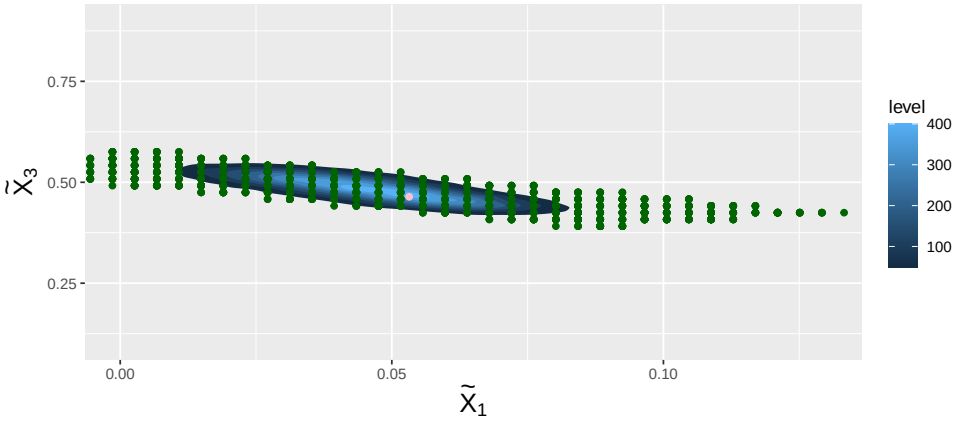

Figure 5.15: Joint marginal posterior density of $(\boldsymbol{X}_1^*, \boldsymbol{X}_3^*)$, with the possible vectors of these two explanatory variables which are able to produce the desired response vector $\boldsymbol{Y}_1^* = (0.48, 0.24)$ (given the underlying simulation on which the model is based) also included (dark green points). The pink point indicates the elements of joint posterior mode of $\boldsymbol{X}^*$ corresponding to these two variables.

Rerunning the backward modelling algorithm with the same desired response vector of $\boldsymbol{Y}_1^* = (0.48, 0.24)$ along with this adjustment to the prior distribution, provides sufficient improvement of the convergence of the backward modelling. A maximum upper bound of the PSRF estimate for any

backward model is 1.08 to 3 significant figures, indicating that all backward models in the algorithm have converged. Moreover, the minimum estimate of effective sample size for any dimension of the corresponding input vector $\boldsymbol{X}^*$ from any backward model (and any parallel chain) is 210 (to 3 significant figures), indicating that sufficient levels of mixing occur in all backward models. The MCMC output is now considered to be the posterior distribution and is examined further below.

As explained above, the posterior distribution of $\boldsymbol{X}^*$ should align with those vectors that are identified in Figure 5.14; alternatively, the posterior density could be more concentrated within the identified subspace—in other words, the backward model may not identify the entire subspace in its posterior density, but it may have high confidence in a region within the subspace. The posterior density is presented in a similar way to the presentation of the region of green points in Figure 5.14—that is, the joint marginal posterior density of $(\boldsymbol{X}_1^*, \boldsymbol{X}_3^*)$ is provided in Figure 5.15 and the joint posterior density of $(\boldsymbol{X}_2^*, \boldsymbol{X}_3^*)$ is presented in Figure 5.16. Note that the joint marginal posterior density of $(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*)$ is omitted here, as the two plots mentioned here provide all of the information needed. It is clear from both plots that the joint marginal posterior densities are a good representation of the input vectors that have been identified as being able to produce the desired response vector $\boldsymbol{Y}_1^* = (0.48, 0.24)$ given the underlying model. In both Figure 5.15 and Figure 5.16 the joint marginal posterior densities do not capture every possible input vector, and instead are more concentrated on a subspace of these vectors. Note that, in both plots, the joint posterior mode (pink point) is the joint posterior mode of $\boldsymbol{X}^* = (X_1^*, X_2^*, X_3^*)$, with the respective elements extracted from the vector and plotted; this is as opposed to identifying the joint marginal posterior mode of the explanatory variables considered in each plot.

Given the plots of the joint marginal posterior density of (X_1^*, X_3^*) and (X_2^*, X_3^*), it is highly likely that the joint posterior mode that has been identified, given by $\boldsymbol{X}^*_{\text{mode},1} = (0.0532, 0.332, 0.464)$, is a suitable candidate vector for producing the desired response vector (note that subscript '1' in the mode indicates that this mode corresponds to the first desired response vector of $\boldsymbol{Y}^* = (0.48, 0.24)$).

Without knowledge of the underlying relationship of the simulation, this would not be possible to know, so the additional step of producing values of the response variable given this posterior mode is carried out here. In this case, the distribution of predictions of the response variables is 2-dimensional, and so there is uncertainty in both response variables. It is possible that the uncertainty is correlated, which is why the forward and

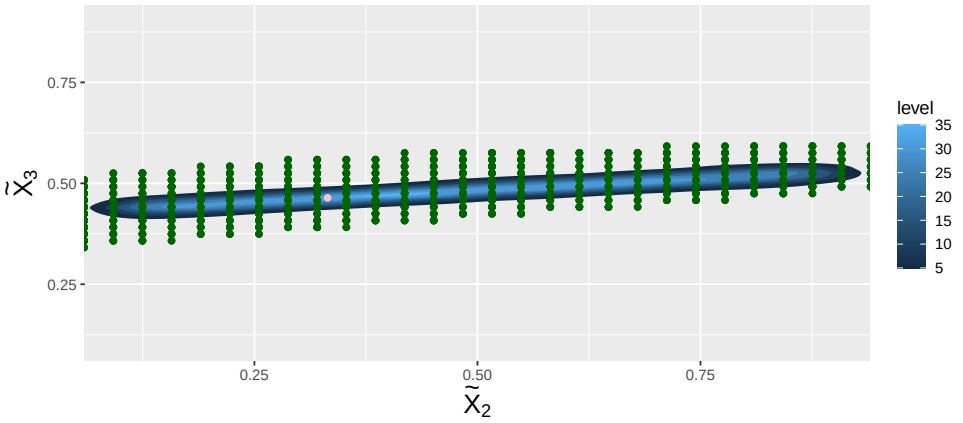

Figure 5.16: Joint marginal posterior density of $(\boldsymbol{X}_2^*, \boldsymbol{X}_3^*)$, with the possible vectors of these two explanatory variables which are able to produce the desired response vector $\boldsymbol{Y}_1^* = (0.48, 0.24)$ (given the underlying simulation on which the model is based) also included (dark green points). The pink point indicates the elements of joint posterior mode of $\boldsymbol{X}^*$ corresponding to these two variables.

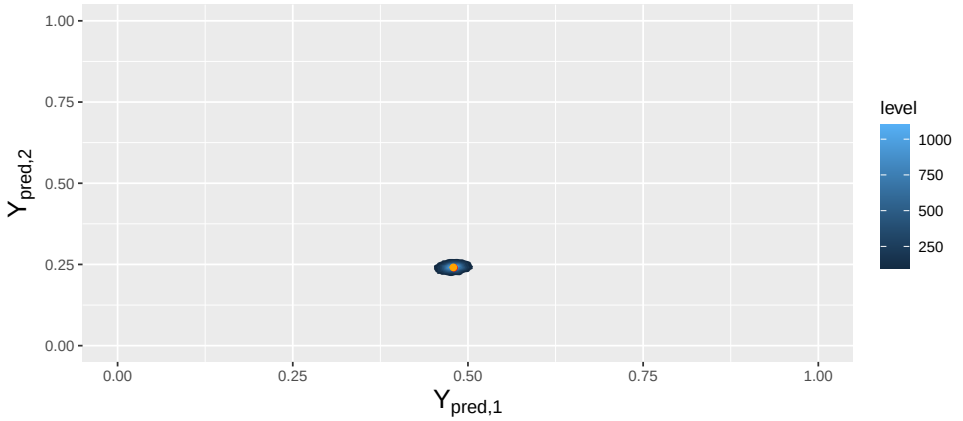

Figure 5.17: The joint density $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode},1})$, where the backward model has been fitted with the desired response vector $\boldsymbol{Y}^* = (0.48, 0.24)$ (red point). The orange point indicates the mode of $\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode},1}$.

backward models account for this, and this will also be accounted for in the plot of the 2-dimensional density of predictions of the response vector. This distribution of predictions is found using the forward model posterior distribution, given by

$$\begin{pmatrix} Y_{1,\text{pred},s} \\ Y_{2,\text{pred},s} \end{pmatrix} = \begin{pmatrix} \beta_{0,1,s} & \beta_{1,1,s} & \beta_{2,1,s} & \beta_{3,1,s} \\ \beta_{0,2,s} & \beta_{1,2,s} & \beta_{2,2,s} & \beta_{3,2,s} \end{pmatrix} \begin{pmatrix} 1 \\ X^*_{1,\text{mode},1} \\ X^*_{2,\text{mode},1} \\ X^*_{3,\text{mode},1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,s} \\ \epsilon_{2,s} \end{pmatrix}, \tag{5.1.5.1}$$

where $\boldsymbol{X}^*_{\text{mode},1} = (X^*_{1,\text{mode},1}, X^*_{2,\text{mode},1}, X^*_{3,\text{mode},1}) = (0.0532, 0.332, 0.464)$. The joint density of the predictions of the response variable given the posterior mode, denoted by $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode},1})$, is plotted in Figure 5.17, with the mode of the predictions highlighted as the orange point, given by $(0.479, 0.241)$. The location and uncertainty around the mode of $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode},1})$ indicate that the backward model has suggested a candidate vector that is highly likely to produce vectors of the response that are very close to the desired response. Given the posterior distribution from the forward modelling, which had clearly captured the underlying relationship chosen for the model, it is not a surprise that this backward model is able to recover a suitable candidate vector for $\boldsymbol{X}^*$, in this case of a desired response vector which is likely to occur for several combinations of the input variables.

A more challenging task for this backward model is the desired response vector $\boldsymbol{Y}^*_2 = (0.95, 0.48)$, which is located on the far right of the plot in Figure 5.13. As can be seen in this figure, this desired response vector is much less likely to occur given any combination of the input variables (given those input variables taking values in the range of values already explored in the forward modelling), suggesting the subspace of input vectors will be more restricted than in the previous case. A visualisation of this is provided in Figure 5.18 (i.e., given $Y^*_{1,2} = 0.95$ and $Y^*_{2,2} = 0.48$), analogous to the plot in Figure 5.14. This increase in difficulty of achieving the desired response vector has lead to more of a restriction of the first input variable (see the middle plot of Figure 5.14), and the possible values for the third input variable are also heavily restricted.

The backward model for the case of $\boldsymbol{Y}^* = (0.95, 0.48)$ is now fitted, with the prior distributions of $X^*_1 \sim \text{U}(0.01, 0.13)$, $X^*_2 \sim \text{U}(0.1, 0.9)$ and $X^*_3 \sim \text{U}(0.1, 0.9)$. The upper bounds of PSRF estimates for each backward model are checked to confirm convergence to the posterior distribution—the largest such upper bound is 1.03 to 3 significant figures, indicating convergence for each backward model. With the minimum effective sample size estimate

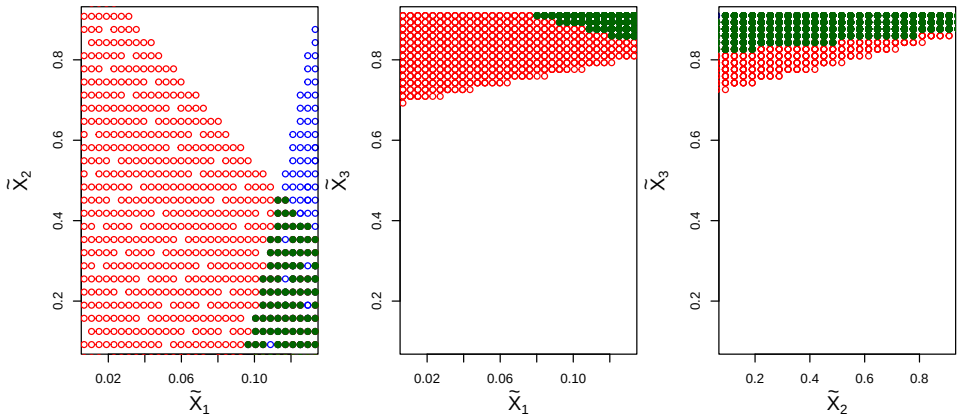

Figure 5.18: Plots of the three input variables demonstrating points which can either produce the desired response value $Y^*_{1,2} = 0.95$ (blue points), the desired response value $Y^*_{2,2} = 0.48$ (red points), or both desired response values $(0.95, 0.48)$ simultaneously (green points). The left plot provides vectors for (X^*_1, X^*_2), the middle plot for (X^*_1, X^*_3) and the right plot for (X^*_2, X^*_3).

for any variable from and backward model run being 899.6, each backward model has also sufficiently mixed.

The complete posterior distribution of $\boldsymbol{X}^*$ is now considered, specifically with two plots of the joint marginal posterior densities of (X^*_1, X^*_3) and (X^*_2, X^*_3), provided in Figure 5.19 and 5.20 respectively. Both plots have the vectors indicated in Figure 5.18 superimposed, to be able to compare the posterior density of $\boldsymbol{X}^*$ with vectors discovered from the underlying simulation. This comparison indicates the backward model has been able to identify suitable values for each of the three input variables. The full range of possible values of X^*_2 has not been explored in the backward model, but it is clear from both Figures 5.19 and 5.20 that the joint posterior mode of $\boldsymbol{X}^*$, given by $\boldsymbol{X}^*_{\text{mode},2} = (0.129, 0.268, 0.896)$ lies within the possible range of values for each of the respective variables (subscript '2' in the posterior mode references the second desired response vector, $\boldsymbol{Y}^*_2 = (0.95, 0.48)$).

This joint posterior mode of $\boldsymbol{X}^*_{\text{mode},2} = (0.129, 0.268, 0.896)$ is tested using the forward model posterior distribution, to discover whether the mode is a suitable candidate vector for the corresponding desired response vector. A joint distribution of predictions for the two response variables is provided, given by $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{X}^*_{\text{mode},2})$, estimated analogously to Equation 5.1.5.1, replacing the posterior mode $\boldsymbol{X}^*_{\text{mode},1}$ with the posterior mode $\boldsymbol{X}^*_{\text{mode},2}$. The density of this joint distribution of predictions is provided in Figure 5.21.

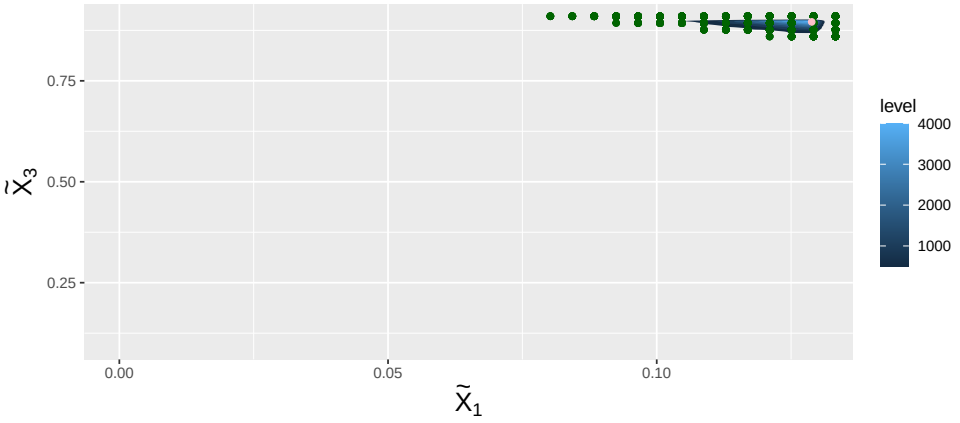

Figure 5.19: Joint marginal posterior density of $(\boldsymbol{X}_1^*, \boldsymbol{X}_3^*)$, with the possible vectors of these two explanatory variables which are able to produce the desired response vector $\boldsymbol{Y}_1^* = (0.95, 0.48)$ (given the underlying simulation on which the model is based) also included (dark green points). The light green point indicates the elements of joint posterior mode of $\boldsymbol{X}^*$ corresponding to these two variables.

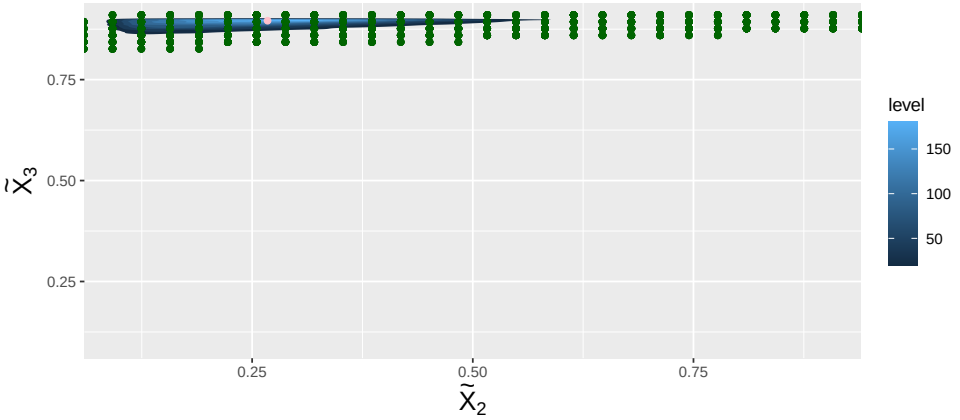

Figure 5.20: Joint marginal posterior density of $(\boldsymbol{X}_2^*, \boldsymbol{X}_3^*)$, with the possible vectors of these two explanatory variables which are able to produce the desired response vector $\boldsymbol{Y}_1^* = (0.95, 0.48)$ (given the underlying simulation on which the model is based) also included (dark green points). The light green point indicates the elements of joint posterior mode of $\boldsymbol{X}^*$ corresponding to these two variables.

The orange point indicates the mode of the joint distribution of predictions, given by $(0.954, 0.480)$ and the red point indicates the desired response vector $(0.95, 0.48)$. Clearly, the joint distribution of predictions of the response variables is centred around the desired response vector, as desired. There appears to be slightly more variation in the predictions of the response variable, comparing with those predictions from the previous case, displayed in Figure 5.17. The likely reason for this is that the model uncertainty becomes larger the further away from the mean values for the variables, which occurs here intentionally to investigate how the models performs in such cases.

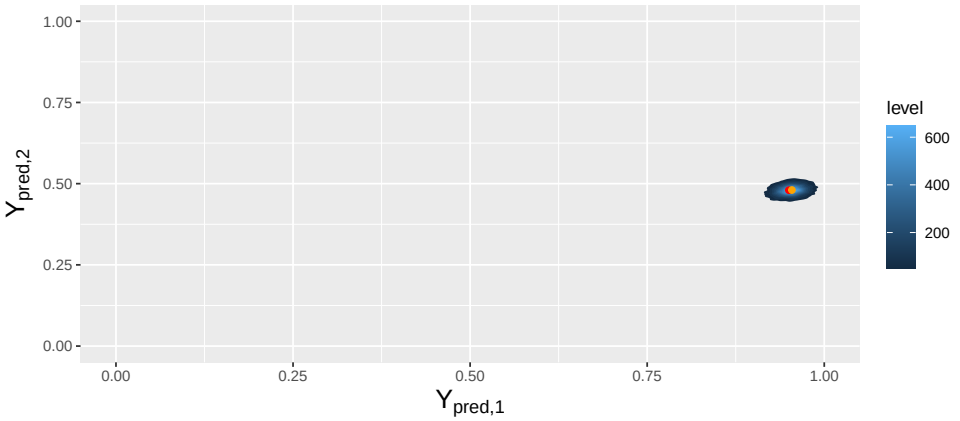

Figure 5.21: The joint density $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode},2})$, where the backward model has been fitted with the desired response vector $\boldsymbol{Y}^* = (0.95, 0.48)$ (red point). The orange point indicates the mode of $\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode},2}$.

The final case of backward modelling considered for this simulation is with the desired response vector $\boldsymbol{Y}^*_3 = (0.25, 0.375)$. This desired response vector identified in Figure 5.13 is not in close proximity to those possible desired response vectors that could be produced given the underlying simulation and the range of values of the three input variables. For this reason, the interest is to see what occurs in the backward modelling in a scenario where a corresponding vector of input values cannot be produced for the desired response vector.

To confirm this, the plot in Figure 5.22 shows possible combinations of the three input variables that can produce the desired response values $Y^*_{1,3} = 0.25$ and $Y^*_{2,3} = 0.375$ separately (the blue points and the red points respectively). In the previous two versions of this plot for the previous desired response vectors, the green points indicated the combinations of the three input variables that could produce both desired response variables

simultaneously; in Figure 5.22, these only occur for combinations of the first two input variables, for the desired response vector $\boldsymbol{Y}_3^* = (0.25, 0.375)$. For these green points, the values of the third input variable are not the same. This is evident based on the middle and right plots, where there is no crossover between the red points and blue points.

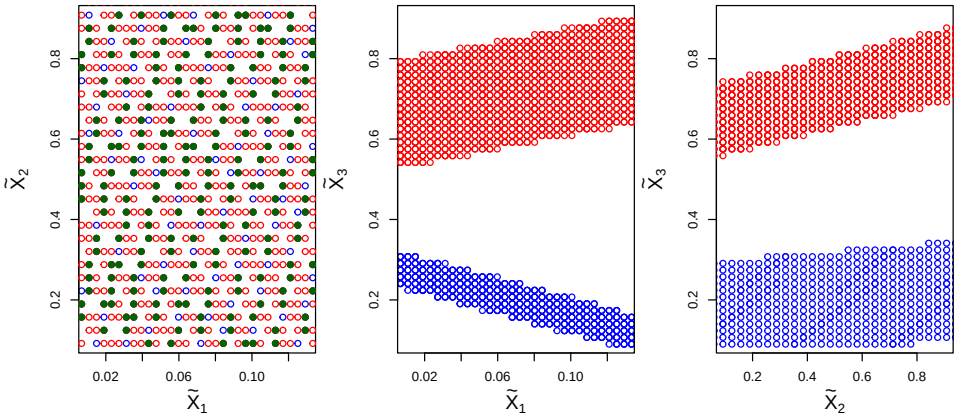

Figure 5.22: Plots of the three input variables demonstrating points which can either produce the desired response value $Y_{1,3}^* = 0.25$ (blue points) or the desired response value $Y_{2,3}^* = 0.375$ (red points). The green points in the first plot indicate that there are combinations of the first two input variables which can produce both desired response values simultaneously, but these are for different values of X_3^*, as indicated in the middle and right plots. The left plot provides vectors for (X_1^*, X_2^*), the middle plot for (X_1^*, X_3^*) and the right plot for (X_2^*, X_3^*).

The backward modelling algorithm is now run for the desired response vector $\boldsymbol{Y}_3^* = (0.25, 0.375)$. The largest upper bound of PSRF estimates from any of the backward models is 1.04 to 3 significant figures, indicating that each backward model has converged to the posterior distribution. There is also an insignificant level of autocorrelation between the samples for all the backward models; the minimum estimate of effective sample size from any backward model and any parallel chain is 695, to 3 significant figures. The MCMC output is now treated as the posterior distribution.

A plot of the joint marginal posterior density $p((X_1^*, X_3^*)|\boldsymbol{Y}_3^*)$ is provided in Figure 5.23. Given there are no dark green points indicating the combinations of the input variables that can produce both desired response values of $\boldsymbol{Y}_3^*$ simultaneously, the blue and red points from the middle plot of Figure 5.22 are provided, which indicate those combinations that can produce

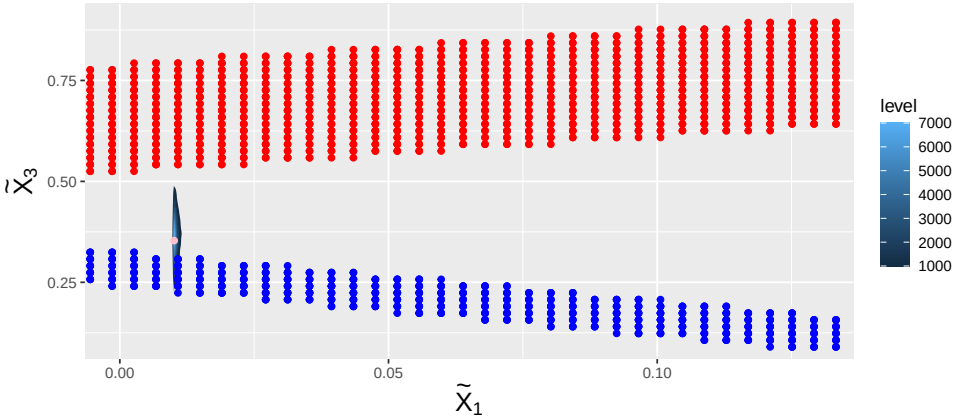

Figure 5.23: Joint marginal posterior density of (X_1^*, X_3^*) given $\boldsymbol{Y}_3^*$, with the possible vectors of these two explanatory variables which are able to produce the desired response value $Y_{1,3}^* = 0.25$ (blue points) or the desired response value $Y_{2,3}^* = 0.375$ (red points). The pink point indicates the elements of joint posterior mode of $\boldsymbol{X}^*$ corresponding to these two variables.

$Y_{1,3}^*$ and $Y_{2,3}^*$ respectively. Interestingly, there seems to be very little variation around the values of the first input variable that are captured in the backward modelling, particularly in comparison with the variation of values for the third input variable captured in the backward modelling. There appears to be some tendency to head towards those blue points, which indicate vectors that can produce the desired value for the first response, $Y_{1,3}^* = 0.25$. This is clarified later when considering the joint distribution of the predictions of the response variables given the posterior mode $\boldsymbol{X}_{\text{mode},3}^* = (0.0101, 0.103, 0.353)$.

The joint marginal posterior density of $p((X_2^*, X_3^*)|\boldsymbol{Y}_3^*)$ is plotted in Figure 5.24. As with the previous figure, the blue and red points identified in Figure 5.22, corresponding to input vectors which can produce the desired response $Y_{1,3}^*$ or $Y_{2,3}^*$ respectively, are also plotted. As in the previous case, it appears that there is some tendency for the backward model to try to produce the desired response for the first response variable, since the posterior density contains values within the region of the blue points.

The joint distribution of predictions of the response variables given the posterior mode $\boldsymbol{X}_{\text{mode},3}^*|\boldsymbol{Y}_3^*$ is now considered. The vectors that build this distribution are computed analogously to those in Equation 5.1.5.1, replacing $\boldsymbol{X}_{\text{mode},1}^*$ with the vector $\boldsymbol{X}_{\text{mode},3}^*$. This joint density of predictions is presented in Figure 5.21, with the orange point indicating the mode of the

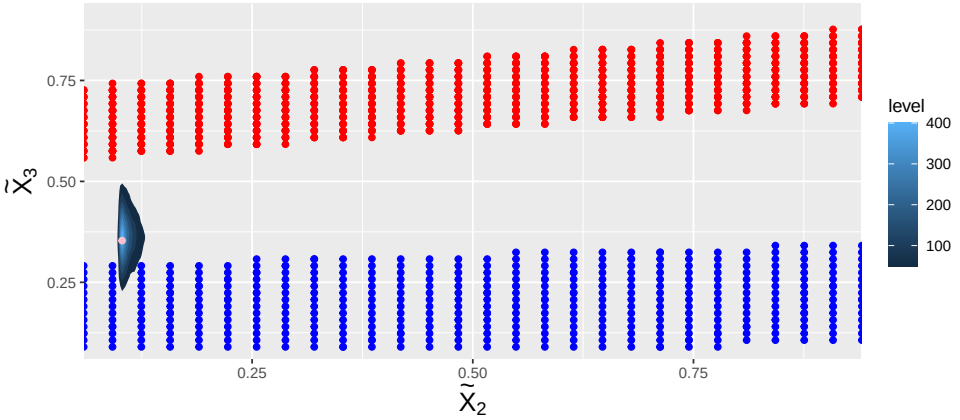

Figure 5.24: Joint marginal posterior density of (X_2^*, X_3^*) given $\boldsymbol{Y}_3^*$, with the possible vectors of these two explanatory variables which are able to produce the desired response value $Y_{1,3}^* = 0.25$ (blue points) or the desired response value $Y_{2,3}^* = 0.375$ (red points). The pink point indicates the elements of joint posterior mode of $\boldsymbol{X}^*$ corresponding to these two variables.

predictions, and the red point indicating the desired response vector $\boldsymbol{Y}_3^*$. It is unsurprising that the predictions of the response variable are not in close proximity with the desired response vector, given the underlying simulation and the reason for choosing this desired response. Interestingly, if this plot is compared with the density of possible response vectors that could be produced in the simulation (see Figure 5.13), it would appear that this joint distribution of predictions given the posterior mode has minimised the distance between the desired response vector $\boldsymbol{Y}_3^*$ and the subspace of possible response vectors from Figure 5.13. This is evidence for the backward modelling doing the best it possibly can—it has been set up to not be able to obtain the desired response vector (in other words, given the range of values of the prior density of $\boldsymbol{X}^*$, it is impossible to produce $\boldsymbol{Y}_3^*$ given the simulation), so the backward model has chosen a candidate vector for $\boldsymbol{X}_3^*$ which produces a vector response that minimises the distance between the $\boldsymbol{Y}_3^*$ and the subspace of possible response vectors.

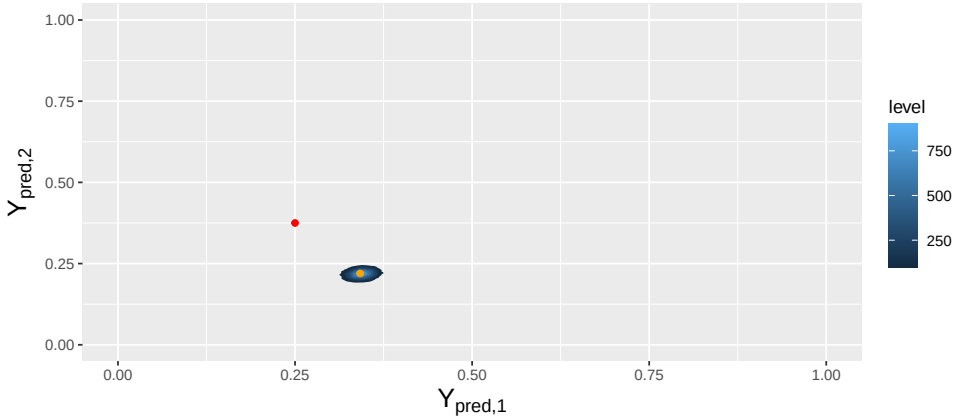

Figure 5.25: The joint density $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{X}^*_{\text{mode},3})$, where the backward model has been fitted with the desired response vector $\boldsymbol{Y}^* = (0.25, 0.375)$ (red point). The orange point indicates the mode of $\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode},3}$.

5.2 Errors-in-variables Gaussian processes on simulated data

In this section, the backward modelling algorithm is explored with simulated data and forward modelling fitted with an EIV GP. That is, the hyperparameters $\boldsymbol{\phi}$ (including the true values of the response and explanatory variables) for an EIV GP prior have been estimated by the joint posterior distribution $p(\boldsymbol{\phi}|Y, X)$, and can be used to predict the response variable(s) at particular true values or true vectors of the input variable(s) with the EIV GP posterior. As described in Section 3.6.7, the EIV GP posterior is utilised in the backward modelling in order to optimise the input variable(s) to be able to produce a desired response value or vector. In this section, the backward modelling is applied to single-output, single-input EIV GPs, as well as EIV MOGPs. The following subsection considers the single-output, single-input EIV GP, whose forward modelling was demonstrated in Sections 4.3.1, 4.3.2 and 4.3.3.

5.2.1 Single-output, single-input EIV GP

The backward modelling process is briefly summarised below for the case of one response variable and one explanatory variable.

Assume a desired response value $Y^* = y^*$ has been provided by some expert; the aim with the backward modelling is to optimise the corresponding

input variable X^* in order to produce this desired response $Y^* = y^*$. Given the posterior distribution of the hyperparameters $p(\boldsymbol{\phi}|Y, X)$, the EIV GP posterior, with consideration of the desired response variable Y^*, is built from the joint distribution

$$\begin{pmatrix} \tilde{\boldsymbol{Y}}_b \\ Y^* \end{pmatrix} \sim \mathrm{N}_{n_g+1}\left(\begin{pmatrix} m_b(\tilde{\boldsymbol{X}}_b) \\ m_b(X_b^*) \end{pmatrix}, \begin{pmatrix} V_b(\tilde{\boldsymbol{X}}_b, \tilde{\boldsymbol{X}}_b) & K(\tilde{\boldsymbol{X}}_b, X_b^*) \\ K(X_b^*, \tilde{\boldsymbol{X}}_b) & v_b(X_b^*, X_b^*) \end{pmatrix} \right),$$

leading to the conditional distribution (EIV GP posterior) of

$$Y^*|X_b^*, \boldsymbol{\phi}_b \sim \mathrm{N}(m_b^*, v_b^*).$$

Note again that the subscript b refers to joint posterior sample b from the hyperparameter posterior distribution $p(\boldsymbol{\phi}|Y, X)$, and that the mean function $m(\tilde{X}_i) = \alpha$ or the covariance kernel k, that inform the vectors and matrices in the joint distribution, are in turn evaluated given the joint posterior sample b in each case. For example, the mean $m_b(X_b^*) = \alpha_b$. The mean and variance m_b^* and v_b^* of the EIV GP posterior are given by

$$m_b^* = m_b(X_b^*) + K_b(X_b^*, \tilde{\boldsymbol{X}}_b)V_b(\tilde{\boldsymbol{X}}_b, \tilde{\boldsymbol{X}}_b)^{-1}(\tilde{\boldsymbol{Y}}_b - m_b(\tilde{\boldsymbol{X}}_b))$$

and

$$v_b^* = v_b(X_b^*, X_b^*) - K_b(X_b^*, \tilde{\boldsymbol{X}}_b)V_b(\tilde{\boldsymbol{X}}_b, \tilde{\boldsymbol{X}}_b)^{-1}K_b(\tilde{\boldsymbol{X}}_b, X_b^*).$$

In the forward modelling, this EIV GP posterior is used to predict the response variable given some new, true value of the explanatory variable, but in the backward modelling, $Y^* = y^*$ is assumed to be observed, and for each joint posterior sample b, some uniform prior distribution is placed on X_b^*, and an MCMC algorithm is fitted in order to find the best values for each X_b^*, and the posterior distributions of each X_b^* are combined (taking random samples from each) to build the posterior distribution of the corresponding input variable X^*, $p(X^*|Y^*, \boldsymbol{\phi})$. For further details the reader is referred to Section 3.6.7.

The EIV GP model considered here is the single-input EIV GP simulation considered in Section 4.3.1, which has been fitted with a reparameterisation of $l = k_l\sigma_\epsilon$ and $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, with the uniform prior distributions $k_l \sim \mathrm{U}(0, 60)$ and $k_{\sigma_k} \sim \mathrm{U}(4, 200)$. The corresponding model fit is that provided in Figure 4.30. As can be seen in either this model fit or in the plot of the simulation function given in Figure 5.26 (which also includes the first desired response value $Y^* = 0.24$ and the possible input values to produce this desired response), the majority of the range of response values that have been considered in the forward modelling are able to be produced by

multiple values of the input variable. This is likely to lead to a multimodal posterior distribution for each X_b^*, and therefore an issue in the convergence of the each of these posterior distributions. As previously discussed, the results from this backward modelling analysis are used despite the lack of convergence in the posterior distributions of X_b^*, and plots of the complete posterior distribution are not likely to provide any information about how the relative probability densities of each posterior mode of X^* (but still do indicate that the backward modelling was able to capture the posterior modes).

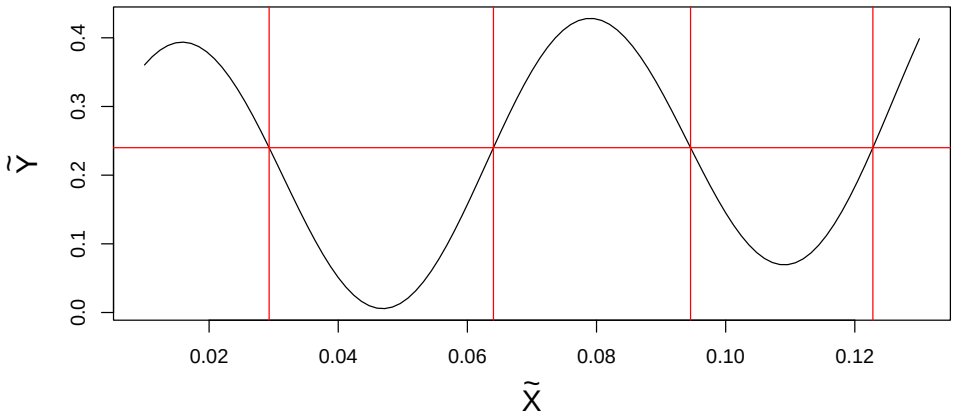

Figure 5.26: A plot of the simulation function for the single-input EIV GP considered in Section 4.3.1, along with a red horizontal line corresponding to the desired response value $Y^* = 0.24$, and four red vertical lines corresponding to the input values $X^* = 0.0293$, $X^* = 0.0640$, $X^* = 0.0946$ and $X^* = 0.123$ (all to 3 significant figures), that can produce the desired response value

The backward modelling algorithm is fitted with the uniform prior distribution $X_b^* \sim \mathrm{U}(0, 1)$, and (for each of 2 parallel chains) the MCMC tuning parameters of an adaptation length of 500 samples, a burn-in length of 15000 samples, 2000 posterior samples stored for each backward model posterior X_b^* having taken every 10^{th} sample from 20000 posterior samples (these 2000 posterior samples are used for checking mixing and convergence of the MCMC output), then 10 posterior samples are randomly selected from each X_b^* posterior distribution to build the complete posterior distribution of X^*. This is firstly fitted with $Y^* = 0.24$, and the PSRF upper bounds from each posterior distribution X_b^* show that a significant majority of the backward models are not converging in this case. The median upper bound of PSRF

over all backward models is 8.25, which is significantly larger than than cut-off of 1.1. A view into one of the backward models that has not converged according to PSRF is given in Figure 5.27, which provides the posterior density of X_b^* for each of the two parallel chains of the backward model given posterior sample $b = 1$ from the forward model. Both chains are significantly different from one another—the first chain has a lot of density at all values of the prior distribution, with a clear peak which is centred at the existing true values from the forward modelling, and the second chain has all of its density concentrated close to one of the possible input values that can produce the desired response (notably the corresponding input value is a relatively-low density value of the distribution). This is not desired behaviour—the optimal outcome would be that both chains recover all four possible input values as posterior modes.

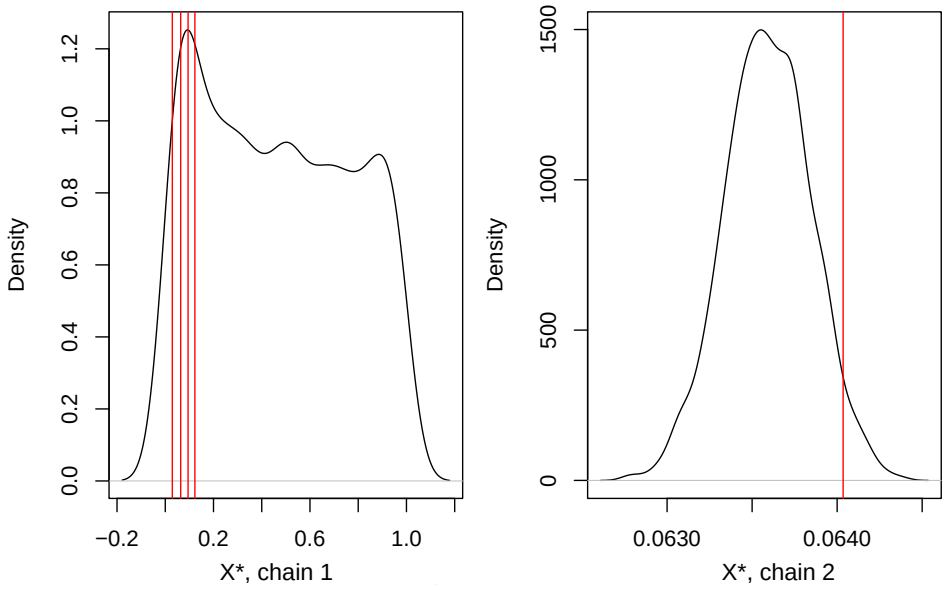

Figure 5.27: The posterior density of X_1^* (so given the forward model posterior sample $b = 1$) for its two parallel chains. It is noted that the model has not converged, with an upper bound of PSRF given by 13.3.

While acknowledging that the MCMC output for each X_b^* does not represent the posterior distribution, samples are still collected from each X_b^* and combined to build the complete 'posterior' distribution of X^* given the desired response value $Y^* = 0.24$. This allows us to confirm that the possible

input values identified in Figure 5.26 are captured as modes in the posterior distribution, noting that the probability density around these modes cannot be assessed with this output. A plot of this posterior density is found in Figure 5.28. This desired response value $Y^* = 0.24$ represents an interesting task for the backward model—this value is found by taking the mean of the 'chosen' true values for the response variable in the underlying simulation. In this particular simulation, where the small differences between the input variable correspond to potentially large differences in the response variable (given estimates of the distance-scaling parameter, see the middle plot of Figure 4.31), the mean prediction of the response variable for input values beyond what are observed in the forward modelling is close to the mean of the response variable (this is true for any GP fit, given the nonparametric behaviour, extrapolation is difficult). For this reason, the mean of the response variable being the desired response then appears to allow for density at almost all input values in the range [0,1] for X^*.

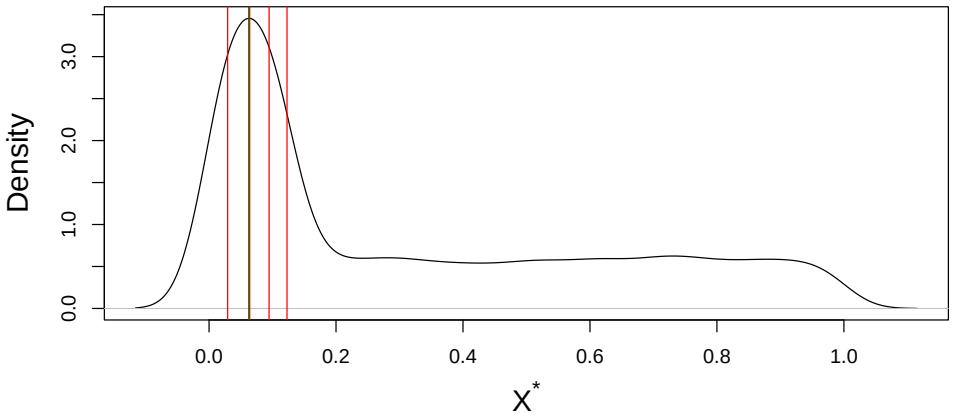

Figure 5.28: The posterior density of X^* given the desired response $Y^* = 0.24$ and the forward model posterior $\boldsymbol{\phi}$. The red lines represent the four possible input values that can produce the desired response given the underlying simulation, and the green vertical line represents the 'posterior' mode, given by 0.0626 to 3 significant figures, which is very close to the possible input value 0.0640.

The backward modelling is carried out again for $Y^* = 0.24$, with the prior distribution $X^* \sim \mathrm{U}(0.01, 0.13)$. The purpose of this is to investigate whether all four possible input values can be captured as modes in the 'posterior' distribution of X^*. The word 'posterior' is placed in quotes because it is likely that the multimodal posterior distribution will not be perfectly

captured by the modelling, since each backward model is not converging to its posterior distribution. This is confirmed by considering the upper bounds of PSRF for each of the $b = 1, \ldots, S_{\text{back}}$ backward models, with the median upper bound being 236 to 3 significant figures. As above, the samples from the MCMC output are still examined, and the 'posterior' density of X^*, having taken 10 samples from the 'posteriors' of X_b^*, is provided in Figure 5.29. Not much can be said about the relative probabilities of each mode, given the lack of convergence for the majority of backward models, but it is evident from Figure 5.29 that averaging over all of the backward models leads to finding the appropriate posterior modes for X^*.

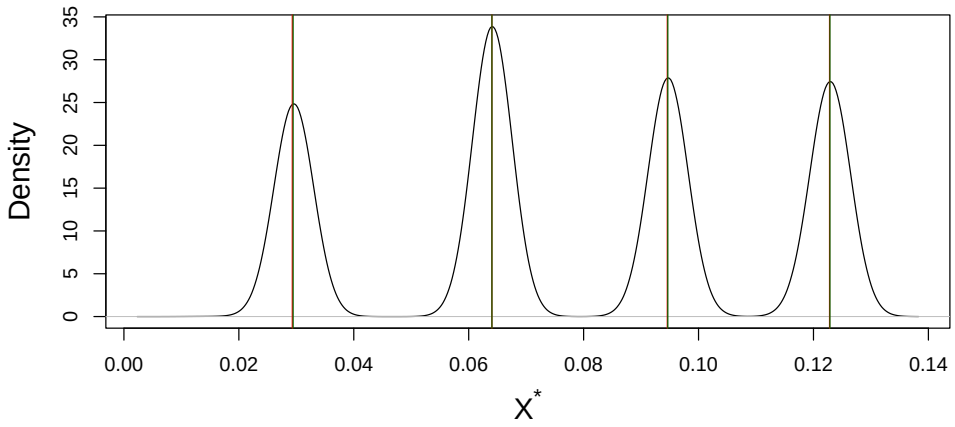

Figure 5.29: The posterior density of X^* given the desired response $Y^* = 0.24$ and the forward model posterior $\boldsymbol{\phi}$. The prior distribution of X_b^* in this case is U(0.01,0.13). The red lines represent the four possible input values that can produce the desired response given the underlying simulation, and the green vertical lines represents the four 'posterior' modes that have been recovered in the backward model; from smallest to largest these are given by 0.0294, 0.0640, 0.0947, and 0.123, to 3 significant figures.

The backward model posterior modes can be checked to see if they are suitable candidate values for X^* using the posterior distribution of the forward model. In this case, given the forward model is fitted using the EIV GP, the EIV GP posterior is used to produce predictions of the response variable at each posterior mode of X^*. Defining X^*_{mode} to be a posterior mode of X^* recovered in the backward model, the EIV GP posterior provides the distribution for each $Y^*_{\text{pred},s}$ given $\boldsymbol{\phi}_s$, given by

$$Y^*_{\text{pred},s} | \boldsymbol{\phi}_s, X^*_{\text{mode}} \sim \text{N}(m^*_{\text{pred},s}, v^*_{\text{pred},s})$$

where

$$m^*_{\text{pred},s} = m_s(X^*_{\text{mode}}) + K_s(X^*_{\text{mode}}, \tilde{\boldsymbol{X}}_s)V_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s)^{-1}(\tilde{\boldsymbol{Y}}_s - m_s(\tilde{\boldsymbol{X}}_s))$$

and

$$v^*_s = v_s(X^*_{\text{mode}}, X^*_{\text{mode}}) - K_s(X^*_{\text{mode}}, \tilde{\boldsymbol{X}}_s)V_s(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{X}}_s)^{-1}K_s(\tilde{\boldsymbol{X}}_s, X^*_{\text{mode}}).$$

Then, the overall distribution for Y^*_{pred} that accounts for the uncertainty in the posterior distribution can be summarised by sampling once from each $Y^*_{\text{pred},s}$ and plotting the density of the samples. These densities, given $X^*_{\text{mode},1} = 0.0294$ and $X^*_{\text{mode},2} = 0.0640$, are given in the same plot in Figure 5.30, where $p(Y^*_{\text{pred},1}|\boldsymbol{\phi}, X^*_{\text{mode},1})$ is the black curve, $p(Y^*_{\text{pred},2}|\boldsymbol{\phi}, X^*_{\text{mode},2})$ is the blue curve, the red vertical line represents $Y^* = 0.24$, the green vertical line represents the mode of $p(Y^*_{\text{pred},1}|\boldsymbol{\phi}, X^*_{\text{mode},1})$ (value of 0.237) and the orange vertical line represents the mode of $p(Y^*_{\text{pred},2}|\boldsymbol{\phi}, X^*_{\text{mode},2})$ (value of 0.240). The densities look almost identical, with the density of the predictions given the second mode shifted across slightly, closer towards the desired response value. The second mode is preferred over the first mode as a candidate value for $Y^* = 0.24$, given that the mode of the predictions for the response variable given the second mode is closer to the desired response.

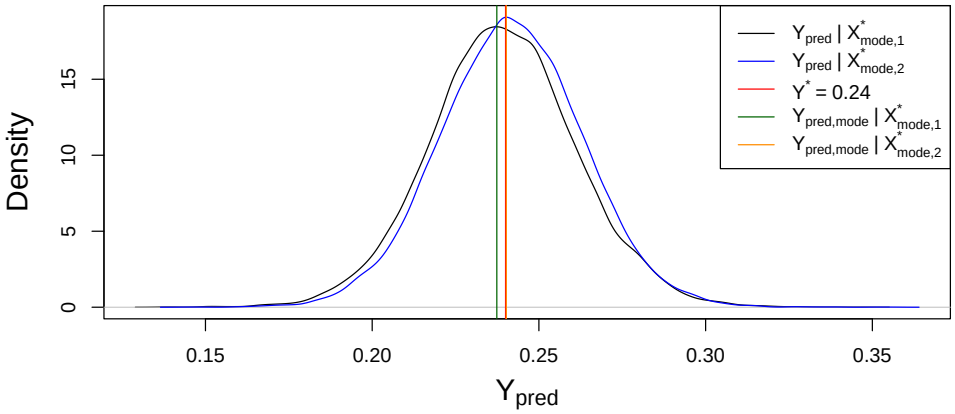

Figure 5.30: The probability density functions $p(Y^*_{\text{pred},1}|\boldsymbol{\phi}, X^*_{\text{mode},1})$ (black curve) and $p(Y^*_{\text{pred},2}|\boldsymbol{\phi}, X^*_{\text{mode},2})$ (blue curve), with the vertical lines corresponding to $Y^* = 0.24$ (red line), the mode of $p(Y^*_{\text{pred},1}|\boldsymbol{\phi}, X^*_{\text{mode},1})$ (green line, value of 0.237) and the mode of $p(Y^*_{\text{pred},2}|\boldsymbol{\phi}, X^*_{\text{mode},2})$ (orange line, value of 0.240).

The analogous probability density functions are provided for the third and fourth modes $X^*_{\text{mode},3} = 0.0947$ and $X^*_{\text{mode},4} = 0.123$ in Figure 5.31.

This figure demonstrates a distinction between the third and fourth modes in terms of the variance in their predictions of the response variable—the black line, representing $p(Y^*_{\text{pred},3}|\boldsymbol{\phi}, X^*_{\text{mode},3})$, has more variance than that of the blue line representing $p(Y^*_{\text{pred},4}|\boldsymbol{\phi}, X^*_{\text{mode},4})$, while both distributions produce a mean prediction that is close to the desired response. Comparing the mode of the predictions given each of the posterior modes of X^*, the mode most likely to produce the desired response of $Y^* = 0.24$ is the second mode $X^*_{\text{mode},2} = 0.0640$.

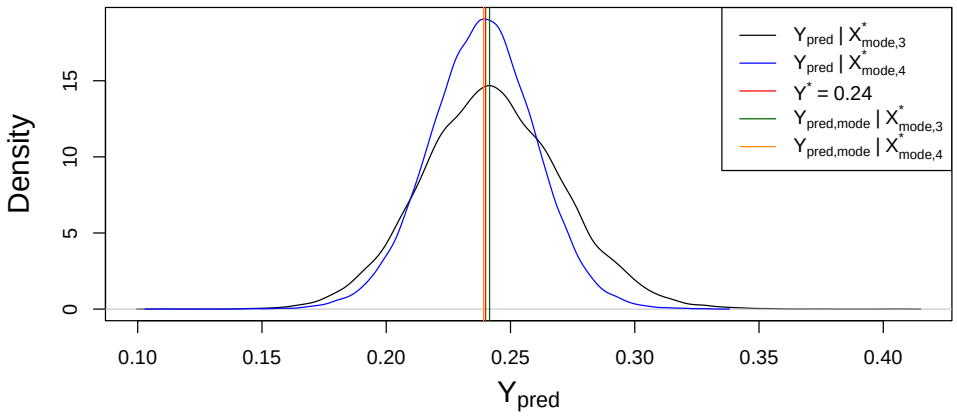

Figure 5.31: The probability density functions $p(Y^*_{\text{pred},3}|\boldsymbol{\phi}, X^*_{\text{mode},3})$ (black curve) and $p(Y^*_{\text{pred},4}|\boldsymbol{\phi}, X^*_{\text{mode},4})$ (blue curve), with the vertical lines corresponding to $Y^* = 0.24$, the mode of $p(Y^*_{\text{pred},3}|\boldsymbol{\phi}, X^*_{\text{mode},3})$ (green line, value of 0.242) and the mode of $p(Y^*_{\text{pred},4}|\boldsymbol{\phi}, X^*_{\text{mode},4})$ (orange line, value of 0.239).

A new desired response value of $Y^* = 0.07$ is considered from the same simulation example, calculated by taking the 15% quantile of the chosen true values of the response variable and rounding to 2 decimal places. This desired response value presents another interesting challenge for the backward model. As indicated by the plot of the underlying simulation in Figure 5.32, with the horizontal line of $Y^* = 0.07$, there are again four possible input values that are candidates for X^*. The two possible modes furthest to the right of the plot in Figure 5.32 are very close in distance to one another—is the backward modelling capable of distinguishing between these two values, or are they so close together than only one peak will be found that is attributed to both values?

The backward modelling algorithm is fitted twice, firstly with $X^*_b \sim \text{U}(0, 1)$ and secondly with $X^* \sim \text{U}(0.01, 0.13)$. The aim in the first case is to discover whether there is any density at values beyond the existing range of

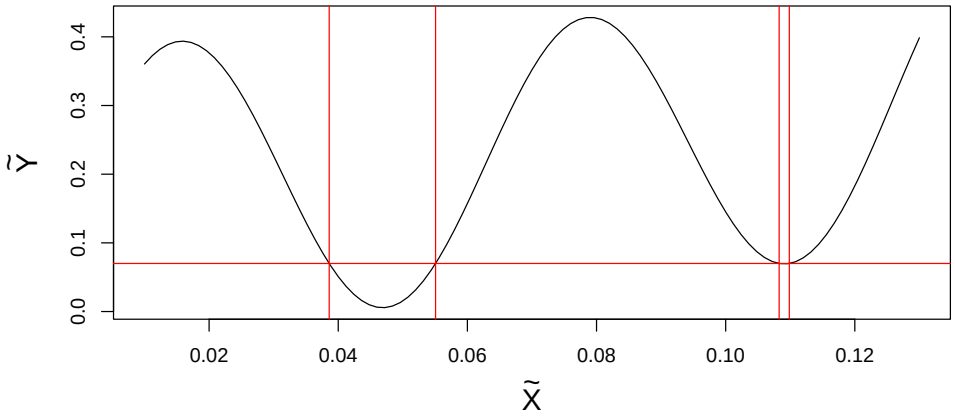

Figure 5.32: A plot of the simulation function for the single-input EIV GP considered in Section 4.3.1, along with a red horizontal line corresponding to the desired response value $Y^* = 0.07$, and four red vertical lines corresponding to the input values $X^* = 0.0386$, $X^* = 0.0551$, $X^* = 0.108$ and $X^* = 0.110$ (all to 3 significant figures), that can produce the desired response value.

true values of the explanatory variable considered in the forward modelling, and the aim in the second case is to discover whether all four possible values can be recovered as posterior modes of X^*. As in the previous multimodal cases, fitting this backward model with $X^* \sim \mathrm{U}(0, 1)$ and $Y^* = 0.07$, the majority of the backward models are deemed to be not converging, given the median value of PSRF upper bounds over all backward models is 7.07. The results are carried over from these runs and analysed here, accepting that the relative probability densities about each mode may not be representative of those of the genuine posterior distribution. The posterior density of X^* given $Y^* = 0.07$ and the prior $X_b^* \sim \mathrm{U}(0, 1)$ is plotted in Figure 5.33. As with the previous example (see Figure 5.28) with $Y^* = 0.24$, there is some density found at extrapolated values (i.e., beyond the range of true values considered in the forward modelling) for X^*. As noted above, the relative probabilities of these 'posterior' distributions in the cases where each backward model is not converging are not reliable—the plot would suggest that there is significantly more density at the modes of X^* than at the extrapolated values of X^*, comparing with the posterior density of Figure 5.28. The posterior density in Figure 5.33 is bimodal, with the corresponding input values provided in Figure 5.32 recovered effectively by the model.

The posterior density of the case with $X_b^* \sim \mathrm{U}(0.01, 0.13)$ is provided

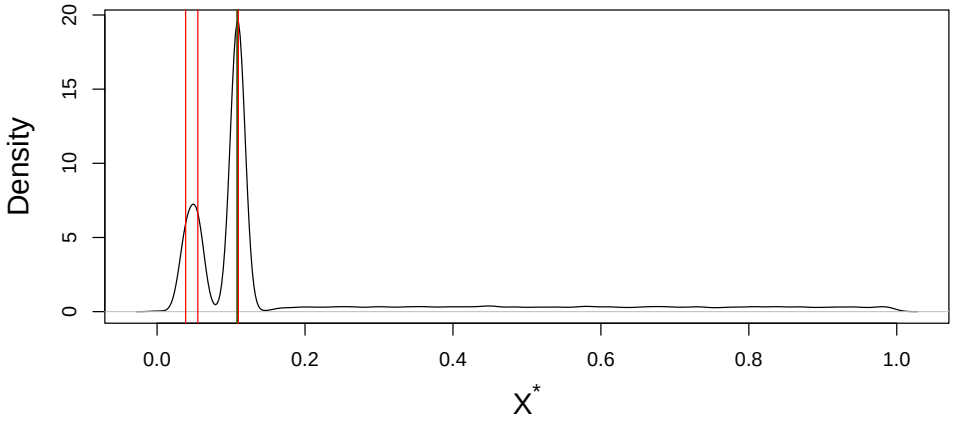

Figure 5.33: The posterior density of $X^* \sim \mathrm{U}(0,1)$ given the desired response $Y^* = 0.07$ and the forward model posterior density $\boldsymbol{\phi}$. The red lines represent the four possible input values that can produce the desired response given the underlying simulation, and the green vertical line represents the 'posterior' mode, given by 0.1081 to significant figures, which is very close to the possible input values 0.1083 and 0.1099 (to four significant figures).

in Figure 5.34. The posterior density appears to have three distinct modes. Unsurprisingly, there is one mode for the two corresponding input values at $X^* = 0.108$ and $X^* = 0.110$, with a posterior mode identified in between these two values of $X^*_{\mathrm{mode},1} = 0.1091$ (to 4 significant figures). Again, there is evidence here to suggest the backward model is recovering appropriate values for the corresponding input variable despite the issues with convergence.

The three posterior modes of $X^*_{\mathrm{mode},1} = 0.1091$, $X^*_{\mathrm{mode},2} = 0.03892$ and $X^*_{\mathrm{mode},3} = 0.0550$ (to 4 significant figures) from the backward model with $Y^* = 0.07$ and $X^*_b \sim \mathrm{U}(0.01, 0.13)$ are now examined using the forward model posterior distribution, to provide predictions of the response variable given each forward model sample s and each posterior mode of X^*. The density of the predictions given the first mode and given the second mode are compared in Figure 5.35—it is clear that the first posterior mode $X^*_{\mathrm{mode},1} = 0.1091$, given the forward model posterior distribution, provides a more accurate prediction of the response variable relative to the desired response value $Y^* = 0.07$, and the predictions of the response variable are much less variable than those for the second posterior mode $X^*_{\mathrm{mode},2} = 0.03892$. The first posterior mode is now compared with the third posterior mode in Figure 5.36 with the density of predictions of

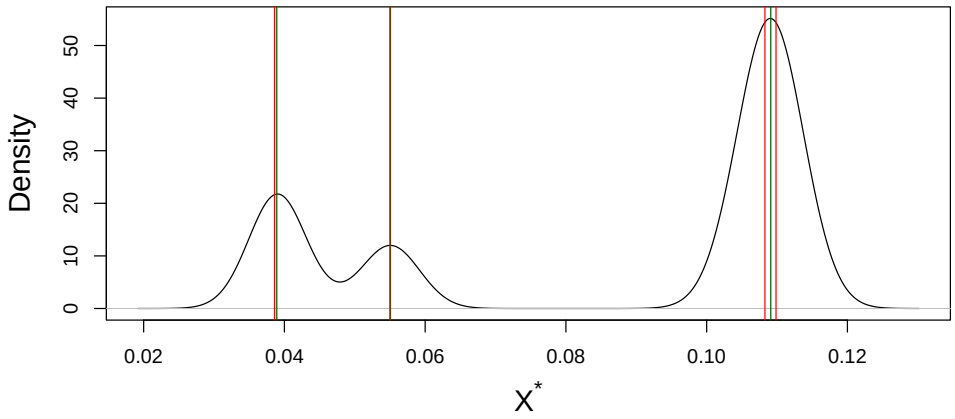

Figure 5.34: The posterior density of $X^* \sim \mathrm{U}(0.01, 0.13)$ given the desired response $Y^* = 0.07$ and the forward model posterior $\boldsymbol{\phi}$. The red lines represent the four possible input values that can produce the desired response given the underlying simulation, and the green vertical line represents the 'posterior' mode, given by 0.1091 to 4 significant figures, which lies in between the two possible input value 0.1083 and 0.1099 (to 4 significant figures).

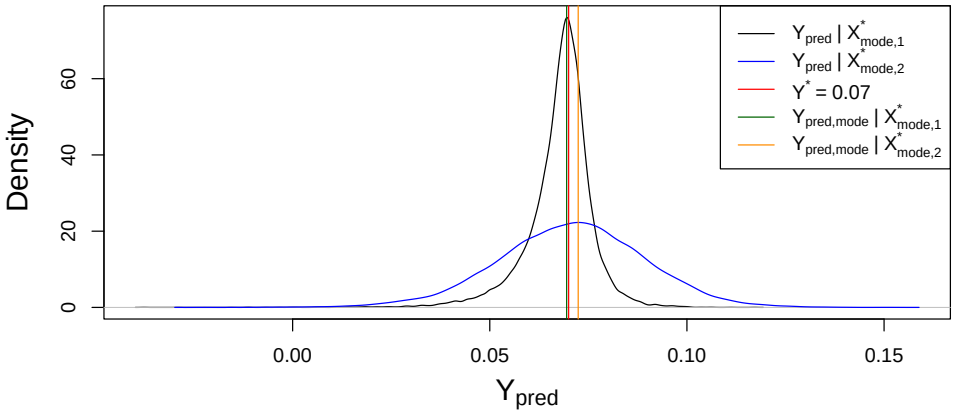

Figure 5.35: The probability density functions $p(Y^*_{\mathrm{pred},1}|\boldsymbol{\phi}, X^*_{\mathrm{mode},1})$ (black curve) and $p(Y^*_{\mathrm{pred},2}|\boldsymbol{\phi}, X^*_{\mathrm{mode},2})$ (blue curve), with the vertical lines corresponding to $Y^* = 0.07$, the mode of $p(Y^*_{\mathrm{pred},1}|\boldsymbol{\phi}, X^*_{\mathrm{mode},1})$ (green line, value of 0.0695) and the mode of $p(Y^*_{\mathrm{pred},2}|\boldsymbol{\phi}, X^*_{\mathrm{mode},2})$ (orange line, value of 0.0724).

the response variable given the forward model posterior distribution $\boldsymbol{\phi}$. In this case, it appears that the third posterior mode performs better than the second posterior mode with respect to the mode of the predictions of the response at those points. The variance of these predictions is very similar. The main conclusion is that the first posterior mode provides the most accurate prediction of the response variable and has the least uncertainty in its prediction, and is recommended as the corresponding input value for X^*.

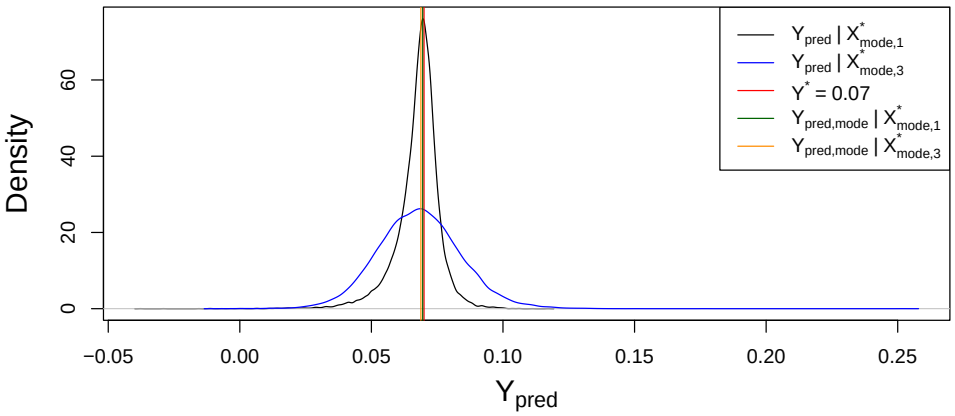

Figure 5.36: The probability density functions $p(Y^*_{\text{pred},1}|\boldsymbol{\phi}, X^*_{\text{mode},1})$ (black curve) and $p(Y^*_{\text{pred},3}|\boldsymbol{\phi}, X^*_{\text{mode},3})$ (blue curve), with the vertical lines corresponding to $Y^* = 0.07$, the mode of $p(Y^*_{\text{pred},1}|\boldsymbol{\phi}, X^*_{\text{mode},1})$ (green line, value of 0.0695) and the mode of $p(Y^*_{\text{pred},3}|\boldsymbol{\phi}, X^*_{\text{mode},3})$ (orange line, value of 0.0688).

The following section considers the backward modelling process being applied to the EIV MOGP. Note that an example of a single-output, multi-input EIV GP backward modelling is not provided here – the EIV MOGP backward modelling considers multiple input variables, and so demonstrates this extension of the process while simultaneously demonstrating the multi-output extension.

5.2.2 Multi-output, multi-input EIV GP

In this section, the backward modelling process is applied to the EIV MOGP. The posterior distribution of the hyperparameters is taken from the forward modelling from Section 4.3.5, that is, the backward modelling here corresponds to the forward model, and underlying simulation, from Section 4.3.5. Below is a brief description of the backward modelling process for the EIV

MOGP.

Having estimated $\boldsymbol{\phi}$, the posterior distribution of the hyperparameters and the true values of the output and input variables, given by $\tilde{Y} = \begin{pmatrix}\tilde{\boldsymbol{Y}}_1 & \tilde{\boldsymbol{Y}}_2\end{pmatrix}$ and $\tilde{X} = \begin{pmatrix}\tilde{\boldsymbol{X}}_1 & \tilde{\boldsymbol{X}}_2\end{pmatrix}$ respectively, the aim of the backward modelling is to identify a candidate vector $\boldsymbol{X}^*$ that can produce some desired response vector $\boldsymbol{Y}^*$. Thus, the conditional distribution of interest is

$$\boldsymbol{Y}^*|\boldsymbol{\phi}, \boldsymbol{X}^* \sim \mathrm{N}_2(\boldsymbol{m}^*, V^*),$$

where

$$\boldsymbol{m}^* = (\alpha_1, \alpha_2)' + \boldsymbol{K}(\boldsymbol{X}^*, \tilde{X})V^{-1}_{\mathrm{MOGP},\tilde{X}}(\mathrm{vec}(\tilde{Y}) - (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'), \qquad (5.2.2.1)$$

and

$$V^* = V_{\mathrm{MOGP},\boldsymbol{X}^*} - \boldsymbol{K}(\boldsymbol{X}^*, \tilde{X})V^{-1}_{\mathrm{MOGP},\tilde{X}}\boldsymbol{K}(\tilde{X}, \boldsymbol{X}^*). \qquad (5.2.2.2)$$

For clarity, note that

$$V_{\mathrm{MOGP},\tilde{X}} = V_{K,1} \otimes K_1(\tilde{X}, \tilde{X}) + V_{K,2} \otimes K_2(\tilde{X}, \tilde{X}) + \begin{pmatrix}\sigma^2_{\epsilon_1} I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma^2_{\epsilon_2} I_{n_g}\end{pmatrix},$$

with the matrix $V_{\mathrm{MOGP},\boldsymbol{X}^*}$ derived analogously, and

$$\boldsymbol{K}(\boldsymbol{X}^*, \tilde{X}) = V_{K,1} \otimes K_1(\boldsymbol{X}^*, \tilde{X}) + V_{K,2} \otimes K_2(\boldsymbol{X}^*, \tilde{X}),$$

and $\boldsymbol{K}(\tilde{X}, \boldsymbol{X}^*) = (\boldsymbol{K}(\boldsymbol{X}^*, \tilde{X}))^T$. For further details of other parameters in the modelling, the reader is referred to Section 3.5.5. While the above distribution is described in terms of how the hyperparameters $\boldsymbol{\phi}$ and the vector $\boldsymbol{X}^*$ influence the response vector $\boldsymbol{Y}^*$, the behaviour of $\boldsymbol{X}^*$ given $\boldsymbol{\phi}$ and $\boldsymbol{Y}^*$ can be investigated by placing a prior distribution on $\boldsymbol{X}^*$ and fixing the desired response at $\boldsymbol{Y}^* = \boldsymbol{y}^*$, and estimating the corresponding input vector with the posterior distribution.

In order to estimate the posterior distribution of $\boldsymbol{X}^*|\boldsymbol{Y}^*, \boldsymbol{\phi}$, the above conditional distribution is investigated for some large S_{back} random samples of the posterior distribution of the hyperparameters and true values of $\boldsymbol{\phi}$, with which a posterior distribution of $\boldsymbol{X}^*_b$ is found for each $b = 1, \ldots, S_{\mathrm{back}}$. The complete posterior distribution of $\boldsymbol{X}^*$ is then built by collecting a small set of samples from each $\boldsymbol{X}^*_b$. As noted in previous sections relating to applying the backward model to simulations, there are convergence and mixing issues in the backward models where multimodal posterior distributions can be recovered. This also occurs here, and as previously done, the results of the backward modelling are still examined despite this.

The backward model is investigated using the desired response vector $\boldsymbol{Y}^* = (18, 237)'$ (before scaling). The main aim of this demonstration was to show that the backward modelling with multiple output variables can lead to interesting posterior distributions for the corresponding input vector, where there could be multiple candidate vectors to produce the desired output. Having considered plots of the underlying true values for the output variables and the input variables (see Figure 5.37), it was clear that the output true values of groups 5 and 12 (cyan and yellow) were relatively close, compared with their corresponding input true values. This indicated that desired output vectors relatively close to those of groups 5 and 12 were likely to have multiple vectors of the input variables that could produce the desired output.

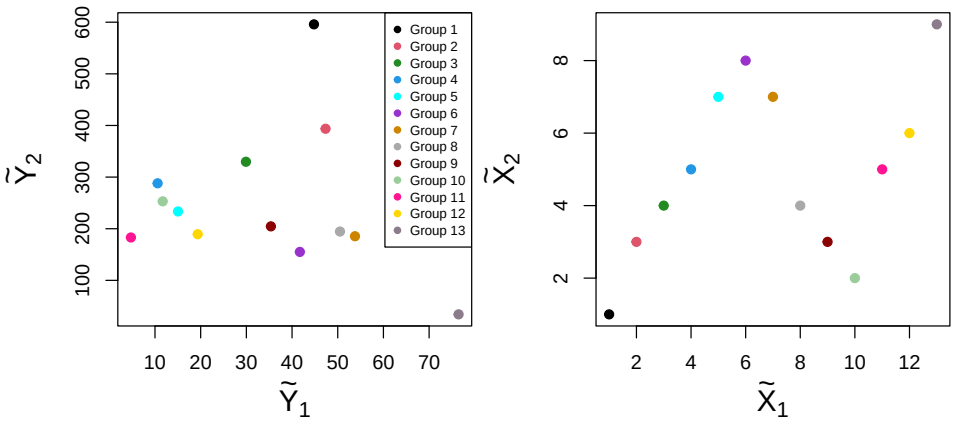

Figure 5.37: The underlying true values from the EIV MOGP simulation built in Section 4.3.5. The left plot corresponds to the output true values (with the second output on the vertical axis, first output on the horizontal axis), and the right plot corresponds to the input true values (second input on the vertical axis, first input on the horizontal). The relative distances between groups 5 and 12 (cyan and yellow) in each plot is noted.

Then, through trial and error, output vectors close to those of groups 5 and 12 were selected, and plots of the input vectors were examined to give an indication as to whether a multimodal distribution would be recovered in the posterior (i.e., whether there were multiple regions of the input space with possible candidate vectors). This plot, for $\boldsymbol{Y}^* = (18, 237)$, is provided in Figure 5.38, which also indicates the locations of the underlying true values of the input variables. The red curve highlights the input vectors that are able to produce the desired value for the first output variable to within 1%

of the value, and the blue curve does the equivalent for the second output. Therefore, the intersections of these curves indicates the input vectors that are optimal, i.e., that we hope the backward model can recover as candidate vectors for the input variables. Expectations must be tempered somewhat – note that the forward modelling learns about the functional relationship at the true values, i.e., the black points. As one moves further away from the true values, the quality of the predictions of the response variables becomes worse (in the sense of the forward modelling), and therefore the likelihood the backward model looks to these regions for candidate vectors becomes smaller. For this reason, it is fair to expect that the possible candidate vectors in close proximity to (0.6,0.34) and (1.23,0.25) would not be recovered in the backward modelling. Instead, the possible candidate vectors close to (0.97,0.23), (0.44,0.74) and (0.6,0.6) could be recovered by the backward modelling.

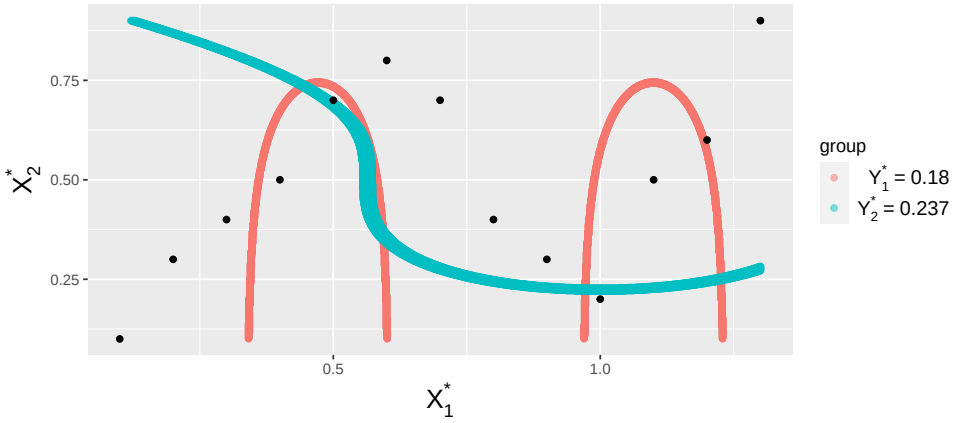

Figure 5.38: Plot of the input space demonstrating input vectors that can produce the desired values for the first output variable (highlighted in red) and for the second output variable (highlighted in blue), with the black points being the underlying true values of the input variables from the simulation.

The backward modelling process is now fitted, with the desired response vector $\boldsymbol{Y}^* = (18, 237)$ (which becomes (0.18,0.237) after scaling) and the posterior distribution of the hyperparameters and true values $\boldsymbol{\phi}$. Moreover, the prior distribution assumed for $\boldsymbol{X}^*$ is a uniform box, with the ranges in each dimension being $[0.1, 1.3]$ (for the first input variable; note that this variable was rescaled in the forward modelling, see Section 4.3.5) and $[0.1, 0.9]$ for the second input variable. The joint posterior density of $\boldsymbol{X}^*$ is

provided in Figure 5.39, along with the red and blue curves from Figure 5.38, as well as the underlying true values of the input variables. Clearly, there are two distinct regions of the posterior distribution of $\boldsymbol{X}^*$ surrounding two posterior modes, with the global mode located at $\boldsymbol{X}^*_{\text{mode},1} = (0.523, 0.671)$ (pink point) and a local mode located at $\boldsymbol{X}^*_{\text{mode},2} = (0.974, 0.228)$ (yellow point). Given the issues with convergence and mixing in the backward modelling, it is noted that the relative densities at these points may not be accurate, and so it is possible that the pink point is just a local mode with the yellow point being a global mode. Despite this caveat, it is evident that the backward modelling is functioning as intended here. Moreover, this plot is the ideal demonstration of the capabilities of the backward modelling process developed in this work. It provides evidence that, given a forward modelling posterior distribution, it is possible to find multiple input vectors which are capable of producing a multivariate desired response.

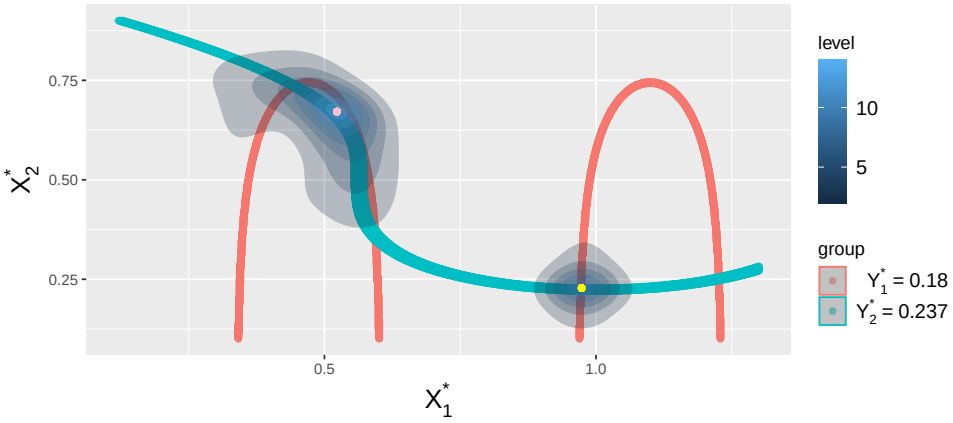

Figure 5.39: Joint posterior density of $\boldsymbol{X}^*|\boldsymbol{\phi}, \boldsymbol{Y}^* = (0.18, 0.237)$, along with the red and blue curves and the underlying input true values (black points) from Figure 5.38.

Typically, it would be assumed that the global posterior mode of $\boldsymbol{X}^*$ would be chosen as the candidate vector, i.e., the best estimate from the modelling at an input vector which could produce the corresponding desired response vector. Given the convergence issues, this would not be assumed here, and the best way to decide which of these two modes is the best candidate would be to run these modes through the EIV MOGP posterior, that is, finding the predictive distribution for the output variables given each mode. In any case, given a multimodal posterior distribution for $\boldsymbol{X}^*$, it would be sensible to estimate the response variables given the multiple modes. Details

of the EIV MOGP posterior distribution can be found in Section 3.5.5; the estimation of the response variables here is roughly analogous to the backward modelling process described earlier, with the variable $\boldsymbol{X}^*$ now fixed to be either $\boldsymbol{X}^*_{\text{mode},1}$ or $\boldsymbol{X}^*_{\text{mode},2}$.

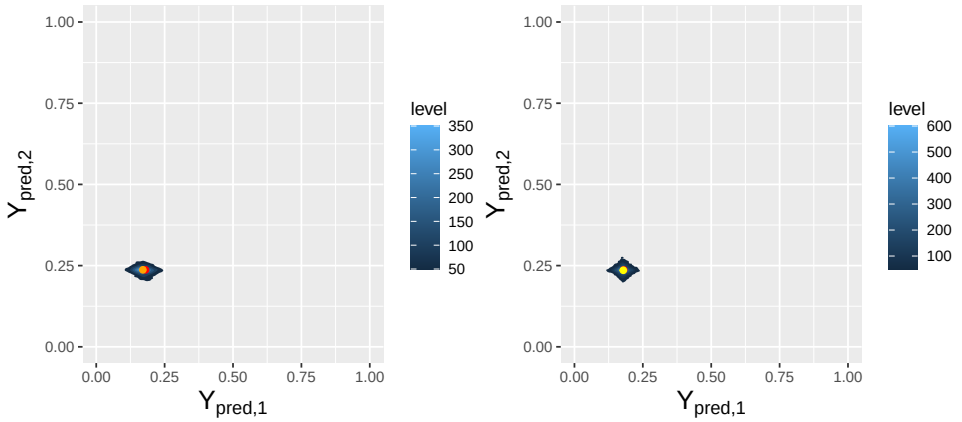

Figure 5.40: Two plots of the joint probability density function of predictions of the response variables, with the left plot being predictions at the global posterior mode $\boldsymbol{X}^*_{\text{mode},1}$, and the right plot being predictions at the local mode $\boldsymbol{X}^*_{\text{mode},2}$. The desired response vector $\boldsymbol{Y}^* = (0.18, 0.237)$ is given by the red point in each plot, and the mode of the distribution of the predictions is given by the pink point in the left plot, and the yellow point in the right plot.

The EIV MOGP posterior, providing joint predictions of the response variable given some input vector, is summarised by the joint density functions given in Figure 5.40 (left plot is the predictive distribution given the global mode $\boldsymbol{X}^*_{\text{mode},1}$, right plot is the analogous plot given the local mode $\boldsymbol{X}^*_{\text{mode},2}$). The red point in both plots is the desired response vector $\boldsymbol{Y}^* = (0.18, 0.237)$, and the modes of the predictive distribution are given by $\boldsymbol{Y}_{\text{pred},1} = (0.170, 0.237)$ (left plot) and $\boldsymbol{Y}_{\text{pred},2} = (0.179, 0.237)$ (right plot) to 3 significant figures. It is clear from the plots that the predictive distribution corresponding to the second mode $\boldsymbol{X}^*_{\text{mode},2} = (0.974, 0.228)$ has a mode which is slightly closer to the desired response vector, where the difference between the quality of the modes of the predictive distribution is in the prediction of the first output variable. In both cases, the uncertainty in the prediction is similar. It is unsurprising, given that the posterior distribution of $\boldsymbol{X}^*$ identified vectors that were known to be able to produce the desired response vector, that the two posterior modes of $\boldsymbol{X}^*$ are suitable candidate

vectors given the EIV MOGP forward modelling (with a slight preference for the local mode $\boldsymbol{X}^*_{\text{mode},2}$, which provides a more accurate prediction of the desired response vector). As noted above, this example demonstrates the capability of the backward modelling process developed in this work, with a multimodal posterior distribution identified for $\boldsymbol{X}^*$, where both modes are capable of producing the desired response vector accurately.

5.3 Powder flow data

Having demonstrated that the backward modelling algorithm has been able to capture corresponding input values as posterior modes of X^* for simulated data sets, it is now tested on the real-world powder flow data that has been investigated in the forward modelling in Sections 4.4.3 and 4.4.4. The following subsection considers the backward modelling process for the EIV BR, which is applied using the fitted forward model identified in Section 4.4.3, and the content of Section 5.3.2 considers the backward modelling process for the EIV GP. In both cases, the fitted forward models given the informed prior distributions are considered.

5.3.1 Errors-in-variables Bayesian regression for powder flow data

Having taken into consideration the highly plausible overfitting issues with the six-explanatory-variable forward model that was initially suggested by the backwards variable selection in Section 4.4.3, it was concluded that a more appropriate model for this set of data is the two-explanatory-variable model with CBD and SE as the two input variables. The corresponding fitted forward model is now investigated for the purpose of optimising the input variables for a given desired response vector, i.e., optimal values of CBD and SE are to be found that correspond to desired values for both tapped density and angle of repose simultaneously.

In this section, the backward modelling algorithm is applied to this model, given the estimated posterior distribution of the (relevant) model parameters $\boldsymbol{\theta}$, where

$$\boldsymbol{\theta} = (\beta_{0,1}, \beta_{1,1}, \beta_{2,1}, \beta_{0,2}, \beta_{1,2}, \beta_{2,2}, T_\epsilon),$$

and given some desired response vector $\boldsymbol{Y}^* = \boldsymbol{y}^*$ for TD and AoR. Desirable values for the two response variables have been elicited from Dr Candice Majewski. For tapped density, these were provided with respect to the powder

PA2200 which is one of the powders considered in the forward modelling. Dr Majewski recommended to use the expected range for the tapped density from PA2200 as the desired values for tapped density. In this case, the maximum value of the observed tapped density values from PA2200 is taken to be the desired response value, given by 0.51 (to 2 decimal places). For angle of repose, a good basis for its values is

- below 25 degrees is excellent,
- between 25 and 30 degrees is good,
- between 30-40 is passable.

An extra caveat with this information is that, while 25 degrees is considered excellent (with respect to producing freely flowing powders), there is some concern that powders that flow too freely could be prone to depositing poorly onto a powder bed. For this reason, Dr Majewski recommends values between 25 and 30 degrees as desirable. For this reason, a value of 27.5 degrees is chosen as the desired response value for angle of repose (scaled onto 0.275 for modelling purposes), where some attention is required as to what values of angle of repose are produced given the eventual posterior mode for $\boldsymbol{X}^*$.

The desired response vector is chosen to be $\boldsymbol{Y}^* = (0.51, 0.275)$. Before implementing the backward modelling process, the desired response vector $\boldsymbol{Y}^*$ is plotted alongside the joint posterior densities of the true values of each group for AoR against TD. This plot is provided on the left-hand side in Figure 5.41, and indicates that Powders 2, 3, 4 and 6 appear to lie in close proximity to the desired response vector, with Powder 2 being the closest. On the right side of this plot are the joint posterior densities of the true values for each group for SE against CBD. Given the proximity of the desired response vector relative the the response values of existing groups, it is expected that the posterior density of the corresponding input vector provides vectors of CBD and SE that are close in proximity to the true input vectors of these powders – that is, roughly speaking, CBD values between 0.43 and 0.47, and SE values between 0.55 and 0.7. This expectation is made given the forward modelling is based on an additive relationship between the two explanatory variables for both response variables.

The prior distribution of $\boldsymbol{X}^*$ is now determined. For both explanatory variables, Dr Majewski recommends considering the ranges of CBD and SE utilised in the forward modelling, while suggesting that there is some preference for SE values closer to 5 than 8. The prior distribution is chosen to

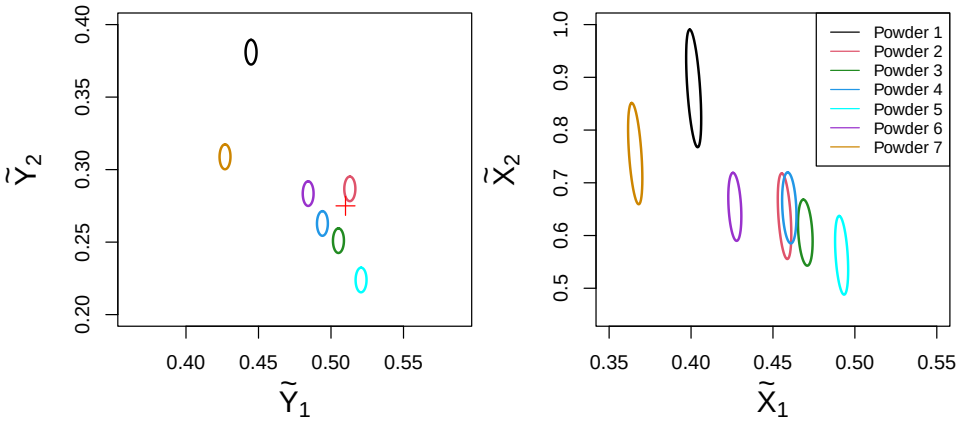

Figure 5.41: Two plots relating the true values of the response variables (left plot) and the true values of the input variables (right plot). The plot of true values of the response variables also includes the desired response vector $\boldsymbol{Y}^* = (0.51, 0.275)$ (red plus). The ellipses in each case represent joint 95% credible ellipses for each group.

be a uniform box, with $X^*_{1,b} \sim \text{U}(0.35, 0.55)$ and $X^*_{2,b} \sim \text{U}(0.5, 0.8)$ for CBD and SE respectively, with values scaled onto [0,1]. Therefore, the backward models to be fitted for $b = 1, \ldots, S_{\text{back}}$ (with $S_{\text{back}} = 2000$ being a large number of posterior samples from the forward model posterior distribution) are

$$\begin{pmatrix} 0.51 \\ 0.275 \end{pmatrix} = \begin{pmatrix} \beta_{0,1,b} & \beta_{1,1,b} & \beta_{2,1,b} \\ \beta_{0,2,b} & \beta_{1,2,b} & \beta_{2,2,b} \end{pmatrix} \begin{pmatrix} 1 \\ X^*_{1,b} \\ X^*_{2,b} \end{pmatrix} + \begin{pmatrix} \epsilon^*_{1,b} \\ \epsilon^*_{2,b} \end{pmatrix}, \qquad (5.3.1.1)$$

where $\boldsymbol{\epsilon}^*_b = (\epsilon^*_{1,b}, \epsilon^*_{2,b})' \sim \text{N}(\boldsymbol{0}, T_{\epsilon,b})$, with the subscript b denoting a posterior sample of the parameter from the forward model posterior distribution. Moreover, for each $b = 1, \ldots, 2000$, $X^*_{1,b} \sim \text{U}(0.35, 0.55)$ and $X^*_{2,b} \sim \text{U}(0.5, 0.8)$ are the prior distributions for the corresponding input variables, for which a posterior distribution is estimated for each b, and posterior samples are drawn from each $\boldsymbol{X}^*_b$ and combined to build the complete distribution of $\boldsymbol{X}^*|\boldsymbol{\theta}, \boldsymbol{Y}^*$.

The initialisation for the MCMC algorithm for each b is briefly discussed. In the simulated-data cases, a Latin hypercube sampling is carried out to determine unique initial values for each $\boldsymbol{X}^*_b$ that cover the input space, where the same initial vector for $\boldsymbol{X}^*_b$ was chosen for each of the two parallel chains that were run for each backward model b. In order to provide a more robust

method for demonstrating the convergence of each backward model run, unique input vectors are chosen for each parallel chain of each backward model run in this case of the backward modelling applied to the real data. In effect, a Latin hypercube sampling is carried out with dimension 2 (for the two input variables) and 4000 samples, determined by there being 2 parallel chains for each of the 2000 backward models. The initial vectors of the corresponding input variables are then randomly selected from the Latin hypercube sampling.

The backward modelling algorithm is firstly run with the MCMC tuning parameters, for each backward model b, of

- an adaptation phase of length 500 samples,
- a burn-in phase of length 15000 samples,
- 2000 posterior samples stored, taking every 10^{th} sample from 20000 posterior samples,
- two parallel chains for each backward model b.

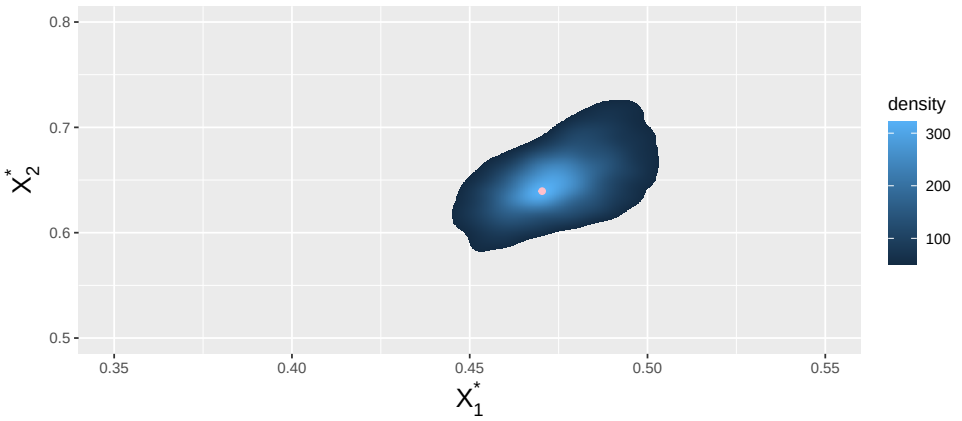

Figure 5.42: The posterior density $p(\boldsymbol{X}^*|\boldsymbol{\theta}, \boldsymbol{Y}^*)$ for the real data, given the desired response vector of $\boldsymbol{Y}^* = (0.51, 0.275)$, where the pink dot represent the joint posterior mode of $\boldsymbol{X}^*|\boldsymbol{\theta}, \boldsymbol{Y}^*$. This backward model posterior density function is based on the EIV BR modelling.

Fitting the backward modelling algorithm with these tuning parameters does not lead to convergence in all backward models. The tuning parameters are adjusted until convergence is observed for each backward model b; this occurs for an adaptation phase of length 1500 samples, a burn-in phase

of length 50000 samples, and 10000 posterior samples stored having taken every 15^{th} sample from 150000 posterior samples. In this case, the MCMC output from backward model run 485 has not converged, while the remaining backward models have converged to the posterior distribution. The maximum upper bound of PSRF from backward model run 485 is 1.17, suggesting the model is relatively close to the convergence cutoff of 1.1. Running this backward model iteration with an adaptation phase of 2000 instead of 1500 provides MCMC output which has converged to the posterior distribution, and so all backward models have now converged, and sufficient levels of mixing are observed in each case. Consequently, 10 samples are taken from each backward model posterior distribution $p(\boldsymbol{X}_b^*|\boldsymbol{\theta}_b, \boldsymbol{Y}^*)$ to build the complete posterior distribution of $\boldsymbol{X}^*|\boldsymbol{\theta}, \boldsymbol{Y}^*$, which is now examined. The posterior density $p(\boldsymbol{X}^*|\boldsymbol{\theta}, \boldsymbol{Y}^*)$ is provided in Figure 5.42, with the posterior mode of the distribution, $(0.4704, 0.6394)$ to 4 decimal places, represented by the pink dot. The posterior distribution is notably unimodal, which is unsurprising given that each backward model run b was able to converge with relative ease, and given that an additive linear predictor, with no interaction or polynomial terms, describes the relationship between the outputs and the inputs. Interestingly, with respect to the true values plotted in Figure 5.41, the vectors of CBD and SE recovered in the backward modelling are slightly larger in both dimensions than those initially expected. The posterior mode of $(0.4704, 0.6394)$, which is chosen as the candidate value, still appears to be appropriate.

The posterior mode $\boldsymbol{X}^*_{\text{mode}} = (0.4704, 0.6394)$ is now tested using the forward model posterior distribution, to provide a distribution of predictions of the response variables $\boldsymbol{Y}_{\text{pred}}$, with each prediction sample given by

$$\begin{pmatrix} Y_{1,\text{pred},s} \\ Y_{2,\text{pred},s} \end{pmatrix} = \begin{pmatrix} \beta_{0,1,s} & \beta_{1,1,s} & \beta_{2,1,s} \\ \beta_{0,2,s} & \beta_{1,2,s} & \beta_{2,2,s} \end{pmatrix} \begin{pmatrix} 1 \\ X^*_{1,\text{mode}} \\ X^*_{2,\text{mode}} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,s} \\ \epsilon_{2,s} \end{pmatrix}.$$

Note that the terms $X^*_{1,\text{mode}}$, $X^*_{2,\text{mode}}$ are given by each dimension of the joint posterior mode of $\boldsymbol{X}^*|\boldsymbol{\theta}, \boldsymbol{Y}^*$, as opposed to being the marginal posterior modes of $X_1^*|\boldsymbol{\theta}, \boldsymbol{Y}^*$ and $X_2^*|\boldsymbol{\theta}, \boldsymbol{Y}^*$. The joint probability density function of the distribution of predictions $\boldsymbol{Y}_{\text{pred}}$ is provided in Figure 5.43. The proximity between the mode of the distribution $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode}})$ (given by the orange point (0.5077,0.2687) in Figure 5.43) and the desired response vector $\boldsymbol{Y}^*$ (given by the red point in Figure 5.43) indicates that the posterior mode of $\boldsymbol{X}^*_{\text{mode}}$ is a suitable candidate vector for $\boldsymbol{X}^*$, with the desired response vector $\boldsymbol{Y}^* = (0.51, 0.275)$ being a high-density vector of the joint distribution of predictions. That is, the backward model recommends that a CBD

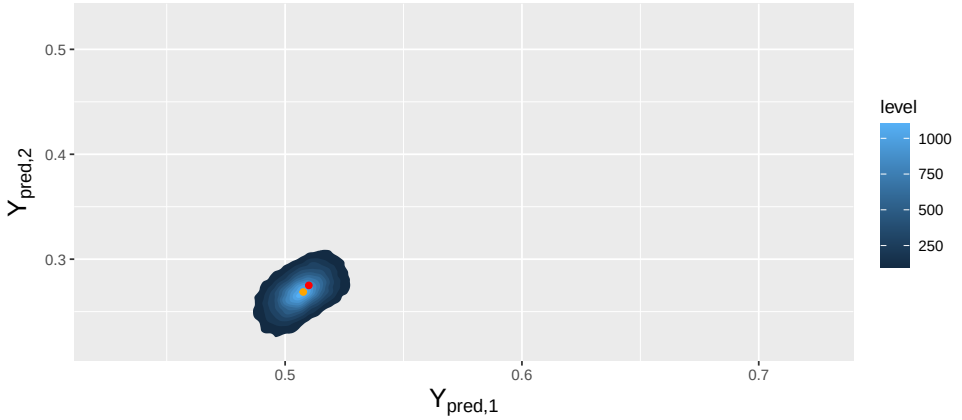

Figure 5.43: The joint probability density function $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\theta}, \boldsymbol{X}^*_{\text{mode}})$ for the real data, given the joint posterior mode of $\boldsymbol{X}^*_{\text{mode}} = (0.4704, 0.6394)$ determined in the backward modelling given the desired response vector of $\boldsymbol{Y}^* = (0.51, 0.275)$. The orange point represents the mode of the joint distribution of predictions, given by $\boldsymbol{Y}_{\text{pred,mode}} = (0.5077, 0.2687)$, and the red point represents the desired response vector $\boldsymbol{Y}^*$. This predictive density function is based on the EIV BR modelling.

value of 0.4704 g/ml and an SE value of 6.394 mJ/g are the optimal values for producing the desired values of tapped density (0.51 g/ml) and angle of repose (27.5 degrees) simultaneously. Following the suggestion from Dr Majewski that any values of AoR lying with 25-30 degrees are 'good', the probability of AoR lying within this range given the forward modelling posterior distribution and the posterior mode of $\boldsymbol{X}^*$ is estimated to be 64.0%.

5.3.2 Errors-in-variables Gaussian processes for powder flow data

Using the forward model posterior distribution estimated in Section 4.4.4, having fitted the EIV MOGP to the powder flow data with CBD and SE as the two input variables, the backward modelling process is now investigated. The forward model posterior distribution estimated for this model is based on the informed prior distributions discussed in Sections 4.4.3 and 4.4.4, having elicited expert information regarding the measurement error variability for the response variables TD and AoR, as well as for the explanatory variables CBD, SE and BFE. Moreover, expert information has been elicited for the range of true values of these variables, assuming the

powders are appropriate for use in Laser Sintering.

In the backward modelling, as discussed in Section 5.3.1, the desired response vector to be investigated is $\boldsymbol{Y}^* = (0.51, 0.275)$, after scaling. Furthermore, the aim of the backward modelling process for this data set is to identify the optimal values of the explanatory variables (i.e., the variables measured from the FT4 powder rheometer) in order to produce a powder which both flows optimally and deposits optimally onto the powder bed. Having identified the model with CBD and SE from the EIV BR forward modelling as the best-fitting model, the EIV MOGP was fitted with TD and AoR as the output variables and CBD and SE as the input variables, so the task here is to find optimal values for CBD and SE, as carried out in the previous section.

The backward model for the EIV MOGP from Section 3.6.7 is described again here. Having fitted the forward model for the EIV MOGP, the true values $\tilde{Y} \in \mathbb{R}^{7\times 2}$ and $\tilde{X} \in \mathbb{R}^{7\times 2}$, as well as the hyperparameters for the GP, have been estimated by the posterior distribution. The parameter vector $\boldsymbol{\phi}$, including the true values and the hyperparameters, is given by

$$\boldsymbol{\phi} = (\tilde{Y}, \tilde{X}, \boldsymbol{\alpha}, \boldsymbol{\sigma}_\epsilon, \boldsymbol{l}, \boldsymbol{\sigma}_k, \boldsymbol{\lambda}, \rho_{V_K}),$$

with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, $\boldsymbol{\sigma}_\epsilon = (\sigma_{\epsilon_1}, \sigma_{\epsilon_2})$, $\boldsymbol{l} = (l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2})$, $\boldsymbol{\sigma}_k = (\sigma_{k_1}, \sigma_{k_2})$. Note that measurement error precisions are dropped from the conditioning as they do not impact 'new' true values.

For the backward modelling process, a large random sample S_{back} of draws are taken from the forward model posterior distribution $\boldsymbol{\phi}$, to investigate, for each $b = 1, \ldots, S_{\text{back}}$, the conditional distribution of $\boldsymbol{Y}^* | \boldsymbol{\phi}_b, \boldsymbol{X}_b^*$ (i.e., the GP posterior), which is derived from the joint distribution

$$\begin{pmatrix} \text{vec}(\tilde{Y}_b) \\ \boldsymbol{Y}^* \end{pmatrix} \sim \text{N}_{2n_g+2} \left(\begin{pmatrix} \boldsymbol{\alpha}_{1,b} \\ \boldsymbol{\alpha}_{2,b} \\ \alpha_{1,b} \\ \alpha_{2,b} \end{pmatrix}, \begin{pmatrix} V_{\text{MOGP},\tilde{X},b} & \boldsymbol{K}_b(\tilde{X}_b, \boldsymbol{X}_b^*) \\ \boldsymbol{K}_b(\boldsymbol{X}_b^*, \tilde{X}_b) & V_{\text{MOGP},\boldsymbol{X}^*,b} \end{pmatrix} \right).$$

The block matrices given in the covariance matrix of the above joint distribution are defined by

$$V_{\text{MOGP},\tilde{X},b} = V_{K,1,b} \otimes K_{1,b}(\tilde{X}_b, \tilde{X}_b) + V_{K,2,b} \otimes K_{2,b}(\tilde{X}_b, \tilde{X}_b) + \begin{pmatrix} \sigma^2_{\epsilon_1,b} I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma^2_{\epsilon_2,b} I_{n_g} \end{pmatrix},$$

$$V_{\text{MOGP},\boldsymbol{X}^*,b} = V_{K,1,b} \otimes K_{1,b}(\boldsymbol{X}_b^*, \boldsymbol{X}_b^*) + V_{K,2,b} \otimes K_{2,b}(\boldsymbol{X}_b^*, \boldsymbol{X}_b^*) + \begin{pmatrix} \sigma^2_{\epsilon_1,b} I_{n_g} & 0_{n_g \times n_g} \\ 0_{n_g \times n_g} & \sigma^2_{\epsilon_2,b} I_{n_g} \end{pmatrix},$$

$$\boldsymbol{K}_b(\boldsymbol{X}_b^*, \tilde{X}_b) = V_{K,1,b} \otimes K_{1,b}(\boldsymbol{X}_b^*, \tilde{X}_b) + V_{K,2,b} \otimes K_{2,b}(\boldsymbol{X}_b^*, \tilde{X}_b),$$

and finally $\boldsymbol{K}_b(\tilde{X}_b, \boldsymbol{X}_b^*) = \boldsymbol{K}_b(\boldsymbol{X}_b^*, \tilde{X}_b)^T$. Thus, using standard distribution theory, the conditional distribution $\boldsymbol{Y}^*|\boldsymbol{\phi}_b, \boldsymbol{X}_b^*$ is given by

$$\boldsymbol{Y}^*|\boldsymbol{\phi}_b, \boldsymbol{X}_b^* \sim \mathrm{N}_2(\boldsymbol{m}_b^*, V_b^*),$$

where

$$\boldsymbol{m}_b^* = (\alpha_{1,b}, \alpha_{2,b})' + \boldsymbol{K}_b(\boldsymbol{X}_b^*, \tilde{X}_b) V^{-1}_{\mathrm{MOGP},\tilde{X},b}(\mathrm{vec}(\tilde{Y}_b) - (\boldsymbol{\alpha}_{1,b}, \boldsymbol{\alpha}_{2,b})'),$$

and

$$V_b^* = V_{\mathrm{MOGP},\boldsymbol{X}^*,b} - \boldsymbol{K}_b(\boldsymbol{X}_b^*, \tilde{X}_b) V^{-1}_{\mathrm{MOGP},\tilde{X},b} \boldsymbol{K}_b(\tilde{X}_b, \boldsymbol{X}_b^*).$$

Using the backward modelling process, with $\boldsymbol{Y}^*$ fixed at the vector $(0.51, 0.275)$ and, for each $b = 1, \ldots, S_{\mathrm{back}}$, using the uniform box prior distribution on $\boldsymbol{X}_b^*$ with $X_1^* \sim \mathrm{U}(0.35, 0.55)$ and $X_2^* \sim \mathrm{U}(0.5, 0.8)$, the corresponding input vector can be optimised using MCMC simulation for the purpose of producing the desired response vector. This provides a posterior distribution for $\boldsymbol{X}_b^*|\boldsymbol{\phi}_b, \boldsymbol{Y}^*$ for each b, from which a small sample is drawn from each and combined together to give the overall posterior distribution of $\boldsymbol{X}^*|\boldsymbol{\phi}, \boldsymbol{Y}^*$.

The backward modelling algorithm is fitted with the MCMC tuning parameters (for each backward model iteration) of an adaptation phase of length 1000, a burn-in phase of length 50000, then storing every 20^{th} sample from 200000 posterior draws to given 10000 posterior samples. This is carried out for two parallel chains to be able to test the convergence of the MCMC output to the posterior distribution for each backward model iteration. With these tuning parameters, it is observed that 32 of the 2000 backward models do not converge to the posterior distribution. Of those 32 backward models iterations that do not converge, 9 are deemed to be 'close' to convergence, which is being defined here as having an upper bound for PSRF between 1.1 and 1.15, either for both dimensions of $\boldsymbol{X}^*$, or given one dimension of $\boldsymbol{X}^*$ has already converged. For these 9 cases, it would be possible to run the MCMC simulation for 'slightly' longer (and/or a longer burn-in period) to obtain convergence. Running the simulation for slightly longer is hard to define, but given the PSRF convergence statistics, running the simulation for 1.15 times the number of samples ought to be appropriate (but is not carried out here). It remains that 23 backward model iterations are not close to convergence, suggesting a larger issue at hand that is likely related to the backward model convergence issues from previous sections when it is expected that the posterior of $\boldsymbol{X}^*$ is multimodal.

Moreover, of those 32 backward models that do not converge, there are some issues with the mixing of the MCMC output. In four of those models, the minimum estimate of effective sample size for either dimension of $\boldsymbol{X}^*$ is below the cutoff for three of the models; the cutoff in this case is the value 20, given by $5m$, where m is twice the number of parallel chains. A fourth model has a minimum estimate of 26.7, which lies close to the cutoff. A possible solution for these effective sample size issues could be to run the models for longer, storing more posterior samples to increase the effective sample size. Given the issues with convergence to the posterior distribution, a longer burn-in period, as well as storing more posterior samples from the MCMC output, could also be recommended. An alternative solution, which is also suggested for the convergence issues where the MCMC is not exploring enough of the distribution, could be to increase the proposal variance in the current MCMC algorithm, or choose a different MCMC algorithm. This is discussed further in Chapter 6.

Because a small proportion of the backward model iterations are affected by these issues (1.8%), the posterior samples of $\boldsymbol{X}_b^*$ for the affected iterations are not used to build the complete posterior distribution of $\boldsymbol{X}^*|\boldsymbol{\phi}, \boldsymbol{Y}^*$. This distribution is instead built using the backward model iterations that do not have convergence or mixing issues. It would be recommended here to explore the alternative solution noted above, so as not to lose information about the backward model, but this is not carried out here, as I had limited time.

The posterior density $p(\boldsymbol{X}^*|\boldsymbol{\phi}, \boldsymbol{Y}^*)$ is displayed in Figure 5.44. The marginal posterior density of X_1^*, corresponding to SE, appears to be more spread out over its prior distribution relative to the spread of the CBD marginal posterior density. The peak level of density (the mode) exists in the region towards the centre of the plot, which is close in proximity to the true value joint posterior densities for multiple groups (estimated in the forward modelling – see Powder 2 and Powder 4 in Figure 4.61). The posterior mode is identified as the pink point, given by $\boldsymbol{X}^*_{\text{mode}} = (0.4590, 0.6130)$, each to 4 decimal places.

The posterior mode of $\boldsymbol{X}^*_{\text{mode}} = (0.4590, 0.6130)$ is now tested using the EIV MOGP posterior, given the posterior distribution of $\boldsymbol{\phi}$ estimated in the forward modelling. That is, predictions of the response variable $\boldsymbol{Y}_{\text{pred},s}$ (for each sample of the posterior distribution) corresponding to the input vector $\boldsymbol{X}^*_{\text{mode}}$ are sampled from the conditional distribution

$$\boldsymbol{Y}_{\text{pred},s}|\boldsymbol{\phi}_s, \boldsymbol{X}^*_{\text{mode}} \sim \text{N}_2(\boldsymbol{m}_s^*, V_s^*)$$

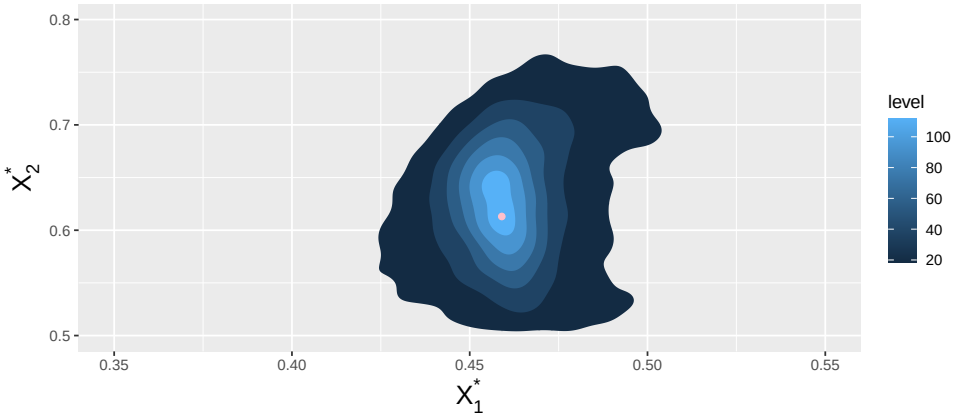

Figure 5.44: The posterior density of $p(\boldsymbol{X}^*|\boldsymbol{\phi}, \boldsymbol{Y}^*)$ given the forward model posterior distribution estimated with the EIV GP, where the pink point represents the posterior mode of $\boldsymbol{X}^*|\boldsymbol{\phi}, \boldsymbol{Y}^*$. The desired response vector is given by $\boldsymbol{Y}^* = (0.51, 0.275)$. This backward model posterior density function is based on the EIV GP modelling.

where

$$\boldsymbol{m}_s^* = (\alpha_{1,s}, \alpha_{2,s})' + \boldsymbol{K}_s(\boldsymbol{X}_{\text{mode}}^*, \tilde{X}_s)V_{\text{MOGP},\tilde{X},s}^{-1}(\text{vec}(\tilde{Y}_s) - (\boldsymbol{\alpha}_{1,s}, \boldsymbol{\alpha}_{2,s})'),$$

and

$$V_s^* = V_{\text{MOGP},\boldsymbol{X}_{\text{mode}}^*} - \boldsymbol{K}_s(\boldsymbol{X}_{\text{mode}}^*, \tilde{X}_s)V_{\text{MOGP},\tilde{X},s}^{-1}\boldsymbol{K}_s(\tilde{X}_s, \boldsymbol{X}_{\text{mode}}^*).$$

For each $s = 1, \ldots, 20000$, one vector of joint predictions of $\boldsymbol{Y}_{\text{pred},s}$ is sampled, and each vector is combined to build the joint distribution of predictions of $\boldsymbol{Y}_{\text{pred}}$, given the forward model posterior distribution $\boldsymbol{\phi}$ and the posterior mode $\boldsymbol{X}_{\text{mode}}^*$. The joint probability density function $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\phi}, \boldsymbol{X}_{\text{mode}}^*)$ is plotted in Figure 5.45. The testing of the posterior mode using the forward model indicates that the backward model has effectively found an input vector which is suitable for producing the desired response vector $\boldsymbol{Y}^* = (0.51, 0.275)$. In context, the backward model for the EIV GP recommends that a value of 0.4590 g/ml and an SE value of 6.130 mJ/g are optimal values for producing the desired values of 0.51 g/ml for tapped density and an angle of 27.5 degrees for angle of repose. The outcomes of the backward modelling for the EIV BR and the EIV GP are compared in Section 5.4.

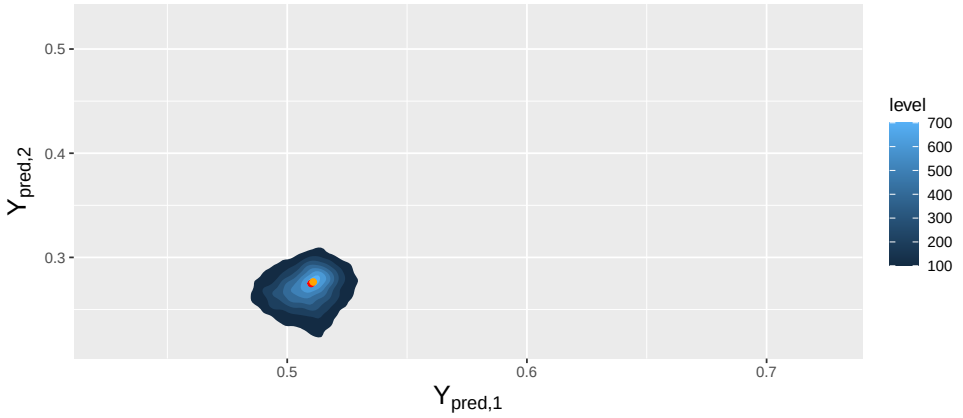

Figure 5.45: The joint probability density function $p(\boldsymbol{Y}_{\text{pred}}|\boldsymbol{\phi}, \boldsymbol{X}^*_{\text{mode}})$ for the real data, given the joint posterior mode of $\boldsymbol{X}^*_{\text{mode}} = (0.4590, 0.6130)$ determined in the backward modelling given the desired response vector of $\boldsymbol{Y}^* = (0.51, 0.275)$. The orange point represents the mode of the joint distribution of predictions, given by $\boldsymbol{Y}_{\text{pred,mode}} = (0.5108, 0.2764)$, and the red point represents the desired response vector $\boldsymbol{Y}^*$. This predictive density function is based on the EIV GP modelling.

5.4 Method comparison using real data

The backward modelling outcomes are now of interest for both methods. The first difference between the methods of EIV BR and EIV GP from the backward modelling was the convergence of the MCMC output for each backward model iteration to its posterior distribution. From Section 5.3.1, the EIV BR backward modelling was able to converge to the posterior distribution for each of the $s = 1, \ldots, 2000$ backward model iterations. This was not the case with the EIV GP, with 32 of the 2000 backward model iterations not showing convergence, as noted in Section 5.3.2. In some of these cases, the backward model iterations are very close to convergence (with psrf upper bounds rounding to values between 1.10 and 1.15), but other cases show a more extreme lack of convergence. This highlights an issue with the sampling in the case of the backward models of the EIV GP (and in fact more complicated EIV BR models), which is discussed further in Chapter 6. Furthermore, relating to the assessment of the MCMC output, an issue of mixing occurs for the EIV GP backward model outputs. This is also discussed further in Chapter 6. For ease of understanding, the posterior density of the corresponding input vector $\boldsymbol{X}^*$ given the desired

response vector $\boldsymbol{Y}^* = (0.51, 0.275)$ (after scaling) **and** given the forward model posterior distribution from EIV BR is given by $p(\boldsymbol{X}^*_{\text{BR}}|\boldsymbol{\theta}_{\text{BR}}, \boldsymbol{Y}^*)$, and the posterior density function $p(\boldsymbol{X}^*_{\text{GP}}|\boldsymbol{\phi}_{\text{GP}}, \boldsymbol{Y}^*)$ is the analogous function given the EIV GP forward model posterior distribution. Moreover, the posterior modes of these distributions are given by $\boldsymbol{X}^*_{\text{BR,mode}}$ and $\boldsymbol{X}^*_{\text{GP,mode}}$. A comparison between the posterior densities $p(\boldsymbol{X}^*_{\text{BR}}|\boldsymbol{\theta}_{\text{BR}}, \boldsymbol{Y}^*)$ (left plot) and $p(\boldsymbol{X}^*_{\text{GP}}|\boldsymbol{\phi}_{\text{GP}}, \boldsymbol{Y}^*)$ (right plot) is provided in Figure 5.46.

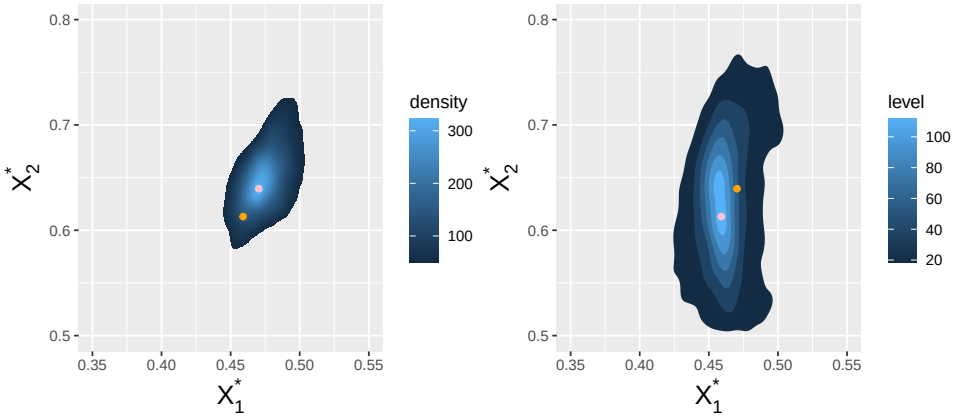

Figure 5.46: A comparison between the posterior densities of $p(\boldsymbol{X}^*_{\text{BR}}|\boldsymbol{\theta}_{\text{BR}}, \boldsymbol{Y}^*)$ (left plot), i.e., the posterior density of the corresponding input vector estimated by the EIV BR backward modelling, given the EIV BR forward model posterior distribution and the desired response vector $\boldsymbol{Y}^* = (0.51, 0.275)$, and $p(\boldsymbol{X}^*_{\text{GP}}|\boldsymbol{\phi}_{\text{GP}}, \boldsymbol{Y}^*)$, the analogous distribution found with the EIV GP. The pink point in each plot is the posterior mode of the respective distribution, and the orange point is the posterior mode found using the other method.

The immediate visual comparison to make is that the posterior density given the EIV BR is much more concentrated than in the case given the EIV GP. More specifically, the marginal posterior density of the SE variable is much more concentrated for the EIV BR. An explanation for this is likely to be related to the observation drawn from Figure 4.61, that there are multiple groups whose true value uncertainty for SE is significantly larger in the case of the EIV GP than for the EIV BR. Note that the posterior true values for the response variables are practically equivalent from the two modelling methods. Furthermore, consider again the comparison between the true values of the response variables inspected in Section 5.3.1, plotted along with the desired response vector $\boldsymbol{Y}^* = (0.51, 0.275)$ in Figure 5.41, which sug-

gested that the corresponding input vector posterior density would be similar to the true values estimated for Powders 2, 3, 4 and 6. It appears that both posterior densities $p(\boldsymbol{X}^*_{\text{BR}}|\boldsymbol{\theta}_{\text{BR}}, \boldsymbol{Y}^*)$ and $p(\boldsymbol{X}^*_{\text{GP}}|\boldsymbol{\phi}_{\text{GP}}, \boldsymbol{Y}^*)$ closely resemble the joint posterior densities of the true values of these powders for CBD and SE from Figure 4.61. Finally, it is clear that the two modelling methods have found distinct candidate vectors for $\boldsymbol{X}^*$, with the posterior density in the EIV BR case concentrated around its mode of $\boldsymbol{X}^*_{\text{BR,mode}} = (0.4704, 0.6394)$, and in the EIV GP case, it appears that there is an optimal 'range' of values for SE between 0.6 and roughly 0.65, given values of CBD between 0.45 and 0.46, with the posterior mode of $\boldsymbol{X}^*_{\text{GP,mode}} = (0.4590, 0.6130)$.

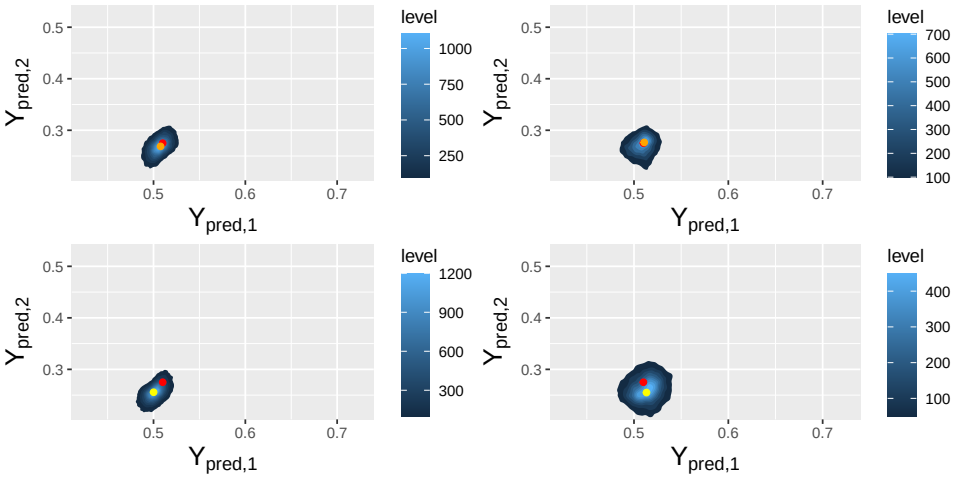

Figure 5.47: Four plots of the joint probability density function for predictions of the response variable given the forward modelling and the posterior mode of the corresponding input vector $\boldsymbol{X}^*$. The first column of plots corresponds to the predictive density of $\boldsymbol{Y}_{\text{pred}}$ given the EIV BR modelling, and the second column corresponds to the analogous density given the EIV GP modelling. The first row of plots demonstrate these densities given the posterior mode of $\boldsymbol{X}^*$ for the corresponding backward modelling (e.g., the plot in row 1, column 1 is the density function given $\boldsymbol{\theta}_{\text{BR}}, \boldsymbol{X}^*_{\text{BR,mode}}$), whereas the second row shows $\boldsymbol{Y}_{\text{pred}}$ given the posterior mode of $\boldsymbol{X}^*$ from the other modelling method. The red point in each plot represents $\boldsymbol{Y}^*$, the orange points represent the predictive mode of the responses given the corresponding backward modelling, and the yellow points represent the predictive mode of the responses given the other backward modelling.

In both cases, the joint distribution of predictions of the response vari-

ables given the posterior modes of $\boldsymbol{X}^*$ and the forward model posterior distribution was estimated in Sections 5.3.1 and 5.3.2 for the EIV BR and EIV GP modelling respectively. The joint probability density functions of these predictions are provided here in Figure 5.47 (first row of plots, with EIV BR density function on the left and EIV GP on the right), above the joint probability density functions of predictions from each modelling method, given the posterior mode identified by the other modelling method. Comparing the plots from the first row, it is noted that the mode of the predictive distribution for the EIV GP (orange dot in top-right plot) is closer to $\boldsymbol{Y}^*$ (red dot in all plots) than that for the EIV BR (orange dot in top-left plot). The Euclidean distances between the mode of the joint prediction distribution and the desired response vector is calculated in both cases. For the EIV BR case, a distance of 0.00669 to 3 significant figures is calculated, given a difference of 0.00233 g/ml between TD values (i.e., between the first element of the mode of the joint prediction distribution and the desired TD value), and a difference of 0.00627 degrees between AoR (over 100) values. In comparison, the distance between the mode of the joint predictive distribution given the EIV GP modelling and the desired response vector is 0.00167 to 3 significant figures, given differences of 0.000845 g/ml for TD and 0.00143 degrees for AoR (over 100). Using the joint predictive distribution in each case, the probability of being within a given distance of the desired response vector can also be calculated. As noted in Section 5.3.1, it is desirable for the value of SE to be within the range of 25 and 30 degrees (0.25 and 0.3 after scaling); given the EIV BR modelling and $\boldsymbol{X}^*_{\text{BR,mode}}$, the probability that AoR lies in this range is 64%, compared with 71% given the EIV GP modelling and $\boldsymbol{X}^*_{\text{GP,mode}}$. So, for multiple comparisons between the two modelling methods, it appears that the EIV GP has performed better in the backward modelling.

Another comparison between the modelling methods is made when considering the correlation between the predictions of TD and of AoR. From Figure 5.47, it is clear that there is some positive correlation between the estimates of TD and AoR given either of the posterior modes. On the other hand, there appears to be practically zero correlation between the predictions of these variables in the EIV GP modelling cases. This can be investigated further, by estimating the correlation between predictions of the response variables from Figures 4.59 and 4.60 given some input vectors. For example, the plot in Figure 5.48 provides the correlation between TD and AoR given the input vectors where $X_1 \in [0.35, 0.55]$ and $X_2 = -1.5X_1 + 1.325$ (with the correlation plotted against the first input variable CBD). The black line corresponds to the correlation in the EIV BR,

and the red line corresponds to the correlation in the EIV MOGP. It is clear that the correlation between TD and AoR is captured effectively in the EIV BR fitted model, whereas practically zero correlation is captured in the EIV MOGP.

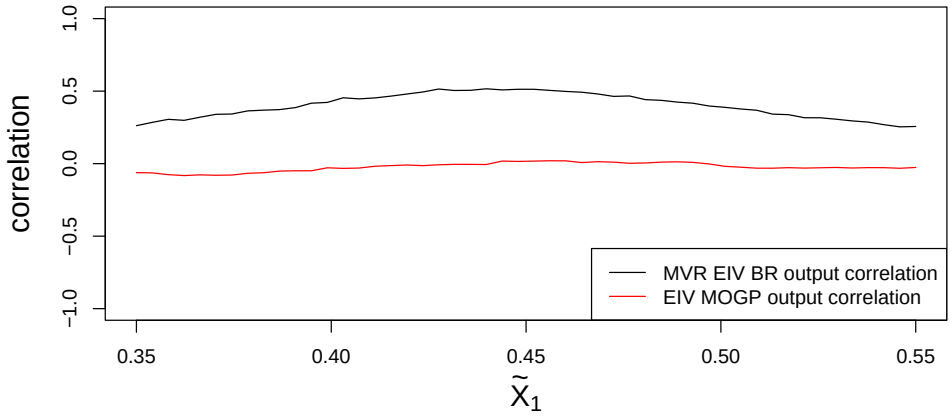

Figure 5.48: A plot of the correlation between TD and AoR at fifty input vectors, created by a sequence of CBD values in the range [0.35,0.55], with SE values equal to $X_2 = -1.5X_1 + 1.325$. The black line corresponds to the correlation in the EIV BR, and the red line in the EIV GP.

The second row of plots in Figure 5.47 displays the density of predictions given the posterior mode estimated in the other modelling method. The first of these plots, corresponding to predictions of response variables from the EIV BR forward modelling given the posterior mode $\boldsymbol{X}^*_{\text{GP,mode}}$, is relatively similar to the plot directly above, of the joint probability density of predictions given $\boldsymbol{X}^*_{\text{BR,mode}}$. This is unsurprising, given the additive relationship in the linear model, and the relative distance between the posterior modes $\boldsymbol{X}^*_{\text{BR,mode}} = (0.4704, 0.6394)$ and $\boldsymbol{X}^*_{\text{GP,mode}} = (0.4590, 0.6130)$ (note that both elements of the EIV BR mode are larger than those of the EIV GP mode). Furthermore, the marginal joint posterior densities of the slope terms from the EIV BR forward modelling are provided in Figure 5.49 to highlight the relationship between the influence of CBD and the influence of SE on each response variable. The left plot of Figure 5.49 provides the joint posterior density of the model coefficients $\beta_{\text{SE,TD}}$ and $\beta_{\text{CBD,TD}}$ (that is, the slope terms of SE and CBD in the linear predictor for TD), and the right plot provides the analogous joint posterior density for the slope terms in the linear predictor for AoR. Evidently, there is a strong positive relationship between the slope terms in both cases. The highest-density region for the

slopes relating to TD show that a change of value of CBD has a larger influence than a change of value of SE, so in order to produce larger values of TD, it is most likely that a larger value of CBD is required. Similarly, the highest density region of the slopes relating to AoR indicate that changes of SE have a larger influence than changes in CBD, thus to produce larger values of AoR, it is most likely that a larger value of SE is required. Further evidence of this is found when comparing the plots of the first column of Figure 5.47; both elements of the mode $\boldsymbol{X}^*_{\text{GP,mode}}$ are smaller than those of $\boldsymbol{X}^*_{\text{BR,mode}}$, hence the predictions of TD and AoR are both smaller given $\boldsymbol{X}^*_{\text{GP,mode}}$ than given $\boldsymbol{X}^*_{\text{BR,mode}}$. Furthermore, this would suggest a better candidate value from the posterior of $\boldsymbol{X}^*_{\text{BR}}|\boldsymbol{Y}^*, \boldsymbol{\theta}_{\text{BR}}$ could be chosen with a larger CBD and a larger SE.

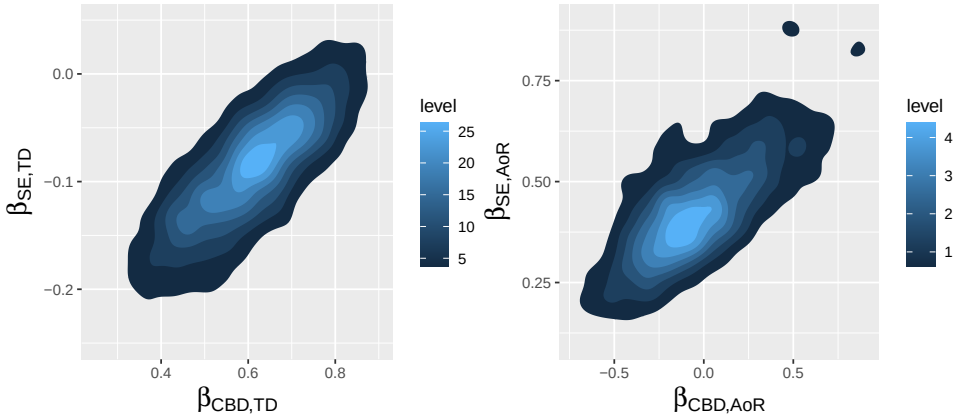

Figure 5.49: Joint posterior densities of the slope terms in each linear predictor in the EIV BR model, fitted with informed priors with TD and AoR as the response variables and CBD and SE as the input variables. The left plot corresponds to the slopes in the linear predictor of TD, and the right plot to the slopes in the linear predictor of AoR.

In the following chapter, possible continuations of the work are discussed, relating to either issues that arise in this work that should be resolved, or suggestions of new avenues on which this work could be explored.

Chapter 6

Further Work

This penultimate chapter details the possible avenues on which the work in this thesis could be progressed. The avenues are divided into two sections; the first section highlights the possible development of the existing methods applied already in this work, while the second section introduces new topics that have not yet been explored in this thesis.

6.1 Development of existing work

The first development noted here relates to the issues of convergence (as well as mixing) that primarily exist in cases of multimodal posterior distributions. The examples of this issue in this work occur in the backward modelling with the simulation examples, where there are multiple 'optimal' input values (or vectors for cases of multiple explanatory variables), which induces multimodality in the posterior distribution of X^* (or $\boldsymbol{X}^*$). Ultimately, the reason that the model does not 'converge' is that the MCMC sampler becomes stuck in a particular input subspace (i.e., in the immediate vicinity of one of the possible optimal input values) for the entirety of the sampling. Since the sampler is stuck in the vicinity of an optimal input value, the sampler is, in a sense, not tempted to explore the input space further, and so other optimal input values are not captured by the model. This leads to a lack of convergence, readily detected by the PSRF, when the parallel chains for the MCMC converge to different optimal input values, and each chain does not explore other areas of the input space. For examples of this in this work, see Sections 5.1.2 and 5.2. As noted in these sections, a possible fix that should be applied in these cases is to use a sampler that implements tempered transitions, which forces the sampler to explore the

other modes of the posterior distribution. This can be applied in JAGS (the software that applies MCMC simulation utilised in this work) using its mix module (see Plummer et al. (2003)).

An alternative solution to the convergence issues would be to utilise the Hamiltonian Monte Carlo (HMC) algorithm, as opposed to the Gibbs sampler that is applied in this work. The HMC algorithm is discussed extensively in Neal (2011) and Betancourt (2017), and implements a physics-based sampling method, with the main goal of improving the directionality of the sampling, particularly in high-dimensional problems (models with a large number of stochastic parameters) using a momentum parameter. Most notably, a variation of this algorithm called NUTS (No-U-Turn sampler) is implemented in the computing software *Stan* (see Gelman et al. (2015)). In Lalchand & Rasmussen (2020), it is noted that the HMC algorithms have an advantage over other MCMC algorithms since the random-walk behaviour does not feature. Conversely, there are other papers which, when comparing the statistical software JAGS (variants of Metropolis-Hastings algorithms) with Stan, suggest that this is either case dependent, or simply the algorithms are hard to separate in terms of performance. For example, Hecht et al. (2021) claim that, for the multi-level intercept-only model, given by

$$y_{i,j} \sim \mathrm{N}(\theta_i, \sigma_\epsilon^2), \theta_i \sim \mathrm{N}(\mu, \sigma_\theta^2),$$

JAGS performs better (according to the metric effective sample size per second to measure MCMC efficiency) when considering the classic parameterisation given above; on the other hand, Stan performs better for this metric in the covariance- and mean-based parameterisation (an alternative parameterisation which is omitted here, but can be found in Hecht et al. (2021)). There is some suggestion in Monnahan et al. (2017) that Stan performs better when fitting more complex models, which was investigated using population ecology models (see Monnahan et al. (2017) for more details). Given the multimodal posterior distribution issues that occur in some examples in this work (considering simulated data), and the work in the literature recommending HMC when estimating parameters in high dimensions, it is recommended that the continuation of this work considers applying this algorithm.

It is noted that the model assessment for the EIV BR is different to that for the EIV GP, in that the approximate LOO-CV-IC methods are not the same. The Pareto-smoothed importance sampling LOO-CV-IC (PSIS-LOO-CV-IC) is applied to the EIV BR to compare predictive performances of models, whereas the mixed LOO-CV-IC is utilised for the EIV GP. While

these approximations are beneficial simply by saving time, they both have drawbacks. For the EIV BR, the PSIS-LOO-CV-IC hinges on the estimates for the shape parameter k for the generalised Pareto distribution, which is applied to smooth the upper tail of importance ratios in the importance sampling. The work in Vehtari et al. (2015) demonstrates that, for estimates of the shape parameter greater than 0.7, the number of required posterior samples for convergence of the importance weights becomes significantly large (it is unclear exactly how large this is, but a graph in the paper suggests it is greater than 10^5 for a shape parameter estimate of 0.7, which quickly becomes 10^{10} for a shape parameter estimate of 0.8, which is clearly infeasible). Without convergence, the smoothing of the importance ratios becomes unreliable, thus the estimate of predictive performance becomes unreliable. As noted in Section 4.4.3, the estimates of the parameter k are all greater than 0.7. For the EIV GP, the mixed LOO-CV-IC based on $n_g = 7$ groups has the potential to provide an overestimate of the exact LOO-CV-IC predictive performance. With the limited size of the data set, the estimates of hyperparameters such as the model error standard deviation σ_ϵ (for a single output variable), the distance-scaling parameter l and the signal standard deviation parameter σ_k could lead to overly optimistic estimates of the response variables in the LOO GP posterior.

Generally speaking, while it is recommended to use the approximate versions of LOO-CV-IC for larger data sets, the size of this data set with $n_g = 7$ could indicate that an exact LOO-CV-IC could be used to estimate predictive performance. However, the errors-in-variables data structure causes further issues with computation time in the exact LOO-CV-IC, for both the EIV BR and EIV GP modelling methods. In the exact LOO-CV-IC method *without* consideration of measurement error on both the explanatory variables and response variables, the statistical model should be fitted n times, in each case estimating the LOO posterior distribution given $n-1$ data points, and assessing predictive performance by testing the model on the 1 'left-out' data point. With consideration of the hierarchical modelling structure (as well as replicate measurements), e.g. with the EIV BR, this number of model fits becomes $n_g S$, for S posterior samples from the LOO posterior distribution; the true values of each 'left-out' (test) group must be estimated given the LOO posterior distribution. Therefore, for each of the $i = 1, \ldots, n_g$ LOO model fits (i.e., for each estimation of the posterior distribution given $n_g - 1$ training groups), a subsequent model must be fitted in order to estimate the true values of the test group (whilst simultaneously estimating the log-likelihood for predictive performance). Furthermore, to properly account for the uncertainty in the LOO posterior distribution, the

subsequent model must be fitted given a large number of posterior samples S.

Evidently this is time consuming, and so an alternative is sought. One possibility is to fit the model $n_g + 1$ times, with n_g fits corresponding to fitting the model with $n_g - 1$ groups (i.e., leaving out a different group each time), and the final fit corresponding to fitting the model with all n_g groups, which would provide the true values for each test group. This would not have an influence on the fitted model for each of the n_g LOO posterior distributions, which are all trained given $n_g - 1$ groups. There is some chance that the posterior density of the true values of the test group, given the model fit with n_g groups, differs significantly from the analogous density given the model fit with $n_g - 1$ groups. That is, the posterior density of the true values $p(\tilde{Y}_{i'}, \tilde{X}_{i'} | \tilde{\boldsymbol{Y}}_{-i'}, \tilde{\boldsymbol{X}}_{-i'})$ may differ from the posterior density $p(\tilde{Y}_{i'}, \tilde{X}_{i'} | \tilde{\boldsymbol{Y}}_{\text{LOO},-i'}, \tilde{\boldsymbol{X}}_{\text{LOO},-i'})$, with the subscript LOO indicating the training values are estimated in the LOO setting, and the subscripts $i = i'$ indicating the test group, and $-i'$ indicating the training groups. Estimating the true values of the test group i' using the model fitted with all groups is likely to be an overconfident estimate of the true values of the test group. Alternatively, the model could be fitted n_g times, each time training on a different combination of the $n_g - 1$ groups, as is standard for the exact LOO-CV-IC. Then, estimate the distributions of true values $(\tilde{Y}_{i'}, \tilde{X}_{i'})$ for the test group i' by sampling from the true values of the group from the $n_g - 1$ model fits in which the test group i' is a training group. In this case, the estimate of $(\tilde{Y}_i, \tilde{X}_i)$ is likely to be an underconfident estimate of the true values of the test group. Further experimentation would be required to see which alternative is more appropriate, but both cases are likely to be better estimates of the predictive performance than the approximate cases, while still reducing the time consumption of the model comparison process relative to the exact case significantly.

Given that the models in this work are fitted to data with only seven groups (so the relationship between the variables must be found with seven data points), it seems logical to suggest that the analysis should be repeated with more data. With the current data, the overfitting issues with the more complicated models restricts the modelling process, with only relatively few models (simpler models) in consideration. It is plausible that these more complicated models (with more explanatory variables and/or more polynomial terms) are desirable, as these would provide more possibilities for engineering insight into model structure.

Given the application of the modelling process (both forward and backward), this work identifies values for the key input variables of CBD and SE

of 0.4704 g/ml and 6.394 mJ/g respectively as optimal (based on the available data and the EIV BR modelling with informed prior distributions) with respect to producing a powder which flows optimally and deposits optimally onto a powder bed. While the limitations of the amount of data and of possible 'excluded' input variables from the FT4 powder rheometer relating to shear cell testing and wall friction tests must be noted, the backward modelling could be tested in a literal sense by asking a powder manufacturer to design a powder based on the qualities from the inputs noted above, to see if the powder possesses the qualities from the outputs noted above, providing an effective powder for Laser Sintering.

It was also established from Section 2.3 that other powder properties that are not investigated here may have an influence on the powder flow properties. More specifically, the work of Vetterli (2019) considers the influence of 'intrinsic' and 'extrinsic' particle properties on the powder flow properties. It is possible that the properties measured by the powder rheometer do not account for these, hence they could somehow be explicitly incorporated into the modelling.

6.2 New avenues to explore

The possible collection of more data noted above leads on to further possibilities with the overarching theme of experimental design. Experimental design is a statistical approach to data collection which involves identifying the best possible settings, samples and conditions for an experiment, in order to maximise the amount of information obtained. Obviously, the amount of information that can be obtained from an experiment correlates with the number of observations of the experiment; simply put, to gain more information, we can take more observations of the experiment. However, in practical terms this is not always feasible, usually because of time and cost constraints. In order to gain the most amount of information while trying, in a sense, to minimise the time and cost of the experiment, an experimental design is developed. A simple example could be to identify values of key input variables, possibly input setting of the Laser Sintering machine, or related to the powders considered in this work, which could be used in experiments in order to reduce the uncertainty on the quality of a 3D-printed object from Laser Sintering. Experimental design could lead into new areas to be explored in AM, by suggesting sets of inputs which have quite uncertain effects and may potentially lead to desirable output values.

The idea of designing experiments leads to a further possibility, which

has notably been explored very recently in multiple papers in the AM literature (see Section 2.3), that being the use of Bayesian optimisation (BO), a process for arriving at optimal decisions about future input values for an experiment specifically in order to learn the most information about maximising an output variable. This links to the work in this thesis, as the typical method for learning about the behaviour of a function in BO is Gaussian processes. Once the function relating the output variable(s) to the input variables is estimated, one has the choice of multiple methods for discovering the values of the input variables which is most likely to provide the largest improvement in understanding the function. These methods include expected improvement and knowledge gradient, which are discussed further in Frazier & Wang (2016).

Chapter 7

Conclusions

The work set out in this thesis provides a statistically-sound modelling process for investigating AM-related data, with two specific aims: firstly, identifying which combination of input variables provides the best understanding of the response variables, and, secondly, 'inverting' this relationship to find optimal values for these input variables to produce desired values of the responses. Two modelling processes were investigated—errors-in-variables Bayesian regression (EIV BR), and errors-in-variables Gaussian processes (EIV GP)—with which both of the aims above were achieved. These methods differ in their assumptions of the relationship between the response variables and the output variables. The parametric method of EIV BR assumes a predetermined relationship between the variables that is linear in its model coefficients, whereas the the nonparametric method of EIV GP makes no such assumption. Both modelling methods considered multiple explanatory variables and multiple response variables simultaneously to answer questions regarding both of the aims above.

The first of two notable elements of these modelling methods is the fully-Bayesian setting in which the models were fitted, that is, treating all parameters in the modelling as random variables and estimating them with a posterior distribution. Secondly, the structure of observed data of these explanatory variables and response variables in this work was that there were multiple 'replicate' measurements on multiple groups for each of the variables, with the measurements known to be affected by some error. As mentioned above, these structures were dealt with by implementing errors-in-variables models, which are handled capably by the Bayesian modelling, which treats the modelling parameters as random variables (therefore, it is straightforward to estimate 'true' values for each of the groups).

Before applying these methods to real-world data, both were 'trained' extensively using simulated data, assuring that the modelling processes were performing as intended. For the EIV BR, this training process proceeded by simulating data for several possible linear predictors, each differing in their complexity, and comparing the posterior distribution from the fitted models with the 'chosen' values from the underlying simulation (such as the model coefficients, the measurement error variability, etc.). In cases where there were clear issues with the modelling, the issues were further investigated by considering the posterior densities of particular parameters, and appropriate adjustments to the prior distributions were implemented to improve the model fit. For the EIV GP, multiple simulations were considered with differing underlying functions that related the response variable(s) to the input variable(s), where the underlying functions were difficult for the EIV GP to recover. These functions represented extreme cases, and so the modelling process performing as intended provided evidence for the EIV GP to be successful in any case. Again, appropriate adjustments to prior distributions were carried out where necessary. Furthermore, the backward modelling processes were also trained using simulated data, with 'would-be' desired response values selected based on the 'chosen' true values from the underlying simulation, again to test the model with extreme cases. In both the forward and backward modelling, for both modelling methods, and with simulated data, it was demonstrated that the relationships between the response variable(s) and input variable(s) were recovered effectively, as well as the candidate values of the input variable(s) for the given desired response(s).

Moreover, this modelling process has been applied to real-world data, relating to powder behaviour in the AM process Laser Sintering. The two output variables that were utilised in the modelling were tapped density (TD) and angle of repose (AoR), which are good proxies for powder bed deposition and powder flow respectively. By comparing several EIV BR models, each with different combinations of the input variables, it was concluded that the best-fitting EIV BR model is the two-covariate model which assumed an additive relationship between conditioned bulk density (CBD) and specific energy (SE). This was concluded with the help of approximate LOO-CV-IC to evaluate predictive performance and compare this between models. Moreover, these variables were selected given fears of overfitting with more complicated models. The EIV GP with CBD and SE as the input variables was also fitted for comparison; its predictive performance (evaluated using mixed LOO-CV-IC) was inferior to that of the EIV BR model. In addition, due to the possible overfitting issues with more com-

plicated models (given the models were fitted based on seven groups, thus seven true values for each variable), the EIV BR modelling is preferred in this case. This worry of overfitting also contributed to the decision to choose the two-covariate EIV BR model over a more complicated model with six covariates, despite the better performance according to approximate LOO-CV-IC of the six-covariate model.

In order to demonstrate the most complete form of Bayesian modelling, an elicitation was carried out to gather expert information regarding aspects of the real data, using which particular prior distributions in the model were created. The expert information comprised ranges of the variables TD, AoR, CBD, SE and basic flowability energy (BFE) for which 95% of the possible powders could lie within and ranges of those variables for which 95% of *subsamples* of a given powder could lie within. Informed prior distributions were built for the measurement error precision, the between-materials precision of the input variables and the model coefficients, based on this information. More appropriate prior distributions for those variables on which no prior information was acquired were also developed. The improvements in the model fits were noted both visually and numerically, namely with smaller model error standard deviations being estimated with the informed priors.

Having identified the best-fitting 'forward' model, the backward modelling process was carried out. It was identified that desired values for the response variables were 0.51 g/ml for TD and 27.5 degrees for AoR (where values between 25 and 30 degrees were also acceptable). Applying the backward modelling process using the EIV BR with uniform prior distributions over the elicited ranges of the variables CBD and SE, the 'optimal' corresponding values to the desired response vector were estimated to be 0.4704 g/ml and 6.394 mJ/g respectively. Having assumed these optimal values to be the candidate vector for $\boldsymbol{X}^*$, the candidate vector was tested using the forward modelling, to provide a distribution of predictions for the response variables. The mode of this distribution was found to be 0.5074 g/ml and 26.81 degrees for TD and AoR respectively. Additionally, the EIV GP backward modelling was also carried out for demonstration, and further comparison with the EIV BR. The candidate vector (i.e., the posterior mode of $\boldsymbol{X}^*$) identified in the backward modelling using the EIV GP contained values of 0.4590 g/ml and 6.130 mJ/g for CBD and SE respectively. An indication of the quality of the backward modelling was provided by running this candidate vector through the forward modelling for the EIV GP, producing a mode for the distribution of response predictions of 0.5108 g/ml and 27.64 degrees for TD and AoR respectively.

At stages throughout this thesis, limitations of the work were highlighted. These limitations, which are discussed previously in Chapter 6, included the small data set and its consequence of overfitting, the differences between the approximate methods of model evaluation for the EIV BR and the EIV GP, the Pareto smoothing parameter estimates in the approximate LOO-CV-IC, and the selection of variables used from the FT4 powder rheometer. Most notably, due to the dangers of overfitting, EIV BR modelling is preferred to EIV GP modelling for data sets of the size and complexity that were considered here. It is noted that the FT4 powder rheometer offers scope for collecting further data which could also play a role in the continuation of this case study.

In conjunction with the observations made about the existing statistical research in AM, there are clearly several possibilities for applying these robust methods to other aspects of either laser sintering, or even other AM methodologies. The most logical application for these methods would be to investigate the relationship between final-part properties (for example, tensile strength and Young's modulus simultaneously) and processing parameters. While processing parameters relating to printer settings are exact (i.e., there is no measurement error), other processing parameters not related to printer settings, even the variables tapped density and angle of repose investigated here, could be incorporated, which are clearly both measured with error, and influence the final-part properties.

Finally, the next steps set out for this work are clearly laid out in Chapter 6. Given the forward and backward modelling outcomes, a practical test could be implemented by organising the manufacturing of a powder with CBD and SE values outlined above, with the aim of producing a powder which can flow optimally through chutes and hoppers and deposit optimally onto a powder bed. Further data collection, and an investigation of FT4 powder rheometer variables not considered here using the modelling process outlined in this work, would be recommended prior to the test. Another AM-related data collection could be carried out in order to apply the modelling process to optimise other aspects of AM or Laser Sintering, such as the example described in Chapter 6. Additionally, this introduces a possibility to contribute to the statistical research within AM by considering an application of experimental design to real-world problem.

Appendix A

Exact and Approximate leave-one-out cross-validation

This section gives a more detailed demonstration of evaluating Bayesian model fits using leave-one-out cross-validation information criterion (LOO-CV-IC), with both the exact method, and the approximate method using Pareto smoothing.

A.1 Exact leave-one-out cross-validation for EIV BR

The exact LOO-CV-IC is demonstrated here using the simple linear model *without* measurement error on either the response or explanatory variable, defined by

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

for $i = 1, \ldots, n$, where $\epsilon_i \sim \mathrm{N}(0, \tau_\epsilon)$. The parameter vector for this model (assuming no hyperparameters in the modelling) is

$$\boldsymbol{\theta} = (\beta_0, \beta_1, \tau_\epsilon).$$

Firstly, define μ_i to be the mean of this regression model, that is,

$$\mu_i = \mathrm{E}(Y_i) = \beta_0 + \beta_1 x_i. \tag{A.1.0.1}$$

Then, it follows from above that Y_i conditioned on $\boldsymbol{\theta}$ is normally distributed with mean μ_i and precision τ_ϵ. That is,

$$Y_i|\boldsymbol{\theta} \sim \mathrm{N}(\mu_i, \tau_\epsilon). \tag{A.1.0.2}$$

The log-likelihood of the value Y_i given $\boldsymbol{\theta}$ can be investigated, that is, the values

$$\log p(Y_i|\boldsymbol{\theta}).$$

The additional notation of Y_{new} as a future observation of the response variable is introduced, with some corresponding explanatory variable value X_{new}, assumed to be not known. Note that the model has not been trained on the pair of values $(Y_{\text{new}}, X_{\text{new}})$, so the idea here is to calculate some out-of-sample predictive accuracy of the model for external validation, which is given by the log predictive density for Y_{new}; that is,

$$\begin{aligned} \log(p_{\text{post}}(Y_{\text{new}}|X_{\text{new}})) &= \log(\mathrm{E}_{\text{post}}(p(Y_{\text{new}}|\boldsymbol{\theta}, X_{\text{new}})) \\ &= \log \int p(Y_{\text{new}}|\boldsymbol{\theta}, X_{\text{new}}) p_{\text{post}}(\boldsymbol{\theta}) d\boldsymbol{\theta}, \end{aligned} \tag{A.1.0.3}$$

where $p_{\text{post}}(Y_{\text{new}}|X_{\text{new}})$ represents the predictive density of the new data point Y_{new} given the joint posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{x})$ and the corresponding new explanatory variable value X_{new}. The subscript 'post' refers to averaging over the posterior distribution of $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ is a random variable, the log predictive density of Y_{new} is not conditioned on some point estimate of $\boldsymbol{\theta}$, rather the expectation averaging over the entire posterior is taken to account for the uncertainty in the posterior.

Then, since the joint future data point $(Y_{\text{new}}, X_{\text{new}})$ is unknown, the expected (out-of-sample) log predictive density for a single new data point Y_{new} is defined to be

$$\begin{aligned} \text{elpd} &= \mathrm{E}_p(\log(p_{\text{post}}(Y_{\text{new}}|X_{\text{new}})) \\ &= \int \int (\log(p_{\text{post}}(Y_{\text{new}}|X_{\text{new}})) p(Y_{\text{new}}, X_{\text{new}}) dY_{\text{new}} dX_{\text{new}}. \end{aligned} \tag{A.1.0.4}$$

Further, the expected log predictive density for a vector of out-of-sample data points $Y_{\text{new},1}, \ldots Y_{\text{new},n_{\text{new}}}$ given the corresponding explanatory variable values $X_{\text{new},1}, \ldots, X_{\text{new},n_{\text{new}}}$, is defined by the expected log pointwise predictive density given by the sum of the expected log predictive densities for each point, i.e.,

$$\text{elppd} = \sum_{m=1}^{n_{\text{new}}} \mathrm{E}_p(\log(p_{\text{post}}(Y_{\text{new},m}|X_{\text{new},m}))). \tag{A.1.0.5}$$

In an ideal world, some form of the elpd in Equation A.1.0.4 (or elppd in Equation A.1.0.5) would be used to assess the model fit, but the distribution of the new observation Y_{new} is not known, so an alternative based

on something known must be used. That is, some evaluation of the model using the log-likelihoods $\log(p(Y_i|\boldsymbol{\theta}))$.

Since the parameter vector $\boldsymbol{\theta}$ is not known, the log-likelihood $\log(p(Y_i|\boldsymbol{\theta}))$ is also not known, so it is estimated with the help of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{x})$. To evaluate the predictive accuracy of the fitted model, the log pointwise predictive density, given by,

$$\begin{aligned}\text{lppd} &= \log\left(\prod_{i=1}^{n} p_{\text{post}}(Y_i|x_i)\right) = \sum_{i=1}^{n} \log\left(\int p(Y_i|\boldsymbol{\theta}, x_i) p_{\text{post}}(\boldsymbol{\theta}) d\boldsymbol{\theta}\right) \\ &= \sum_{i=1}^{n} \log(p(Y_i|x_i)),\end{aligned} \tag{A.1.0.6}$$

is estimated by the computed log pointwise predictive density, using posterior samples $\boldsymbol{\theta}_s$ for $s = 1, \ldots, S$, where S is the (large) number of samples taken from the posterior distribution. The computed lppd is

$$\text{computed lppd} = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(Y_i|\boldsymbol{\theta}_s, x_i)\right). \tag{A.1.0.7}$$

For large enough S, the lppd and the computed lppd (Equations A.1.0.6 and A.1.0.7 respectively) should be equivalent, as the posterior samples should be large enough to capture the complete posterior distribution.

The computed lppd in Equation A.1.0.7 is a good starting point as an estimate of the elpd. Using the response variable data Y_i in this way will overestimate the elppd for future data (from Equation A.1.0.4 or A.1.0.5) since the model has been built on this data, so an adjustment needs to be made to estimate out-of-sample predictive performance while using the data Y_i. This comes with the help of information criteria, and in particular the preference here is to consider some form of cross-validation, in particular leave-one-out cross-validation.

To calculate the exact LOO-CV-IC involves fitting the model n times (with n being the number of data points), each time leaving out one of the data points to use as a 'test' data point, and training the model on the remaining data points. This introduces n posterior distributions $p(\boldsymbol{\theta}_{-i}|\boldsymbol{Y}_{-i}, \boldsymbol{x}_{-i})$, where $\boldsymbol{\theta}_{-i}$ is defined to be the parameter vector

$$\boldsymbol{\theta}_{-i} = (\beta_{0,-i}, \beta_{1,-i}, \tau_{\epsilon,-i}), \tag{A.1.0.8}$$

and the vectors $\boldsymbol{Y}_{-i}$ and $\boldsymbol{x}_{-i}$ are the observed data without the i^{th} data point. That is,

$$\boldsymbol{Y}_{-i} = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)$$

and

$$\boldsymbol{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Each posterior distribution is summarised by some large number S of posterior samples, leading to the log pointwise predictive density

$$\begin{aligned} \text{lppd}_{\text{LOO-CV}} &= \sum_{i=1}^{n} \log(p_{(\text{post},-i)}(Y_i|x_i)) \\ &= \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(Y_i|\boldsymbol{\theta}_{-i,s}, x_i)\right). \end{aligned} \tag{A.1.0.9}$$

The density $p_{(\text{post},-i)}(Y_i)$ represents the predictive density of the 'test' data point Y_i given the posterior distribution $p(\boldsymbol{\theta}_{-i}|\boldsymbol{Y}_{-i}, \boldsymbol{x}_{-i})$. Finally, the LOO-CV-IC is calculated by multiplying the $\text{lppd}_{\text{LOO-CV}}$ by -2, converting the metric onto the deviance scale. That is,

$$\text{LOO-CV-IC} = -2 \times \text{lppd}_{\text{LOO-CV}} \tag{A.1.0.10}$$

An estimate of the effective number of parameters is given by (for easier comparison with other information criteria)

$$p_{\text{LOO-CV}} = \text{lppd} - \text{lppd}_{\text{LOO-CV}}. \tag{A.1.0.11}$$

Some adjustments are required for the exact LOO-CV and to account for errors-in-variables in both the response variable and explanatory variable. That is, consider now the errors-in-variables Bayesian regression with one input variable with replicate observations on each group, defined in Equations 3.2.3.1 and 3.2.3.2, with the parameter vector $\boldsymbol{\theta}$ defined in Equation 3.3.0.1.

The first main difference to be aware of is that the response variable data Y_i from the non-EIV case are replaced with the true, unobservable values $\tilde{Y}_i$, which must be estimated in the posterior distribution. A subset of the parameter vector $\boldsymbol{\theta}$ (corresponding to the parameter vector defined in Equation 3.3.0.1) must be introduced, since the true values $\tilde{Y}_i$ are part of the parameter vector, and so the conditional statement of $\tilde{Y}_i|\tilde{Y}_i$ occurs when conditioning the true values of the response variable on $\boldsymbol{\theta}$. The parameter vector $\boldsymbol{\phi}_i = (\beta_0, \beta_1, \tilde{X}_i, \tau_\epsilon)$ is therefore introduced, with the log-likelihood of the true value $\tilde{Y}_i$ given $\boldsymbol{\phi}_i$ now of interest. That is, the values

$$\log p(\tilde{Y}_i|\boldsymbol{\phi}_i),$$

for $i = 1, \ldots, n_g$. Moreover, the expectation of $\tilde{Y}_i$ is defined to be $\widetilde{\mu}_i$, i.e.

$$\widetilde{\mu}_i = \mathrm{E}(\tilde{Y}_i) = \beta_0 + \beta_1 \tilde{X}_i, \tag{A.1.0.12}$$

and so the log-likelihood is computed assuming $\widetilde{\mu}_i$ as the mean and τ_ϵ as the precision of the normal distribution.

An adjustment to the log pointwise predictive density from Equation A.1.0.6 is then provided in Equation A.1.0.13, given by

$$\begin{aligned} \mathrm{lppd}_{\mathrm{EIV}} &= \log\left(\prod_{i=1}^{n_g} p_{\mathrm{post}}(\tilde{Y}_i)\right) = \sum_{i=1}^{n_g} \log\left(\int p(\tilde{Y}_i|\boldsymbol{\phi}_i) p_{\mathrm{post}}(\boldsymbol{\phi}_i) d\boldsymbol{\phi}_i\right) \\ &= \sum_{i=1}^{n_g} \log(p(\tilde{Y}_i|Y, X)). \end{aligned} \tag{A.1.0.13}$$

This is then estimated by the computed lppd accounting for measurement error, given by

$$\text{computed } \mathrm{lppd}_{\mathrm{EIV}} = \sum_{i=1}^{n_g} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(\tilde{Y}_{i,s}|\boldsymbol{\phi}_{i,s})\right).$$

In the case of this EIV Bayesian regression model, the exact LOO-CV is slightly more complicated based on the need to estimate the unobservable true values on the response variable, as well as the covariate. That is, in the exact case, we need to learn about $\beta_{0,-i}$, $\beta_{1,-i}$ and $\tau_{\epsilon,-i}$ just from the true values $\tilde{\boldsymbol{Y}}_{-i}$ and $\tilde{\boldsymbol{X}}_{-i}$ (which we also learn about simultaneously, with help from their respective observations), and not from $\tilde{Y}_i$ and $\tilde{X}_i$. But, while finding the log-likelihood of the true value $\tilde{Y}_i$ given $\beta_{0,-i}$, $\beta_{1,-i}$, $\tau_{\epsilon,-i}$ and $\tilde{X}_i$, we need to simultaneously estimate the true values $\tilde{Y}_i$ and $\tilde{X}_i$.

More specifically, for a fixed $i = i'$, the posterior distribution $p(\boldsymbol{\theta}_{-i'}|Y_{-i'}, X_{-i'})$ must firstly be estimated, where $\boldsymbol{\theta}_{-i'}$ is defined to be the parameter vector

$$\boldsymbol{\theta}_{-i'} = (\tilde{\boldsymbol{Y}}_{-i'}, \tilde{\boldsymbol{X}}_{-i'}, \beta_{0,-i'}, \beta_{1,-i'}, \tau_{\epsilon,-i'}, \tau_{\tilde{X},-i'}, \tau_{\eta,-i'}, \tau_{\delta,-i'}, \tau_{\beta_{0,-i}}, \tau_{\beta_{1,-i}}), \tag{A.1.0.14}$$

and the matrices $Y_{-i'}$ and $X_{-i'}$ are the matrices of the observed data without

the replicate measurements of the $(i')^{\text{th}}$ group. That is,

$$X_{-i'} = \begin{pmatrix} X_{1,1} & \cdots & X_{1,n_c} \\ \vdots & \ddots & \vdots \\ X_{i'-1,1} & \cdots & X_{i'-1,n_c} \\ X_{i'+1,1} & \cdots & X_{i'+1,n_c} \\ \vdots & \ddots & \vdots \\ X_{n_g,1} & \cdots & X_{n_g,n_c} \end{pmatrix},$$

and equivalently for $Y_{-i'}$. Secondly, using the posterior distribution $p(\boldsymbol{\theta}_{-i'}|Y_{-i'}, X_{-i'})$, the relationships

$$\tilde{Y}_{i'} = \beta_{0,-i'} + \beta_{1,-i'}\tilde{X}_{i'} + \epsilon_{i'}, \tag{A.1.0.15}$$

$$Y_{i',j} = \tilde{Y}_{i'} + \eta_{i',j}, \tag{A.1.0.16}$$

and

$$X_{i',k} = \tilde{X}_{i'} + \delta_{i',k}, \tag{A.1.0.17}$$

are utilised to find the posterior distribution $p(\tilde{Y}_{i'}, \tilde{X}_{i'}|\boldsymbol{\theta}_{-i'}, Y_{i'}, X_{i'})$ (for $j = 1, \ldots, n_r$ and $k = 1, \ldots, n_c$). The random error terms in Equations A.1.0.15, A.1.0.16 and A.1.0.17 take normal distributions with mean 0 and precisions $\tau_{\epsilon,-i'}$, $\tau_{\eta,-i'}$ and $\tau_{\delta,-i'}$ respectively. Moreover, the prior distribution $\tilde{X}_{i'} \sim \mathrm{N}(\mu_X, \tau_{\tilde{X},-i'})$ is utilised, with μ_X fixed. In order to capture the uncertainty of the posterior distribution $p(\boldsymbol{\theta}_{-i'}|Y_{-i'}, X_{-i'})$ when estimating $\tilde{Y}_{i'}$ and $\tilde{X}_{i'}$, the model represented by Equations A.1.0.15, A.1.0.16 and A.1.0.17 is fitted for each posterior sample $s = 1, \ldots, S$, giving S posterior distributions of $p(\tilde{Y}_{i'}, \tilde{X}_{i'}|\boldsymbol{\theta}_{-i}, Y_i, X_i)$, from which a single posterior sample is taken, and these samples are combined together to acquire the complete posterior distribution of $p(\tilde{Y}_{i'}, \tilde{X}_{i'}|\boldsymbol{\theta}_{-i}, Y_{i'}, X_{i'})$, from which the marginal posterior distribution $p(\tilde{Y}_{i'}|\boldsymbol{\theta}_{-i}, \tilde{X}_{i'}, Y_{i'}, X_{i'})$ is obtained. This posterior distribution is comprised of S posterior samples, which are used for finding Equation 3.4.2.1, given by

$$\begin{aligned} \mathrm{lppd}_{\text{LOO-CV-EIV}} &= \sum_{i=1}^{n_g} \log(p_{(\text{post},-i)}(\tilde{Y}_i)) \\ &= \sum_{i=1}^{n_g} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(\tilde{Y}_{i,s}|\boldsymbol{\theta}_{-i,s}, \tilde{X}_{i,s})\right). \end{aligned}$$

It is noted that exact LOO-CV for models which factor in measurement error for both the response variable and explanatory variable are even

more time-consuming—for S posterior samples and n_g groups, Sn_g models must be fitted using MCMC. An approximate LOO-CV method is therefore preferable, and is provided in Appendix A.2.

A.2 Approximate leave-one-out cross-validation with Pareto smoothing

This section describes the process for approximating LOO-CV-IC, with discussions of importance sampling which lead to the approximate method of PSIS-LOO-CV-IC.

Raw importance sampling was first described in Gelfand et al. (1992)—the expected log predictive density adjusted for leave-one-out cross-validation is defined as

$$\text{elpd}_{\text{loo}} = \sum_{i=1}^{n} \log(p(Y_i|\boldsymbol{Y}_{-i}, x_i, \boldsymbol{x}_{-i})) \tag{A.2.0.1}$$

where

$$p(Y_i|\boldsymbol{Y}_{-i}, x_i, \boldsymbol{x}_{-i}) = \int p(Y_i|\boldsymbol{\theta}_{-i}, x_i) p(\boldsymbol{\theta}_{-i}|\boldsymbol{Y}_{-i}, \boldsymbol{x}_{-i}) d\boldsymbol{\theta}_{-i}. \tag{A.2.0.2}$$

The notation of the conditioning on the explanatory variable values of $(\boldsymbol{x}_{-i}, x_i)$ in Equation A.2.0.1, as opposed to just conditioning on $\boldsymbol{x}$, is for ease of understanding, with correspondence to the posterior distribution $p(\boldsymbol{\theta}_{-i}|\boldsymbol{Y}_{-i}, \boldsymbol{x}_{-i})$ in Equation A.2.0.2.

Given conditional independence of n data points in the model, then Equation A.2.0.2 can be evaluated with samples $\boldsymbol{\theta}_s$ from the posterior conditional distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{x})$, that is, the posterior conditional distribution *given all the data points* $\boldsymbol{Y}$, using importance ratios

$$r_{i,s} = \frac{1}{p(Y_i|\boldsymbol{\theta}_s)} \propto \frac{p(\boldsymbol{\theta}_s|\boldsymbol{Y}_{-i})}{p(\boldsymbol{\theta}_s|\boldsymbol{Y})}. \tag{A.2.0.3}$$

This leads to the importance sampling leave-one-out predictive distribution for the 'test' data point Y_i

$$p(Y_i|\boldsymbol{Y}_{-i}) \approx \frac{\sum_{s=1}^{S} r_{i,s} p(Y_i|\boldsymbol{\theta}_s)}{\sum_{s=1}^{S} r_{i,s}} = \frac{1}{\frac{1}{S}\sum_{s=1}^{S} \frac{1}{p(Y_i|\boldsymbol{\theta}_s)}}. \tag{A.2.0.4}$$

The issue with the raw importance sampling is that the variance of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{x})$ is likely to be smaller than that of the posterior

distributions $p(\boldsymbol{\theta}|\boldsymbol{Y}_{-i}, \boldsymbol{x}_{-i})$—because of this, the distribution of importance ratios defined in Equation A.2.0.3 can have a large variance (specifically a long right tail, discussed in Peruggia (1997) and Epifani et al. (2008)).

One potential modification to this is suggested in Ionides (2008) where the weighted importance ratios, given by

$$w_s = \min(r_s, \sqrt{S}\bar{r}), \tag{A.2.0.5}$$

replace the importance ratios r_s. It is true that these weighted importance ratios provide finite variance. Vehtari et al. (2015) show in some experiments that these weighted importance ratios introduce a bias.

This is taken a step further, leading to the application of Pareto smoothing to the importance ratios, first discussed in Vehtari et al. (2015). Speaking generally about the Pareto distribution, it is named after Italian economist Pareto, who observed that the distribution of wealth in society is such that 80% of society's wealth is held by 20% of the society's population, leading to the 80-20 rule (Pareto principle). The Pareto distribution has probability density function of

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & \text{if } x \geq x_m, \\ 0 & \text{otherwise.} \end{cases} \tag{A.2.0.6}$$

A plot of the probability density function produces a curve in the positive quadrant which has domain $[x_m, \infty)$ for $x_m > 0$ and range $(0, \alpha]$ for $\alpha > 0$. Finally, the probability density function equals $\alpha > 0$ at x_m.

The method of Pareto-smoothed importance sampling involves fitting a generalised Pareto distribution to the largest 20% of the importance ratios, so as to smooth these values and reduce sensitivity to them. The remaining 80% of the importance ratios are left unchanged. The Pareto smoothing is carried out for each data point $i = 1, \ldots, n$, leading to new weights w_i, s for $s = 1, \ldots, S$ and each data point i. Further details of the procedure for fitting the generalised Pareto distribution and carrying out the smoothing are given in Vehtari et al. (2015) and Vehtari et al. (2017). The method does not guarantee stable estimates for the importance ratios—when the estimated shape parameter $\hat{k}$ for the generalised Pareto distribution is greater than 0.7, it is likely that the ratios are unstable.

This leads to the PSIS estimate of the LOO expected log predictive density of

$$\widehat{\text{elpd}}_{\text{PSIS-LOO-CV}} = \sum_{i=1}^{n} \log \left(\frac{\sum_{s=1}^{S} w_{i,s} p(Y_i|\boldsymbol{\theta}_s)}{\sum_{s=1}^{S} w_{i,s}} \right), \tag{A.2.0.7}$$

which, after multiplying by -2 to convert onto the deviance scale, is used to approximate LOO-CV-IC. The effective number of parameters is then given by

$$p_{\text{PSIS-LOO-CV}} = \text{lppd} - \widehat{\text{elpd}}_{\text{PSIS-LOO-CV}}. \tag{A.2.0.8}$$

An adjustment is provided for the consideration of measurement error on both the response and explanatory variables, as in Appendix A.1. The PSIS estimate of the LOO expected log predictive density, accounting for measurement error in the response and explanatory variables, is given by

$$\widehat{\text{elpd}}_{\text{PSIS-LOO-CV,EIV}} = \sum_{i=1}^{n_g} \log \left(\frac{\sum_{s=1}^{S} w_{i,s} p(\tilde{Y}_{i,s} | \boldsymbol{\phi}_{i,s})}{\sum_{s=1}^{S} w_{i,s}} \right),$$

which is used to approximate the exact LOO-CV-IC accounting for measurement error. This statistic is multiplied by -2 so that the values are placed on the deviance scale.

A.3 Exact leave-one-out cross-validation for EIV GP

Firstly, the exact LOO-CV-IC for the EIV GP with one input is presented. That is, the quantities of interest are the log-likelihoods of the 'test' true value $\tilde{Y}_i$ given the 'training' true values $\tilde{\boldsymbol{Y}}_{-i}$ and the model parameters for the Gaussian process, which are 'trained' by the true values $\tilde{\boldsymbol{Y}}_{-i}$. The parameter vector $\boldsymbol{\theta}$ for the EIV GP fitted to all true values is defined as

$$\boldsymbol{\theta} = (\alpha, \sigma_k, \sigma_\epsilon, l, \tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}}, \tau_{\tilde{X}}, \tau_\eta, \tau_\delta).$$

Since the EIV GP model is fitted n_g times in the case of carrying out LOO-CV-IC, this leads to n_g posterior distributions for each iteration of LOO, so the LOO parameter vector $\boldsymbol{\theta}_{-i}$ is defined for each 'test' true value i, that is,

$$\boldsymbol{\theta}_{-i} = (\alpha_{-i}, \sigma_{k,-i}, \sigma_{\epsilon,-i}, l_{-i}, \tilde{\boldsymbol{Y}}_{-i}, \tilde{\boldsymbol{X}}_{-i}, \tau_{\tilde{X},-i}, \tau_{\eta,-i}, \tau_{\delta,-i})$$

is the LOO parameter vector. The hyperparameter vectors, i.e. excluding the true values, is then defined to be $\boldsymbol{\phi}_{-i} = (\alpha_{-i}, \sigma_{k,-i}, \sigma_{\epsilon,-i}, l_{-i})$. The log-likelihood of the 'test' true value $\tilde{Y}_i$ given the 'training' true values is then $\log(p(\tilde{Y}_i | \boldsymbol{\phi}_{-i}, \tilde{\boldsymbol{Y}}_{-i}, \tilde{X}_i))$. This log-likelihood is constructed with the Gaussian process posterior, where the GP prior is given by

$$\begin{pmatrix} \tilde{\boldsymbol{Y}}_{-i} \\ \tilde{Y}_i \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} \boldsymbol{m}(\tilde{\boldsymbol{X}}_{-i}) \\ m(\tilde{X}_i) \end{pmatrix}, \begin{pmatrix} V(\tilde{\boldsymbol{X}}_{-i}, \tilde{\boldsymbol{X}}_{-i}) & \boldsymbol{k}(\tilde{\boldsymbol{X}}_{-i}, \tilde{X}_i) \\ \boldsymbol{k}(\tilde{X}_i, \tilde{\boldsymbol{X}}_{-i}) & v(\tilde{X}_i, \tilde{X}_i) \end{pmatrix} \right), \tag{A.3.0.1}$$

where $\boldsymbol{m}(\tilde{\boldsymbol{X}}_{-i}) \in \mathbb{R}^{n_g-1}$, $V(\tilde{\boldsymbol{X}}_{-i}, \tilde{\boldsymbol{X}}_{-i}) = K(\tilde{\boldsymbol{X}}_{-i}, \tilde{\boldsymbol{X}}_{-i}) + \sigma_\epsilon^2 I_{n_g-1}$, with $K(\tilde{\boldsymbol{X}}_{-i}, \tilde{\boldsymbol{X}}_{-i})$ being the matrix

$$\begin{pmatrix} k_{\text{cov}}(\tilde{X}_1, \tilde{X}_1) & \cdots & k_{\text{cov}}(\tilde{X}_1, \tilde{X}_{i-1}) & k_{\text{cov}}(\tilde{X}_1, \tilde{X}_{i+1}) & \cdots & k_{\text{cov}}(\tilde{X}_1, \tilde{X}_{n_g}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k_{\text{cov}}(\tilde{X}_{i-1}, \tilde{X}_1) & \cdots & k_{\text{cov}}(\tilde{X}_{i-1}, \tilde{X}_{i-1}) & k_{\text{cov}}(\tilde{X}_{i-1}, \tilde{X}_{i+1}) & \cdots & k_{\text{cov}}(\tilde{X}_{i-1}, \tilde{X}_{n_g}) \\ k_{\text{cov}}(\tilde{X}_{i+1}, \tilde{X}_1) & \cdots & k_{\text{cov}}(\tilde{X}_{i+1}, \tilde{X}_{i-1}) & k_{\text{cov}}(\tilde{X}_{i+1}, \tilde{X}_{i+1}) & \cdots & k_{\text{cov}}(\tilde{X}_{i+1}, \tilde{X}_{n_g}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k_{\text{cov}}(\tilde{X}_{n_g}, \tilde{X}_1) & \cdots & k_{\text{cov}}(\tilde{X}_{n_g}, \tilde{X}_{i-1}) & k_{\text{cov}}(\tilde{X}_{n_g}, \tilde{X}_{i+1}) & \cdots & k_{\text{cov}}(\tilde{X}_{n_g}, \tilde{X}_{n_g}) \end{pmatrix}, \tag{A.3.0.2}$$

which has dimension $(n_g - 1) \times (n_g - 1)$. Furthermore, the vector

$$\boldsymbol{k}(\tilde{\boldsymbol{X}}_{-i}, \tilde{X}_i) = \begin{pmatrix} k_{\text{cov}}(\tilde{X}_1, \tilde{X}_i) \\ \vdots \\ k_{\text{cov}}(\tilde{X}_{i-1}, \tilde{X}_i) \\ k_{\text{cov}}(\tilde{X}_{i+1}, \tilde{X}_i) \\ \vdots \\ k_{\text{cov}}(\tilde{X}_{n_g}, \tilde{X}_i) \end{pmatrix}, \tag{A.3.0.3}$$

and $\boldsymbol{k}(\tilde{X}_i, \tilde{\boldsymbol{X}}_{-i}) = (\boldsymbol{k}(\tilde{\boldsymbol{X}}_{-i}, \tilde{X}_i))^T$ are column and row vectors respectively, both of length $n_g - 1$. Then, the posterior distribution of $\tilde{Y}_i | \boldsymbol{\phi}_{-i}, \tilde{\boldsymbol{Y}}_{-i}, \tilde{X}_i$ takes mean m_i^* and variance v_i^*, given by

$$m_i^* = m(\tilde{X}_i) + \boldsymbol{k}(\tilde{X}_i, \tilde{\boldsymbol{X}}_{-i}) V(\tilde{\boldsymbol{X}}_{-i}, \tilde{\boldsymbol{X}}_{-i})^{-1} (\tilde{\boldsymbol{Y}}_{-i} - \boldsymbol{m}(\tilde{\boldsymbol{X}}_{-i})) \tag{A.3.0.4}$$

and

$$v_i^* = v(\tilde{X}_i, \tilde{X}_i) - \boldsymbol{k}(\tilde{X}_i, \tilde{\boldsymbol{X}}_{-i}) V(\tilde{\boldsymbol{X}}_{-i}, \tilde{\boldsymbol{X}}_{-i})^{-1} \boldsymbol{k}(\tilde{\boldsymbol{X}}_{-i}, \tilde{X}_i). \tag{A.3.0.5}$$

Given that some large S posterior samples summarise the posterior distribution of each parameter, this leads to S likelihoods for each 'test' true value given the 'training' true values, leading to the $\text{lppd}_{\text{LOO-CV}}$

$$\text{lppd}_{\text{LOO-CV-IC}} = \sum_{i=1}^{n_g} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(\tilde{Y}_{i,s} | \boldsymbol{\phi}_{-i,s}, \tilde{\boldsymbol{Y}}_{-i,s}, \tilde{X}_{i,s}) \right), \tag{A.3.0.6}$$

where the subscript s represents the s^{th} posterior sample of the parameters.

As discussed for the exact LOO-CV-IC for the EIV Bayesian regression discussed in Section 3.4.2, since the true values $\tilde{Y}_i$ and $\tilde{X}_i$ must also be estimated, a second MCMC must be carried out in order to find the true values

in the likelihoods in Equation A.3.0.6. For a fixed $i = i'$, the relationships

$$\tilde{Y}_{i'} \sim \mathrm{N}(m^*_{i'}, v^*_{i'}), \tag{A.3.0.7}$$

$$Y_{i',j} = \tilde{Y}_{i'} + \eta_{i',j} \tag{A.3.0.8}$$

and

$$X_{i',k} = \tilde{X}_{i'} + \delta_{i',k}, \tag{A.3.0.9}$$

where $\eta_{i',j} \sim \mathrm{N}(0, \tau_{\eta,-i})$ and $\delta_{i',k} \sim \mathrm{N}(0, \tau_{\delta,-i})$, are utilised to find the posterior distribution $p(\tilde{Y}_{i'}, \tilde{X}_{i'}|\boldsymbol{\theta}_{-i'}, Y_{i'}, X_{i'})$, where $Y_{i'}$ and $X_{i'}$ are the observed data for the i^{th} group for the response variable and covariate respectively. The true value $\tilde{X}_{i'}$ takes the prior distribution $\mathrm{N}(\mu_X, \tau_{\tilde{X},-i'})$. In order to take into account the uncertainty of the posterior distribution for the training true values, these relationships are investigated for each posterior sample s, giving S posterior distributions for $\tilde{Y}_{i'}, \tilde{X}_{i'}|\boldsymbol{\theta}_{-i'}, Y_{i'}, X_{i'}$, from which a single posterior sample is taken from each posterior, and these are combined together to give a complete posterior distribution. Finally, the $\text{lppd}_{\text{LOO-CV-IC}}$ in Equation A.3.0.6 is calculated using the samples from this posterior.

A.4 Mixed leave-one-out cross-validation

An alternative to the approximate LOO-CV, explained in Appendix A.2 for the EIV BR, is the mixed LOO-CV, which acts as an intermediate step between carrying out full LOO-CV and PSIS-LOO-CV. Note that the full LOO log-likelihood for posterior sample s of the true value $\tilde{Y}_i$ is given by

$$p(\tilde{Y}_{i,s}|\phi_{-i,s}, \tilde{\boldsymbol{Y}}_{-i,s}, \tilde{X}_{i,s}),$$

where the hyperparameters are trained on the data without the true value (and its replicate measurements) for group i. Note also that the log-likelihood, for the approximate LOO-CV, for posterior sample s of the 'new' true value $\tilde{Y}_{\text{new},i}$ is given by

$$p(\tilde{Y}_{\text{new},i,s}|\boldsymbol{\theta}_s),$$

where the hyperparameters (and the true values for the model) are all estimated based on all the available data.

The mixed LOO is carried out by estimating the hyperparameters (and true values for the model) using all the available data, then finding the LOO log-likelihood for posterior sample s of the true value $\tilde{Y}_i$ based on leaving the data for the i^{th} group out of the log-likelihood, i.e., finding

$$p(\tilde{Y}_{i,s}|\boldsymbol{\phi}_s, \tilde{\boldsymbol{Y}}_{-i,s}, \tilde{\boldsymbol{X}}_s), \tag{A.4.0.1}$$

where the vector $\boldsymbol{\phi}_s$ is posterior sample s for all the hyperparameters, estimated using all the available data. This is essentially analogous to the full LOO-CV, but the hyperparameters are estimated just once, using all of the data. Thus the out-of-sample predictive performance of the EIV GP using mixed LOO-CV-IC is given by

$$\text{lppd}_{\text{mixed-LOO-CV-EIV}} = \sum_{i=1}^{n_g} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(\tilde{Y}_{i,s} | \boldsymbol{\phi}_s, \tilde{\boldsymbol{Y}}_{-i,s}, \tilde{\boldsymbol{X}}_s) \right),$$

Appendix B

EIV GP prior exploration

B.1 Single-input EIV GP

Given the clear relationships noted between the parameters in Section 4.3.1, a reparameterisation of the model, with l and σ_k depending on σ_ϵ being justified. That is, defining

$$l = k_l \sigma_\epsilon \text{ and } \sigma_k = k_{\sigma_k} \sigma_\epsilon. \tag{B.1.0.1}$$

In this case, a uniform prior distribution is placed on the scalars k_l and k_{σ_k}. A suitable starting point can be provided from the first fitted model from Section 4.3.1, where the ratio between each posterior sample of l and σ_ϵ (as well as between σ_k and σ_ϵ) provides an implied posterior distribution for k_l (as well as for k_{σ_k}, with the obvious restriction of considering posterior samples such that $\sigma_\epsilon < 0.075$. A centred 95% credible interval is computed for the implied posterior distributions for k_l and k_{σ_k}, and rough estimates of these are taken to be the limits of their uniform priors. This provides the prior distributions

$$k_l \sim \mathrm{U}(0, 3000)$$

and

$$k_{\sigma_k} \sim \mathrm{U}(5, 1000)$$

where the 95% credible intervals from the implied prior distributions were [0.446,49.5] for k_l and [5.66,838] for k_{σ_k} (to 3 significant figures). There has been some adjustment here to the limits of the uniform prior for k_l to be able to cover future simulation examples cases (or real data cases) where the value of l is much larger—it is still possible that l could be close to 1, which is 50 times larger than the estimate of l we expect to find for the current simulation, and so the upper limit of 3000 is the starting point.

This reparameterised version of the EIV GP is now fitted to the same set of simulated data, with the prior specification

$$\begin{aligned}
\sigma_\epsilon &\sim \text{Gamma}(a_{\sigma_\epsilon}, b_{\sigma_\epsilon}),\\
\sigma_k &= k_{\sigma_k}\sigma_\epsilon,\\
l &= k_l\sigma_\epsilon,\\
k_{\sigma_k} &\sim \text{U}(5, 1000),\\
k_l &\sim \text{U}(0, 3000),\\
\alpha &\sim \text{N}(0, 0.01),\\
\tilde{X}_i &\sim \text{N}(\mu_X, \tau_{\tilde{X}}),\\
\tau_\eta &\sim \text{Gamma}(a_\eta, b_\eta),\\
\tau_\delta &\sim \text{Gamma}(a_\delta, b_\delta),\\
\tau_{\tilde{X}} &\sim \text{Gamma}(a_{\tilde{X}}, b_{\tilde{X}}).
\end{aligned} \tag{B.1.0.2}$$

Again, the shape and rate parameters in the Gamma prior distributions are all set equal to 0.001, apart from for $\sigma_\epsilon \sim \text{Gamma}(1, \frac{1}{0.1})$, and $\mu_X = 0.5$. The MCMC output is assessed for mixing and convergence. The PSRF metric shows that the largest (of all parameters) upper bound on the 95% confidence interval of each estimate is 1.00, so the output has converged to the posterior distribution. The smallest estimate of effective sample size of any parameter from any of the four chains is 17993, which indicates excellent levels of mixing. The MCMC output is now considered to have converged to the posterior distribution.

The posterior densities of σ_ϵ, l and σ_k is provided in Figure B.1. There are clearly issues with these posterior densities, relative to what is expected from these. Most importantly, the values of σ_ϵ (top plot) are large, compared with the cutoff of 0.075 that was identified in the previous EIV GP fit for this simulation. Moreover, the values of l and σ_k are also large, which is no surprise given the priors on their respective scalars and the values found in the marginal posterior of σ_ϵ. The posterior densities of k_l and k_{σ_k} are also provided in Figure B.2. These posterior densities of k_l and k_{σ_k} are expected given the posterior densities of σ_ϵ, l and σ_k observed in Figure B.1. Given the desire to capture larger values of l even with smaller values of σ_ϵ, it still seems necessary to have this wide-ranging prior for the scalar for k_l.

It is not too surprising that this reparameterisation, as fitted here, was unsuccessful in finding an appropriate joint posterior distribution for the hyperparameters. While the uniform priors for the scalars k_l and k_{σ_k} was adjusted for the case that the posterior samples of σ_ϵ were less than 0.075, there has been no adjustment in the prior distribution for σ_ϵ, which can

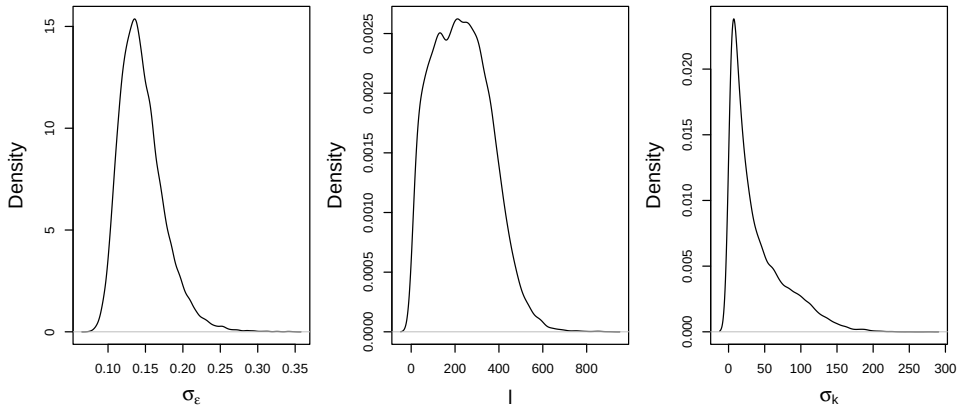

Figure B.1: Posterior density of σ_ϵ (top plot), l (middle plot) and σ_k (bottom plot), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(1, \frac{1}{0.1})$, $l = k_l \sigma_\epsilon$, $k_l \sim \text{U}(0, 3000)$, $\sigma_k = k_{\sigma_k} \sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(5, 1000)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

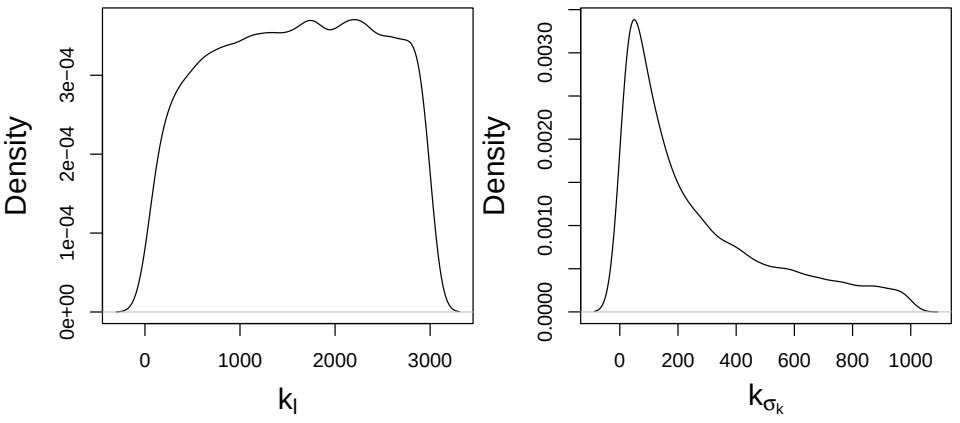

Figure B.2: Posterior densities of k_l (top plot) and k_{σ_k}, corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(1, \frac{1}{0.1})$, $l = k_l \sigma_\epsilon$, $k_l \sim \text{U}(0, 3000)$, $\sigma_k = k_{\sigma_k} \sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(5, 1000)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

clearly still find quite large values of model error standard deviation. An adjustment to the prior distribution of σ_ϵ is discussed further here, firstly with consideration of the 'initial' parameterisation (i.e., not assuming a direct relationship between σ_ϵ and both l and σ_k), and secondly with the reparameterisation of the model.

As the prior distribution on σ_ϵ was first discussed in Section 4.3.1, it was mentioned that the mean could be shifted closer to 0 if it appeared necessary from simulation examples. In truth, given that a model error standard deviation of 0.25 is the maximum that can be considered (given a constant mean for the response variable at 0.5, the model error can then cover the range [0,1], and so setting $1 - 0 = 4\sigma_\epsilon$ gives the rough estimate of 0.25), there should be zero density in the prior distribution to the right of this value. With the prior Gamma$(1, \frac{1}{0.1})$, it was noted that the probability of a sample from the prior being greater than 0.25 is 0.0821, i.e., this occurs over 8% of the time, which is likely allowing for too much density to σ_ϵ at values larger than what can be feasibly measured. An adjustment to the mean of the prior closer to 0 will decrease this probability, or an increase in the shape of the prior. Both of these options are considered in Figure B.3, which displays the posterior densities of four Gamma distributions, along with the probability that a sample from the distribution is larger than 0.25. The four Gamma distributions considered are Gamma$(1, \frac{1}{0.1})$, Gamma$(1, \frac{1}{0.05})$, Gamma$(2, \frac{2}{0.1})$ and Gamma$(2, \frac{2}{0.05})$. Note from the probability statements

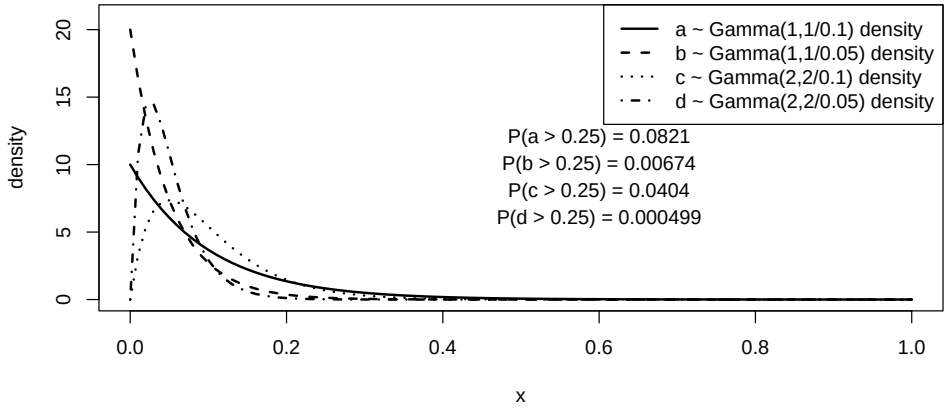

Figure B.3: The probability density function for four Gamma distributions that are considered as possible prior distributions for σ_ϵ. Relevant probability statements are also provided.

the change in the probability as the shape is increased from 1 to 2, and as the

mean is decreased from 0.1 to 0.05. Given we want almost zero probability of the model error standard deviation being greater than 0.25, it seems like a mean for the distribution of 0.1 should be ruled out. This probability is clearly much smaller for the two Gamma distributions with mean 0.05; note that the probability that a random sample is greater than 0.25 is still over 10 times larger for Gamma(1, $\frac{1}{0.05}$) than for Gamma(2, $\frac{2}{0.05}$) (note the difference in probability density at values around 0.15 to 0.2). The other notable difference is the density at values very closer to 0, which is much larger for the Gamma prior with shape 1 (which has its mode at 0). It possibly makes more sense for this reason to use the shape 2 prior, since a shape 1 prior suggests a priori that the most likely value for the model error standard deviation is 0.

The EIV GP model is fitted again to this set of simulated data, firstly with the assumptions that $\sigma_\epsilon \sim$ Gamma(2, $\frac{2}{0.05}$), $\sigma_k \sim$ Gamma(3, $\frac{3}{0.4}$) and $l \sim$ Gamma(2, $\frac{2}{0.25}$), to see whether the reparameterisation considered above is necessary, or the issue the whole time was with the prior on σ_ϵ. The MCMC output is checked for convergence, and shows that the PSRF estimates for l have a mean of 1.29 and upper bound (95% C.I.) of 1.82. There appears to still be significant issues with this version of the EIV GP. The mixing within each parallel chain is comparatively interesting—the minimum effective sample size estimates of the second and third chains are 14032 and 13988 respectively, whereas for chains 1 and 4 these are 90.9 and 42.2 respectively. This suggests that chains 2 and 3 have at least mixed very well, and the posterior distribution may look different to what has been found in chains 1 and 4.

The joint posterior samples of l and σ_ϵ from each of the four parallel chains are considered in Figure B.4. Evidently, the joint posterior samples of l and σ_ϵ for chains 2 and 3 look different from those in chains 1 and 4, and it is clear to see that chains 2 and 3 have mixed much better than chains 1 and 4. It is still interesting to see that there is a lot of posterior density for σ_ϵ at values greater than 0.1 for both the first and fourth chain, although the aim of the change in the prior was to minimise the probability of sampling values greater than 0.25. It is still clear from the Gamma distributions considered for the prior of σ_ϵ in Figure B.3 that the probability density for values between 0.1 and 0.2 is reduced by changing the mean of the prior from 0.1 to 0.05. There is possible evidence here that the density between 0.1 and 0.2 in the prior of σ_ϵ should be reduced further.

The joint posterior samples of σ_k and σ_ϵ are also compared here (see Figure B.5), with a brief note that the relationship between σ_k and σ_ϵ, for values of $\sigma_\epsilon < 0.075$, still exists, and is evident in each parallel chain.

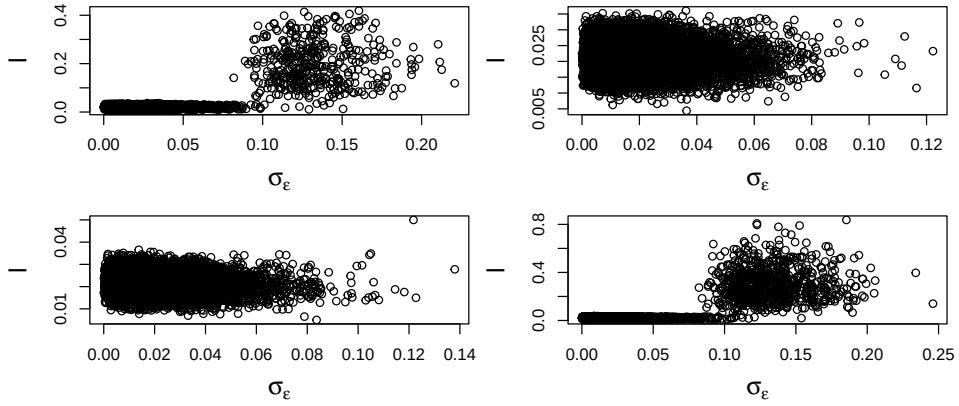

Figure B.4: Four plots of the joint posterior samples of l and σ_ϵ for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

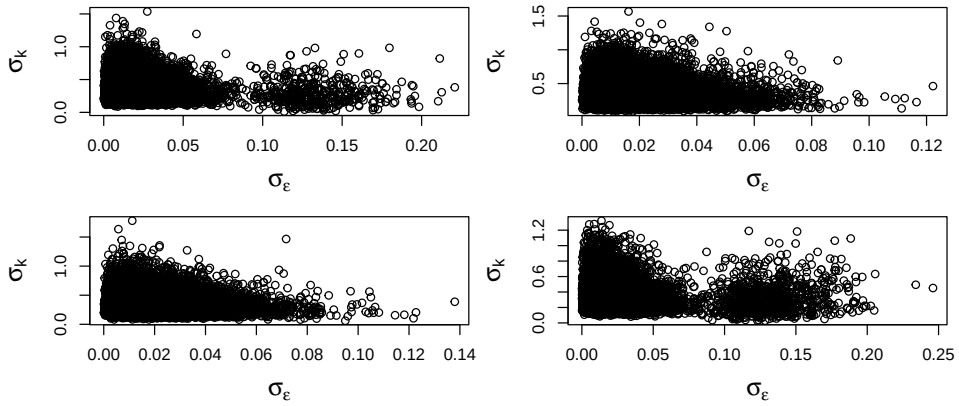

Figure B.5: Four plots of the joint posterior samples of σ_k and σ_ϵ for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

The reparameterisation is considered here again, due to the severity of the convergence issue in the previous attempt. As was carried out before, a rough estimate of the limits on the uniform priors for k_l and k_{σ_k} is gathered from the posterior samples of this model fit by looking at the ratio of the joint posterior samples of l to σ_ϵ, and also of σ_k to σ_ϵ, only considering values of σ_ϵ less than 0.075. An adjustment is then performed in order to be able to find larger values of l which could occur with real data or future simulated data examples, by multiplying the upper limit by 50 (since the mean of l is roughly 0.02, and larger values of l are considered to be around 1). A 95% credible interval over these ratios is given by [0.395,8.73] corresponding to k_l, and [4.90,156] corresponding to k_{σ_k}. The uniform priors are now given by

$$k_l \sim \mathrm{U}(0, 400)$$

and

$$k_{\sigma_k} \sim \mathrm{U}(4, 200).$$

The EIV GP model is fitted again to the same set of simulated data, now with $\sigma_\epsilon \sim \mathrm{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \mathrm{U}(0, 400)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$ and $k_{\sigma_k} \sim \mathrm{U}(4, 200)$. The MCMC output is checked for convergence and mixing. The upper bounds for the 95% confidence interval of PSRF estimates for σ_ϵ, l and σ_k are 7.28, 6.84 and 5.81 respectively, giving a clear indication that the model is not converging. The mixing of the chains also shows there are issues with autocorrelation in the MCMC output—each parallel chain has an effective sample size estimate for one of its parameters that is less than 5.

Given a clear lack of convergence, the posterior density of σ_ϵ is explored for each of the four parallel chains. These are provided in Figure B.6, with the top-left plot corresponding to the first chain, the top-right corresponding to the second chain, the bottom-left corresponding to the third chain, and the bottom-right corresponding to the fourth chain. These plots provide some understanding as to why the model is not converging, as the behaviour of σ_ϵ appears slightly different in each one, or at least the proportion of the density found at values of σ_ϵ around 0.1 to 0.15 is different in each case. In truth, the behaviour is similar, in that the majority of the posterior density is found at values between 0 and (roughly) 0.04, and the rest of the posterior density is found in the region between 0.1 and 0.15. Interestingly, in each case there seems to be zero density between values of roughly 0.05 and 0.09. The posterior densities in chains 3 and 4 are essentially multimodal, with peaks at roughly 0.01 and 0.12. Given the exploration of the GP posterior

samples in Figure 4.28, it is clear that this level of posterior density at larger values of σ_ϵ will correspond to a poor fit for the EIV GP posterior.

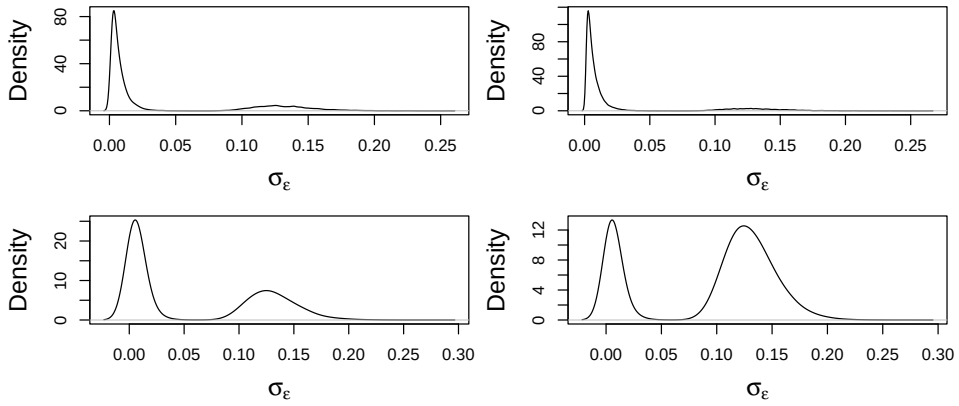

Figure B.6: Four plots of the posterior density of σ_ϵ for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 400)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

The posterior densities of l and k_l for this model fit are displayed in Figures B.7 and B.8 respectively. It is immediately clear from Figure B.7 that the values of l being recovered in the posterior distribution are far too large for this simulation example. Consequently, the posterior densities of k_l in Figure B.8 are unsurprising. At this point, an alteration to the prior density for σ_ϵ would possibly eliminate too much density at values that are still in a way plausible given the scaling of the data onto [0,1], and so the other logical adjustment would be to change the upper bound on the prior distribution of k_l, so that these incredibly large values of l are not possible. This is a compromise for the opposite extreme case, where large values of l close to (or above) 1 are required, as these will still be attainable, but will sacrifice smaller values of σ_ϵ in order to achieve them. In this example, the cutoff for 'suitable' σ_ϵ was found to be 0.075 (that is not to say that we should always want values below 0.075 for σ_ϵ for any example—if in another simulation (or real data) example values of l are required, and the maximum value to be desired for σ_ϵ were 0.075, then the smallest upper limit for the uniform prior on k_l would be $\frac{1}{0.075} = 13.3333$. In this case, there is still a significant amount of leeway for the adjustment to the upper bound for

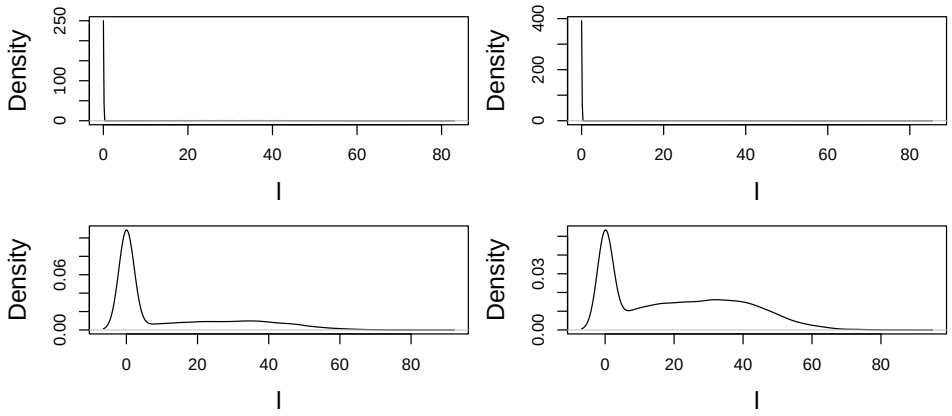

Figure B.7: Four plots of the posterior density of l for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 400)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

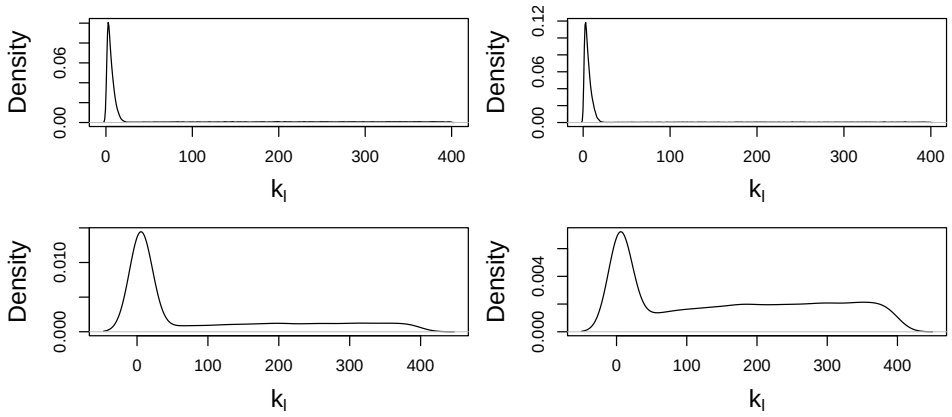

Figure B.8: Four plots of the posterior density of k_l for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 400)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

the uniform prior on k_l. The idea here is now to decrease the upper bound, and fit the model again, and continue to do this until we have a suitable joint posterior distribution for the hyperparameters which produces a suitable fit for the EIV GP posterior. At that point, the corresponding prior specification will be tested on other simulation examples, and adjustments can be made to the prior specification if necessary based on the fits for other examples.

Alterations to the upper bound of k_{σ_k} are not considered here. In this example, it is clear that smaller values of l are to be recovered in the posterior distribution so that the EIV GP is fitted well, along with smaller values of σ_ϵ. We have observed an ideal ratio of k_l within the range [0.395,8.73] from the model fitted with prior specification $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable, and this range was adjusted to account for the possibility that values of l to be recovered are around the value 1 as opposed to the value 0.02. In the case of σ_k, it was clear from looking into possible prior values that we expect to find values smaller than 1 for all examples, and it was actually observed from this model fit that values exceeding 1 are possible—a slight adjustment was made to the identified ratio for k_{σ_k} to account for possibly finding even larger values of σ_k in other cases, so there could be some justification for slightly decreasing this upper bound. The main focus here is on l and k_l, as this has more of an influence on the behaviour of the GP posterior.

The EIV GP model, with prior specification $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 300)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$ and $k_{\sigma_k} \sim \text{U}(4, 200)$ and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable, is now fitted to this set of simulated data. The MCMC output is firstly checked for convergence using PSRF estimates. The largest upper bound for the 95% confidence interval of this estimate for any parameter is 1.13 for σ_ϵ, suggesting that there has been some improvement towards convergence, but this has still not been achieved. The estimates of effective sample size show that the chains are not mixing well, and posterior samples are heavily correlated—the smallest effective sample size estimate for each chain is below 6.

The posterior density of σ_ϵ for each of the four parallel chains is displayed in Figure B.9. While convergence has improved, it appears that the chains are converging to a posterior distribution that is not desired, with 3 of the 4 chains having a lot of posterior density for σ_ϵ at values around 0.1 to 0.2. It is clear that posterior density at these values coincides with a much

larger range of l being possible—the plots of posterior density for l and k_l are omitted here and the model is fitted again with an adjustment to the upper bound on the uniform prior for k_l.

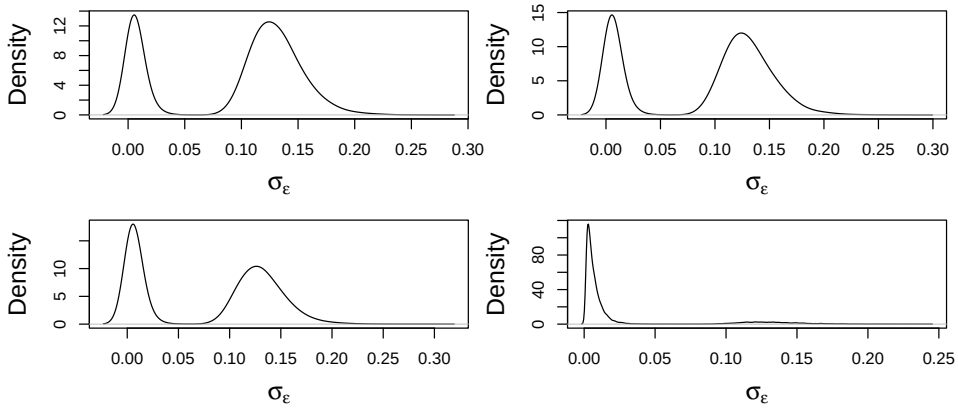

Figure B.9: Four plots of the posterior density of σ_ϵ for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 300)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

The EIV GP model, with prior specification $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 200)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$ and $k_{\sigma_k} \sim \text{U}(4, 200)$ and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable, is now fitted to this set of simulated data. The PSRF shows poor levels of convergence, with the parameters σ_ϵ, l and σ_k having values above 7 for the 95% confidence interval upper bound on this estimate. The effective sample size estimates show that the mixing of the chains is satisfactory for chains 1 and 4, but poor for chains 2 and 3. Some understanding of this can be found when considering the posterior densities of σ_ϵ from each parallel chain (see Figure B.10), where the chains are understandably mixing better if they are not getting stuck in another region (often happens in cases of multimodal posterior distributions, where the chain remains in the 'vicinity' of one mode for a long time, then jumps to the vicinity of the other mode and remains there for a long while. This occurs consistently in backward model cases, and is discussed further in Section 5). A trace plot of the posterior samples confirms this is happening in chains 2 and 3 (see Figure B.11).

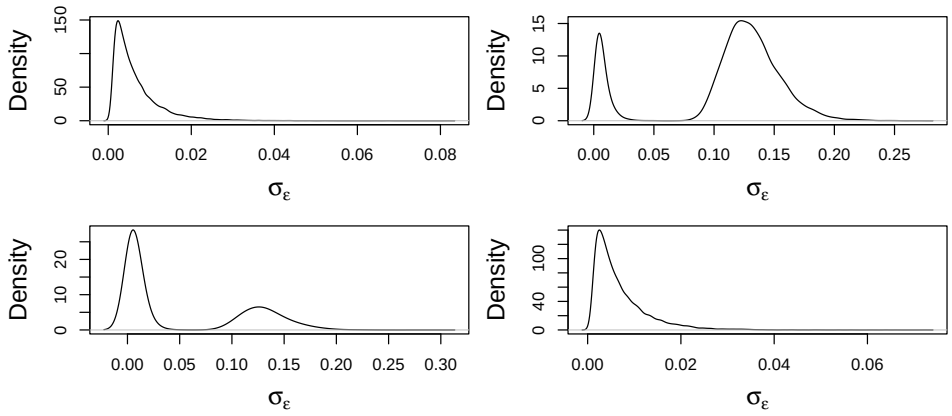

Figure B.10: Four plots of the posterior density of σ_ϵ for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 200)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

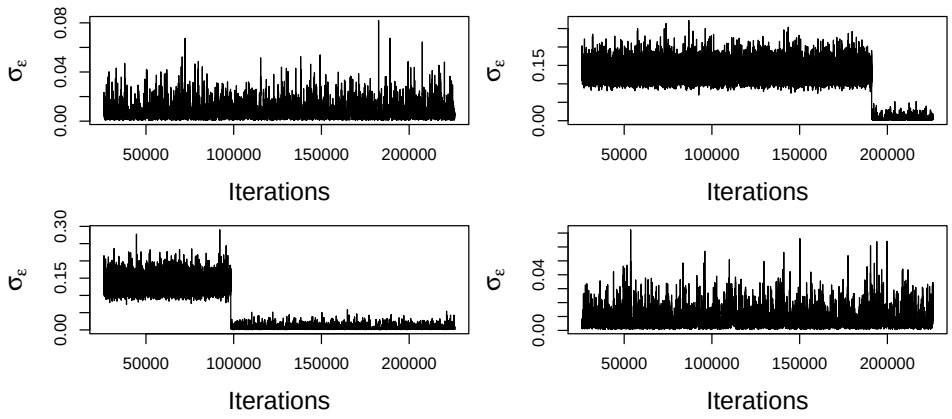

Figure B.11: Four trace plots of the posterior samples of σ_ϵ for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{005})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 200)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

Adjusting the upper bound on the uniform prior of k_l to 100 and fitting the EIV GP model provides a fit where the MCMC output has converged to the posterior distribution, with the upper bound on the 95% confidence interval for PSRF of each parameter being 1.00 to 3 significant figures. The level of mixing in the second parallel chain is poor, with the smallest estimate of effective sample size for any parameter given by 6.33, whereas the other 3 chains have achieved sufficient mixing (smallest estimate for any parameter of 322 from those 3 chains). Given convergence has been observed, the poor effective sample size from one of the four chains is not an issue, though it is still of interest to see what has occurred in that chain.

Of most interest are the posterior densities of σ_ϵ from each of the four parallel chains, plotted in Figure B.10, where it is noted that chains 1 and 3 (top-left and bottom-left plots respectively) are not finding any density at larger values of σ_ϵ (i.e., anything larger than 0.075). In comparison, chains 2 and 4 (top-right and bottom-right plots respectively) are capturing some (relatively minimal, particularly in the case of chain 4) density at larger values of σ_ϵ.

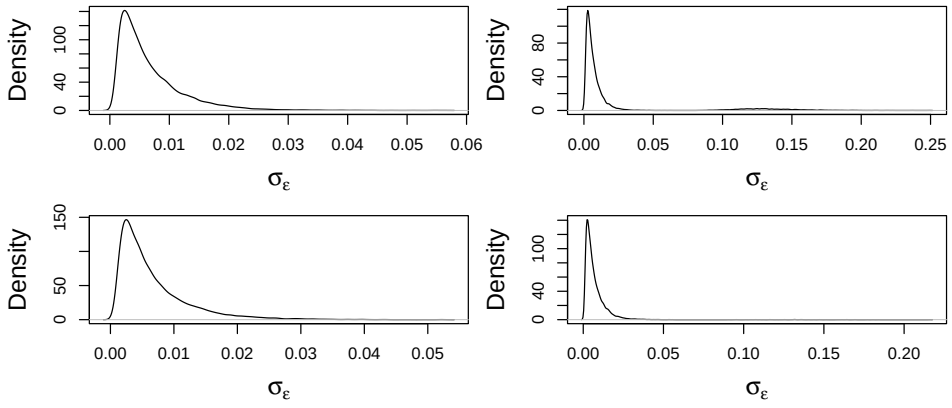

Figure B.12: Four plots of the posterior density of σ_ϵ for each of the four parallel chains (top left is chain 1, top right is chain 2, bottom left is chain 3, bottom right is chain 4), corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $k_l \sim \text{U}(0, 100)$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable.

It is hard to define at what point the posterior distribution is satisfactory so that the prior specification can be considered suitable and tested further on other simulation examples. Given the model above has converged to

the posterior distribution, the effective sample size issue with chain 2 is somewhat alleviated, as the effective sample sizes from each chain can be collected together in the same way the posterior samples can be collected together from each chain. Having some level of posterior density at these 'larger' values of σ_ϵ seems unsatisfactory. A crude estimate for the posterior probability that $P(\sigma_\epsilon > 0.075)$ can be computed using the posterior samples (this does not account for the autocorrelation between samples)—taking the posterior samples from all four chains (of which there are 80000) and counting the number of posterior samples from each chain for which $\sigma_\epsilon > 0.075$ gives a rough estimate of $P(\sigma_\epsilon > 0.075) = 0.0266$, so every 1 in roughly 38 posterior samples will have $\sigma_\epsilon > 0.075$. Furthermore, a 95% credible interval for σ_ϵ (over all posterior samples from all four chains) has an upper bound of 0.101 to 3 significant figures. This helps introduce a sensible cutoff—if the upper bound of the 95% credible interval is less than 0.075, then the posterior distribution is satisfactory. So, the upper limit for the uniform prior on k_l is decreased until the upper bound on the 95% credible interval for σ_ϵ is less than 0.075 (assuming also that the MCMC output has converged and mixed sufficiently well).

The upper limit for the uniform prior on k_l is set to 80, and the EIV GP model is fitted again. The MCMC output has converged to the posterior distribution based on PSRF (upper bound on all estimates rounds to 1.00). The effective sample size estimates show some issues with autocorrelation of posterior samples in chain 2, with the remaining 3 chains performing well. A 95% credible interval for σ_ϵ is given by [0.00120,0.117] to 3 significant figures, and so the upper bound is still above the cutoff of 0.075.

Setting the upper limit for the uniform prior on k_l to 60 provides the desired model fit. The MCMC output has converged to the posterior distribution according to PSRF (largest upper bound for the estimate is 1.00 to 3 significant figures), and the mixing of each chain is at a sufficient level, with the minimum effective sample size estimate for any parameter from any chain being 1403. The 95% credible interval for σ_ϵ over the four parallel chains is [0.00120,0.02160], with the upper bound here being less than 0.075. A plot of the GP posterior is provided in Figure B.13.

B.2 Multi-input EIV GP

The fitting of the multi-input EIV GP is firstly considered with prior specification analogous to that given in Equation 4.3.1.6, with the two main points being that the initial parameterisation, with prior distributions for

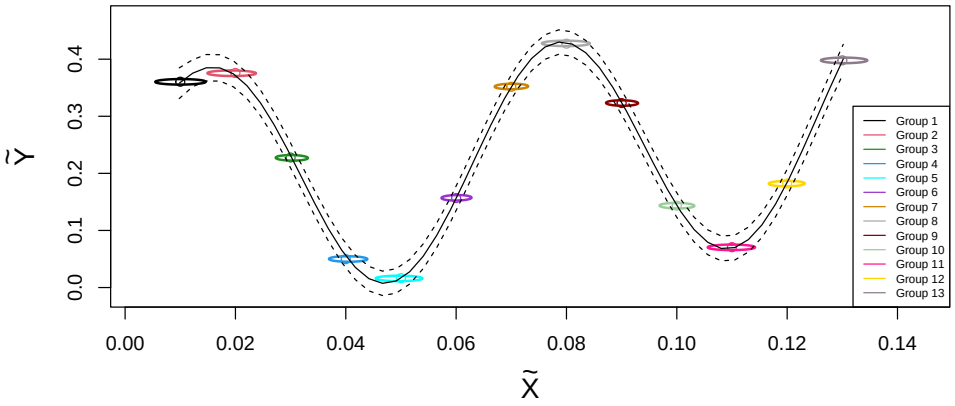

Figure B.13: A 'fitted values' plot for the EIV GP posterior given the joint posterior distribution of the hyperparameters and true values $(\tilde{Y}_i, \tilde{X}_i)'$, corresponding to the model fitted with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l = k_l\sigma_\epsilon$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_l \sim \text{U}(0, 60)$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and uninformed Gamma priors for measurement error precision and between-materials precision for the input variable. The black dotted line represents the mean GP posterior sample over the joint posterior samples of the hyperparameters, and the black dotted lines represent the mean 95% prediction interval of the GP posterior over the joint posterior samples of the hyperparameters. The uncertainty in the true values of the response variable and input variable are demonstrated by the 95% credible ellipses over the joint marginal posterior distribution of each $(\tilde{Y}_i, \tilde{X}_i)'$. The simulation is built from the function defined in Equation 4.3.1.1.

the distance-scaling parameters (in the case there are two with the SE-ARD kernel) and for the signal standard deviation σ_k, is firstly used, and secondly, that the true values are now vectorised corresponding to the multiple inputs, which induces the precision matrices $T_{\tilde{X}}$ and T_δ, as well as their Wishart prior distributions with $S_{\tilde{X}}$ and S_δ as the respective scale matrices. The prior distributions for l_1 and l_2 are both Gamma(2, $\frac{2}{0.25}$) as in Section 4.3.1, and similarly, the prior distributions for σ_k and σ_ϵ are Gamma(3, $\frac{3}{0.4}$) and Gamma(2, $\frac{2}{0.05}$) respectively. As a reminder, the MCMC tuning parameters are an adaptation length of 1000 samples, a burn-in length of 25000 samples, and 20000 posterior samples stored from 200000 posterior samples, taking every 10^{th} sample from the 200000. The model is fitted with four parallel chains.

The multi-input EIV GP is now fitted to this set of simulated data provided in Section 4.3.4. The MCMC output notes that, with an adaptation phase of length 1000, the model does not run in the optimal state with regards to its sampling behaviour. In order to avoid suboptimal sampling behaviour (with regards to efficiency), the recommendation is to run the model again with an increasing number of samples in the adaptation phase, so that the sampling is optimal. This is achieved with an adaptation phase of length 2000. The upper bounds of PSRF estimates are checked for each parallel chain and each parameter, and the maximum of these is 1.01 to 2 decimal places, indicating that the MCMC has converged. The minimum effective sample size estimate is 661 to 3 significant figures, which suggests that the levels of autocorrelation between posterior samples are sufficiently small.

Having confirmed the MCMC output is from the posterior distribution, the joint posterior distribution is now considered. The marginal posterior density of σ_ϵ is provided in Figure B.14. As in the case of Section 4.3.1, where larger values of σ_ϵ cause too much uncertainty in the predictions of the response variable (when considering the initial parameterisation with prior distributions for l and σ_k), this looks to be the case here with the multi-input model. The reparameterised model, aimed to deal with issues of the MCMC output not converging, does not seem to be needed here, given the upper bounds of PSRF, but will be introduced to see if this improves the model fit. A check of the joint posterior samples of l_1 and σ_ϵ for this model with the 'initial' parameterisation, confirms that there is dependency between these parameters, with larger values of σ_ϵ coinciding with larger variance in the values of l_1. This plot of the joint posterior samples of l_1 and σ_ϵ is provided in Figure B.15.

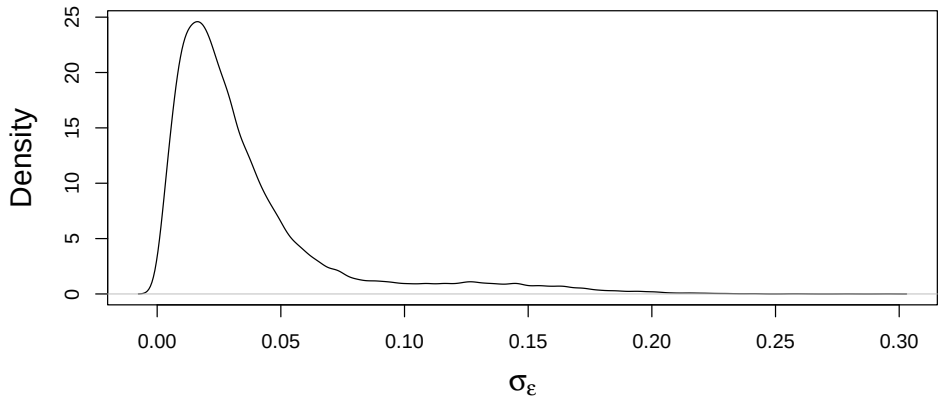

Figure B.14: The marginal posterior density of σ_ϵ, from the multi-input EIV GP with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l_1, l_2 \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$, $T_{\tilde{X}}\text{Wishart}(I_2, 2)$, $T_\delta \sim \text{Wishart}(1000 I_2, 2)$

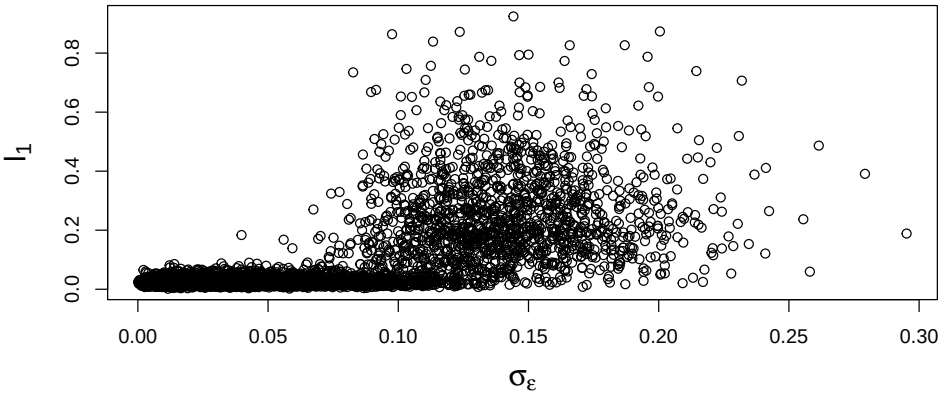

Figure B.15: A plot of the joint posterior samples of l_1 and σ_ϵ, from the multi-input EIV GP with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l_1, l_2 \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$, $T_{\tilde{X}}\text{Wishart}(I_2, 2)$, $T_\delta \sim \text{Wishart}(1000 I_2, 2)$

The multi-input EIV GP is fitted again, with the reparameterisation

$$l_d = k_{l_d}\sigma_\epsilon, \text{ for } d = 1, 2,$$

and

$$\sigma_k = k_{\sigma_k}\sigma_\epsilon,$$

where the the model error standard deviation σ_ϵ still takes the prior distribution Gamma$(2, \frac{2}{0.05})$. The prior distributions for the scalar values are $k_{l_d} \sim \text{U}(0, 60)$ for $d = 1, 2$, and $k_{\sigma_k} \sim \text{U}(4, 200)$, as were found to be effective in Sections 4.3.1 to 4.3.2. With the same MCMC tuning parameters, the adaptation phase of length 2000 is too short, indicating that the sampling behaviour of the MCMC will not be optimal. The model is run again with an increased adaptation phase of length 3000, which provides optimal sampling behaviour in this case. The convergence to the posterior distribution is checked by considering the maximum upper bound of PSRF confidence intervals, over all parameters and over the four chains. The parameter l_1, which is just equal to $k_{l_1}\sigma_\epsilon$, is shown to be not converging with a PSRF upper bound of 1.14, despite the convergence of the two parameters k_{l_1} and σ_ϵ, who have PSRF upper bounds of 1.04 and 1.02 respectively (all the 3 significant figures). The model is run again, with an increase in the burn-in period and the number of stored posterior samples to 30000 and 25000 respectively (from 25000 and 20000 respectively). In this case, the PSRF upper bound of l_1 is now 1.15. Another attempt at increasing the burn-in and stored samples to 35000 and 30000 posterior samples respectively and refitting the model then provides an upper bound of PSRF for l_1 of 1.17, and for σ_k of 1.13.

It seems curious that the convergence of these parameters would become worse as more samples were drawn, and a longer burn-in was run for the model. It is possible that continuing down the avenue of drawing more samples would improve the convergence of the model; on the other hand, other adjustments to the prior specification could be considered, similarly to the case of the EIV BR simulation with two input variables and a full quadratic linear predictor, discussed in Section 4.2.5. That is, weakly informed prior distributions are introduced, which are obtained using the fact that the data is scaled onto the range [0,1].

The weakly informed prior distributions derived in Section 4.2.5, relating to the measurement error precision and between-materials precision, are

utilised here. That is, it is assumed that

$$\tau_\eta \sim \text{Gamma}(1, 6.57 \times 10^{-4}),$$
$$T_\delta \sim \text{Wishart}(S_\delta, 2), \text{ with } S_\delta = \begin{pmatrix} 1522.07 & 0 \\ 0 & 1522.07 \end{pmatrix},$$
$$T_{\tilde{X}} \sim \text{Wishart}(S_{\tilde{X}}, 2), \text{ with } S_{\tilde{X}} = \begin{pmatrix} 15.3664 & 0 \\ 0 & 15.3664 \end{pmatrix}.$$

Additionally, the prior distribution for the GP prior mean α is adjusted from N(0,0.01) to N(0.5,10.8241), where 0.5 is the midpoint of the range [0,1], and the precision value 10.8241 is derived from the probability statement

$$P(0 \leq \alpha \leq 1) = 0.9$$

and that

$$P(-1.64\sigma_\alpha \leq \alpha \leq 1.64\sigma_\alpha) = 0.9,$$

with the latter using the assumption that α is normal at the 90% of its density lies within the range $[-1.64\sigma_\alpha, 1.64\sigma_\alpha]$ (with σ_α being the standard deviation of α). With these adjustments to the prior specification, the initial parameterisation, with $l_1, l_2 \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$ and $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$ is firstly investigated. It is noted that the adaptation phase required a length of 2000 in order to optimise the sampling behaviour of the algorithm. Once more, it is observed that the posterior density of σ_ϵ is capturing larger values of model error standard deviation, which coincide with larger values of l_1, which is indicated by the plot of joint posterior samples in Figure B.16.

The EIV GP model with the reparameterisation was then fitted, with the weakly informed priors discussed above, as well as $l_d = k_{l_d}\sigma_\epsilon$, $\sigma_k = k_{\sigma_k}\sigma_\epsilon$, $k_{l_1}, k_{l_2} \sim \text{U}(0, 60)$ and $k_{\sigma_k} \sim \text{U}(4, 200)$. It is noted that the adaptation length required to optimal sampling behaviour was 3000. The convergence of the MCMC output to the posterior distribution is confirmed, and sufficient levels of mixing are observed for each parameter. Interestingly, the issues with the model capturing larger values of σ_ϵ still exist, but the frequency with which this occurs is small. The plot in Figure B.17 demonstrates that the number of posterior samples of σ_ϵ greater than 0.075 is smaller, and comparing with Figure B.16, the issue of larger l_1 values has also diminished. Moreover, the number of posterior samples for which $\sigma_\epsilon > 0.075$ is 131, and with 80000 posterior samples stored in total (from all four parallel chains), it is estimated that $P(\sigma_\epsilon > 0.075) = 0.00164$ to 3 decimal places. While this issue is not completely eliminated, given that convergence in all parameters

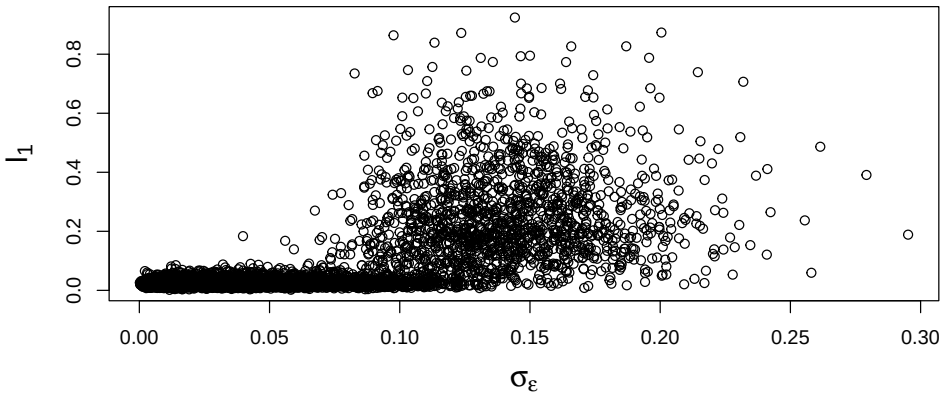

Figure B.16: A plot of the joint posterior samples of l_1 and σ_ϵ, from the multi-input EIV GP with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l_1, l_2 \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_k \sim \text{Gamma}(3, \frac{3}{0.4})$, and weakly informed prior distributions for $T_{\tilde{X}}$, T_δ, τ_η and α.

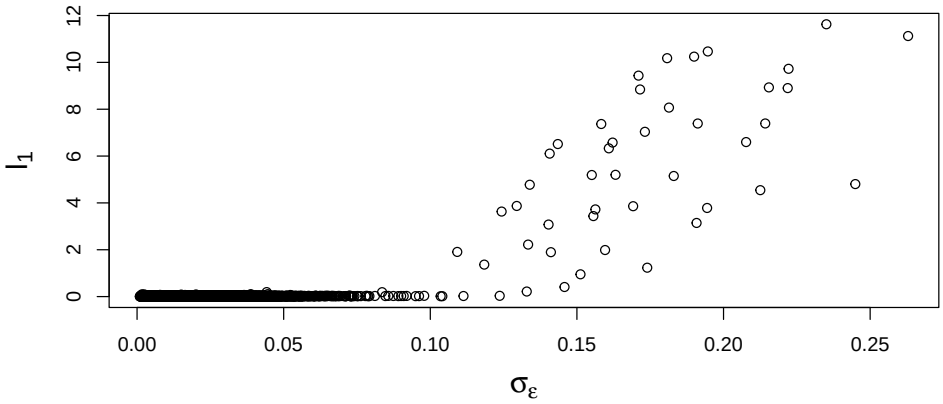

Figure B.17: A plot of the joint posterior samples of l_1 and σ_ϵ, from the multi-input EIV GP with $\sigma_\epsilon \sim \text{Gamma}(2, \frac{2}{0.05})$, $l_d = k_{l_d}\sigma_\epsilon$ (for $d = 1, 2$), $\sigma_k = k_{\sigma_k}$, $k_{l_1}, k_{l_2} \sim \text{U}(0, 60)$, $k_{\sigma_k} \sim \text{U}(4, 200)$, and weakly informed prior distributions for $T_{\tilde{X}}$, T_δ, τ_η and α. Note that these posterior samples are taken from the first parallel chain.

is observed (and so the issue is not prevalent enough to cause issues with the MCMC output), this issue is not further investigated. Some plots of the GP posterior, for various vectors of the input variables, are provided in Section 4.3.4.

B.3 EIV MOGP

The EIV MOGP model outlined in Section 4.3.5 is applied to the simulation outlined in the same section. It is noted that the required adaptation length for optimal sampling behaviour was 2000, and the typical values for the MCMC tuning parameters lead to convergence and sufficient levels of mixing. Summary statistics of the posterior distribution are examined, and suggest a possible issue with the posterior of σ_{ϵ_1}, similarly to the issue noted in the previous two sections of the appendix (that is, the issue when using the uninformed prior distributions for T_δ, $T_{\tilde{X}}$ and τ_η). That is, the model captures slightly larger values of σ_{ϵ_1} that are not appropriate for this example. A plot of the marginal posterior density of σ_{ϵ_1} is provided in Figure B.18, which looks comparatively worse than the case of Figure B.14. This in spite of the fact that the weakly informed priors for α_1, α_2, $\tau_{\eta_1}, \tau_{\eta_2}$, T_δ and $T_{\tilde{X}}$ are implemented.

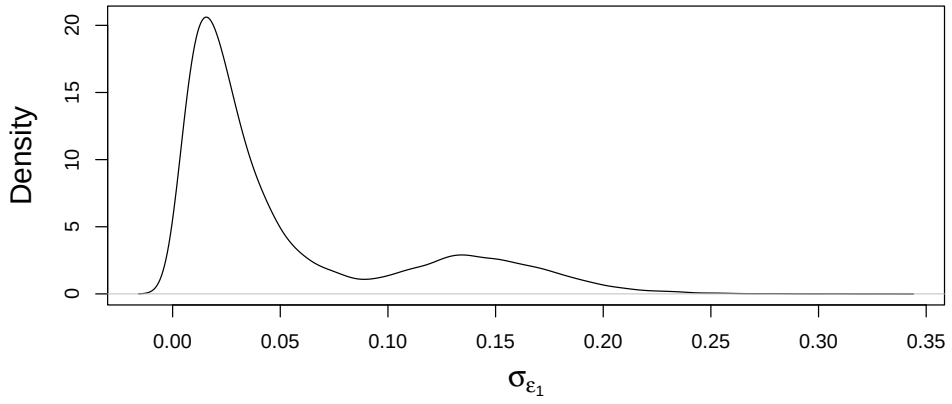

Figure B.18: The marginal posterior density of the σ_{ϵ_1}, from the EIV MOGP with $\sigma_{\epsilon_1}, \sigma_{\epsilon_2} \sim \text{Gamma}(2, \frac{2}{0.05})$, $\alpha_1, \alpha_2 \sim \text{N}(0.5, 10.8241)$, $l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2} \sim \text{Gamma}(2, \frac{2}{0.25})$, $\sigma_{k_1}, \sigma_{k_2} \sim \text{Gamma}(3, \frac{3}{0.4})$, $T_{\tilde{X}}\text{Wishart}(S_{\tilde{X}}, 2)$, $T_\delta \sim \text{Wishart}(S_\delta, 2)$, where $S_{\tilde{X}}$ and S_δ are the scale matrices for the weakly informed priors of $T_{\tilde{X}}$ and T_δ respectively.

Notably, the reparameterisation that is applied in the multi-input, single-output EIV GP from Appendix B.2 is not applied here. Compare the plot in Figure B.16 with those of Figure B.19, which both plot the posterior samples of the distance-scaling parameter corresponding to the first input variable (which is the same input variable in both cases) against the model error standard deviation for the same output variable (with the output variable being the sole output in Figure B.16, and the output variable being the first of two output variables in Figure B.19). There are two plots in the case of Figure B.19, since there are two covariance kernels. As previously discussed with this example, small changes in the input variable correspond to large changes in the output variable, meaning values of the distance-scaling parameter must be relatively small, in order to capture the appropriate model fit. As the model error standard deviation becomes larger, this allows larger values of the distance-scaling parameter to be captured (see Figure 4.28 for a visualisation of this). It had appeared that a cutoff for this behaviour was around $\sigma_\epsilon = 0.075$, where values below this restricted the distance-scaling parameter to be small, and values above this allowed for more variation in the values of the distance-scaling parameter. This does not occur in Figure B.19 – it is evident that there is a lot of variation in values of the distance-scaling parameter (for both $l_{1,1}$ in the left plot, and $l_{2,1}$ in the right plot), even when the model error standard deviation for the first output variable is below the cutoff of 0.075. It must be noted that there does seem to be some a dip in the values of the distance-scaling parameter around this cutoff of $\sigma_{\epsilon_1} = 0.075$, which suggests that there is still some relationship between the distance-scaling parameters and the model error standard deviation.

Two plots of the joint posterior samples of $l_{2,1}$ against $l_{1,1}$ are considered in Figure B.20. Note again that $l_{1,1}$ represents the distance-scaling parameter for first input variable in the SE-ARD kernel for the first covariance function, and $l_{2,1}$ is the analogous parameter for the second covariance function. The left plot provides the posterior samples for $l_{1,1}$ and $l_{2,1}$ without any restrictions on σ_{ϵ_1}, whereas the right plot considers the posterior samples of those distance-scaling parameters with the restriction that $\sigma_{\epsilon_1} < 0.075$. An interesting observation is noted here, that small values of $l_{1,1}$ correspond to large values of $l_{2,1}$, and vice versa, which is exemplified by restricting σ_{ϵ_1} to those values which provide an appropriate model fit. It is plausible that some reparameterisation of the EIV MOGP that considers a relationship between the distance-scaling parameters of the first covariance function and those of the second covariance function, but this is not explored here.

An alternative solution for trying to reduce the model error standard deviation for the first output variable in the EIV MOGP simulation is to

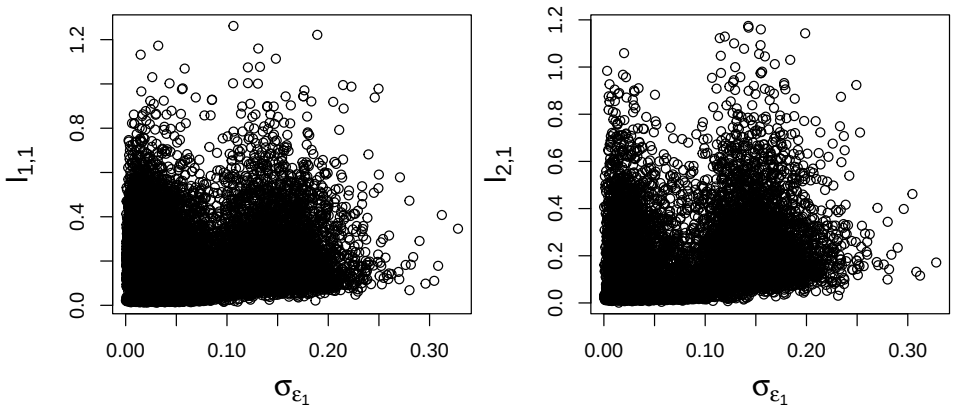

Figure B.19: Left plot presents the joint posterior samples of $l_{1,1}$ against σ_{ϵ_1}, and the right plot presents $l_{2,1}$ against σ_{ϵ_1}, both from the EIV MOGP, with the equivalent prior specification as mentioned in the caption of Figure B.18.

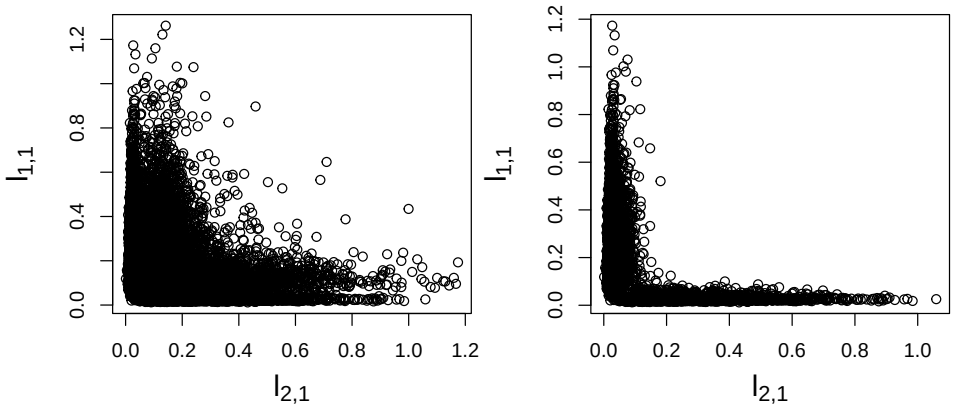

Figure B.20: The joint posterior samples of $l_{2,1}$ against $l_{1,1}$ from the EIV MOGP, with the equivalent prior specification as mentioned in the caption of Figure B.18. The left plot provides those samples without any restrictions on σ_{ϵ_1}, whereas the right plot provides the posterior samples of the distance-scaling parameters where σ_{ϵ_1} is less than 0.075.

adjust the scaling for the first input variable. It has been clarified throughout the investigation of the EIV GP with the simulations used in this work that small values of the distance-scaling parameter are required for the first input variable, in the case where the underlying relationship between this input variable and the output variable is either a sine curve (see Section 4.3.1) or some combination of a sine curve and exponential curve (see Section 4.3.4). The reason for this requirement is due to the small differences in the input variable resulting in large differences in the output variable, and the reasons this requirement is hard to fulfil are multiple – the prior on the model error standard deviation, the prior on the distance-scaling parameter, and also the scaling of the input variable itself. Changes relating to the first two reasons have been considered so far in Appendix B.1 and Appendix B.2, and now a change relating to the scaling of the input variable is considered.

The prior distribution of the distance-scaling parameter that has been effective up to now is the Gamma$(2, \frac{2}{0.25})$ distribution. It is noted that the extreme values of the distance-scaling parameter that are required are still relatively low-density values in this prior distribution. For example, the posterior mode of l_1 in the two-input example is 0.0212 (to 3 s.f.); the probability of sampling a smaller value than this from the distribution Gamma$(2, \frac{2}{0.25})$ is 0.0128, indicating it is an extreme value in this prior distribution. Changing the prior is a possibility, but also note that the right tail of this distribution must be long enough to capture larger values of the distance-scaling parameter. This is not considered here.

Instead, the observed data of the input variable could be scaled differently. While it has been indicated that this example has been designed to be extreme, since the input variable true values are chosen to span relatively small values (in this case, [0.01,0.13]), it is clear that this scaling is causing some issues for the model. A simple solution could be to choose a better way to scale the data. A straightforward solution is provided here, simply by multiplying the observed data for the first input variable by 10, so that they span the range [0.1,1.3]. In a sense, this is a manual override of the data-scaling process, where the main difference now is that the model should (in theory) identify larger values of the distance-scaling parameter, since the function is more stretched in the direction of the input variable (i.e., small differences in the input variable result in smaller difference in the output variable than with the previous scaling). This scaling adjustment is only carried out here; it is recommended that an adjustment to the data-scaling process be considered in the future (see Chapter 6).

The EIV MOGP prior specification is adjusted slightly to account for the different scaling for the first input variable. The diagonal elements of

the scale matrix $S_{\tilde{X}}$ of between-materials precision matrix $T_{\tilde{X}}$ were 15.3664, which was built using the probability statement that $P(0 \leq \tilde{X}_i \leq 1) = 0.95$. This assumption is no longer justified because of the different scaling, so as a compromise, the diagonal elements for this scale matrix are now 6 (with the assumption that the mean of $\tilde{X}_i$ is 0.5, the upper bound of the 95% confidence interval assuming a standard deviation of $\sqrt{\frac{1}{6}}$ is $0.5+1.96\times\sqrt{\frac{1}{6}} = 1.300$ to 4 d.p.). Similarly, the scale matrix S_δ for the measurement-error precision matrix T_δ is adjusted to suggest more variation in the measurement error, and so the diagonal element is changed from 1522.07 to 1000.

With these adjustments, the EIV MOGP is fitted to the data. An adaptation phase of length 2000 samples was required for optimal sampling behaviour in the algorithm, and sufficient levels of mixing and convergence were observed with 20000 posterior samples stored by taking every 10^{th} sample from 200000 draws, after having discarded (i.e., the burn-in) 25000 samples. The marginal posterior densities of the model error standard deviations σ_{ϵ_1} (left plot) and σ_{ϵ_2} (right plot) from this model, as well as the previous fit with scaling the first input variable (roughly) onto [0.01,0.13], are given in Figure B.21. The red curves denotes the posterior densities from the rescaled model, and the black curves correspond to those from the model fitted with the previous scaling. Considering the left plot of Figure B.21 with the posterior densities of σ_{ϵ_1}, it appears that changing the scaling of the first input variable provides a noteworthy improvement in posterior, with the multimodality from the previous fit (black curve) being eliminated in the current fit (red curve). Furthermore, it is noted that $P(\sigma_{\epsilon_1} > 0.075) = 0.0184$ in the current fit, compared with $P(\sigma_{\epsilon_1} > 0.075) = 0.245$ in the previous fit. Conversely, the adjustment to the scaling of the first input variable has notably influenced the parameter σ_{ϵ_2} – the density at values greater than 0.01 is much larger for the current fit. While this compromise is not ideal, the current model fit seems preferable due to the elimination of the multimodality of σ_{ϵ_1}.

The differences between the distance-scaling parameters for the first input variable from the two models are compared in Figure B.22 with marginal posterior density plots. The first row of plots corresponds to the current model fit; the second row corresponds to the previous model fit; the first column corresponds to the distance-scaling parameter $l_{1,1}$; the second column corresponds to $l_{2,1}$. With the difference in model error standard deviation for the first output variable noted above, it is unsurprising that there are differences between the models (so comparing row 1 with row 2) in the posterior densities of the distance-scaling parameters. To clarify, given the

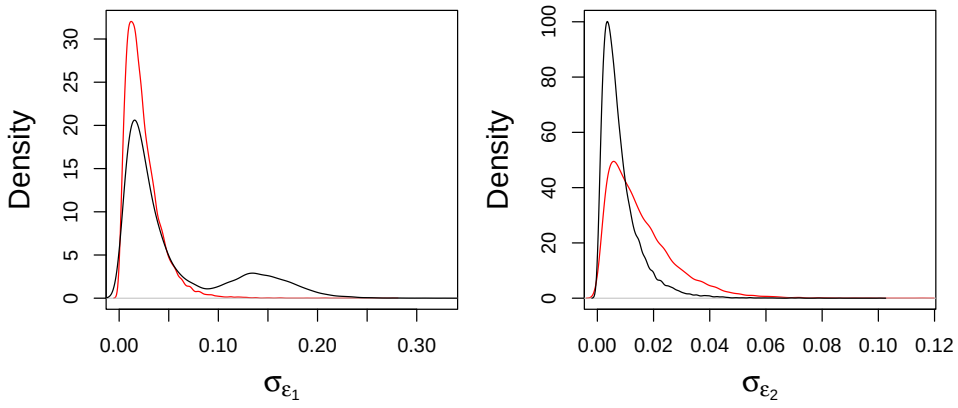

Figure B.21: The marginal posterior densities of σ_{ϵ_1} (left plot) and σ_{ϵ_2} (right plot) from the previous EIV MOGP model fit (black curves) with the old scaling, compared with those from the current model fit (red curves), where the first input variable is unscaled.

only change in the model is to adjust the scaling of the first input variable, specifically by multiplying its data by 10, the model fits would be, in a sense, equivalent if the posterior samples of the distance-scaling parameters were divided by 10. That is, the SE-ARD kernel

$$\exp\left\{-\frac{(\tilde{X}_{1,i}-\tilde{X}_{1,i'})^2}{2l_{1,1}^2}-\frac{(\tilde{X}_{2,i}-\tilde{X}_{2,i'})^2}{2l_{1,2}^2}\right\}$$

is equivalent to

$$\exp\left\{-\frac{\left(\frac{\tilde{X}_{1,i}-\tilde{X}_{1,i'}}{10}\right)^2}{2\left(\frac{l_{1,1}}{10}\right)^2}-\frac{(\tilde{X}_{2,i}-\tilde{X}_{2,i'})^2}{2l_{1,2}^2}\right\}.$$

The plots of the second column are somewhat similar; the posterior mode of roughly 0.25 in the current fit (top-right plot) looks similar to 10 times the posterior mode of the previous fit (bottom-right plot), albeit there is some density beyond values of 0.1 in the bottom-right plot, which is not found in the top-right plot (i.e., little density at values beyond 1). In contrast, the posterior densities for $l_{1,1}$ are notably different. Firstly, the posterior mode of the current fit (top-left plot, roughly a value of 0.25) is the smaller of two modes in the previous fit (bottom-left plot, roughly a value of 0.025),

where it looks like the values are different by a factor of 10. The secondary peak of the top-left plot, and the primary peak of the bottom-left plot, are notably different, considering the difference of a factor of 10 in the values. The rough maximum value of the top-left plot is 1.5, which appears close to the posterior mode in the bottom-left plot. Having changed the scaling of the input variable, the model is essentially able to find relatively smaller values of the distance-scaling parameter, which has improved the model fit. This is also in part due to the prior distribution Gamma$(2, \frac{2}{0.25})$, where the probability of observing a value larger than 0.1 in the prior, which would be unsuitable for the previous scaling of the first input variable, is 0.809, whereas the probability for the value of 1, which would be equally unsuitable for the current scaling of the first input variable, is 0.00302.

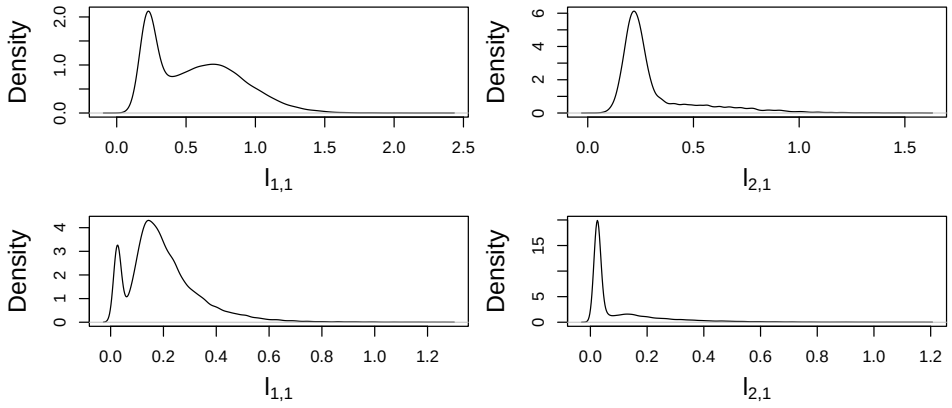

Figure B.22: The marginal posterior densities of $l_{1,1}$ (first column) and $l_{2,1}$ (second column) from the current EIV MOGP model fit (first row) and the previous EIV GP model fit (second row).

Finally, the EIV MOGP posteriors are compared, that is, the predictions of the response variable for given vectors of the input variables. The details for how this is carried out are given in Section 4.3.5. The first of these figures, Figure B.23, is for the previous model fit; this demonstrates the mean predictions of the response variables (solid black curve) as well as the uncertainty of these predictions (95% predictions intervals, given by dotted black lines), which can be compared with the underlying values of the response variables (i.e., the values of the response variable produced by the underlying simulation function, given by the red curve). The predictions of the response variables are carried out jointly using the EIV MOGP, then plotted separately, with the first column corresponding to the first response variable,

and the second column corresponding to the second response variable. The first row of plots indicates predictions of the response variable given the input vectors that satisfy the equation $X_2 = \frac{20}{3}X_1 + \frac{1}{30}$, with $X_1 \in [0.01, 0.13]$. Similarly, the second row of plots is predictions for the input variables satisfying the straight line $X_2 = -\frac{20}{3}X_1 + \frac{29}{30}$, with $X_1 \in [0.01, 0.13]$. This figure is compared with Figure B.24, which demonstrates the analogous plots for the current model fit (adjusting the equation relating the second and first input variables for which the outputs are predicted). It is evident that there is an improvement in the predictions for the first output variable, but more uncertainty is observed in the predictions for the second output variable.

In essence, what this shows is that it is very difficult to find a prior distribution for the distance-scaling parameter that is appropriate for an input variable whose observed data, for example, range from 0.01 to 0.13 as well as for another input variable whose observed data range from 0.1 to 1, for example. As suggested here, a manual override of the data-scaling process could be a possible workaround for this, but a more preferable solution would be to explore a different method for scaling the data, so as to ensure that the range of [0,1] is sufficiently covered. In effect, the data standardisation carried out here, at least for the EIV GP, may be insufficient at standardising.

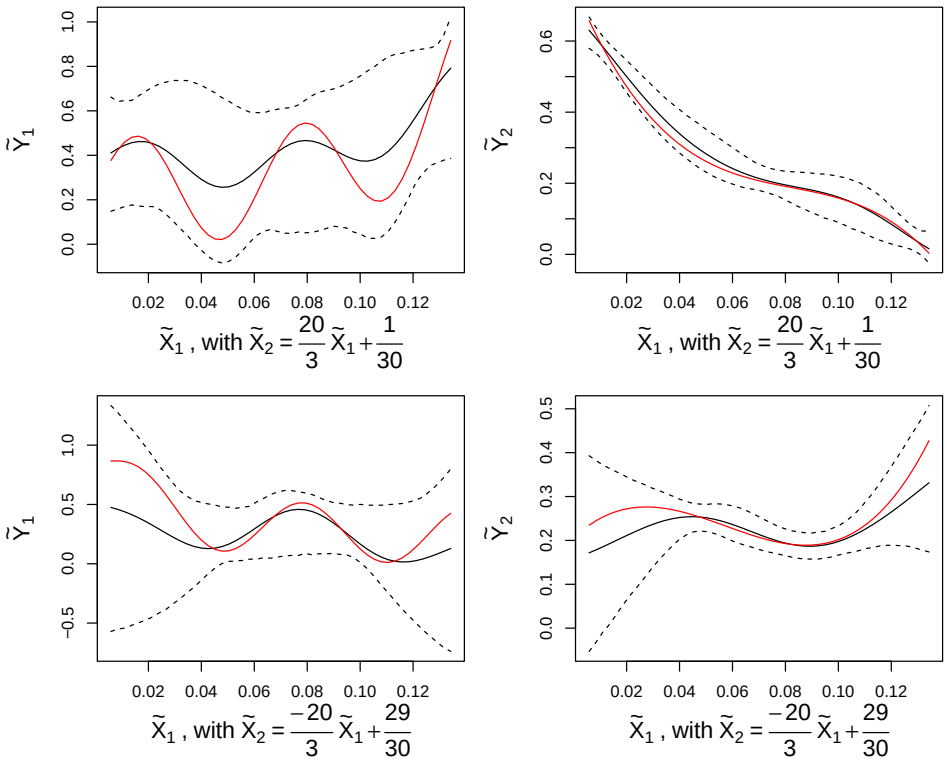

Figure B.23: Joint predictions of the output variables given the lines $X_2 = \frac{20}{3}X_1 + \frac{1}{30}$ (first row) and $X_2 = -\frac{20}{3}X_1 + \frac{29}{30}$ (second row). The predictions of the first output variable are given in column 1, and for the second output variable in column 2. The solid black lines provides the mean of the predictions, the dotted black lines provide 95% prediction intervals, and the red lines provide the underlying values of the output variables provided by the simulation function.

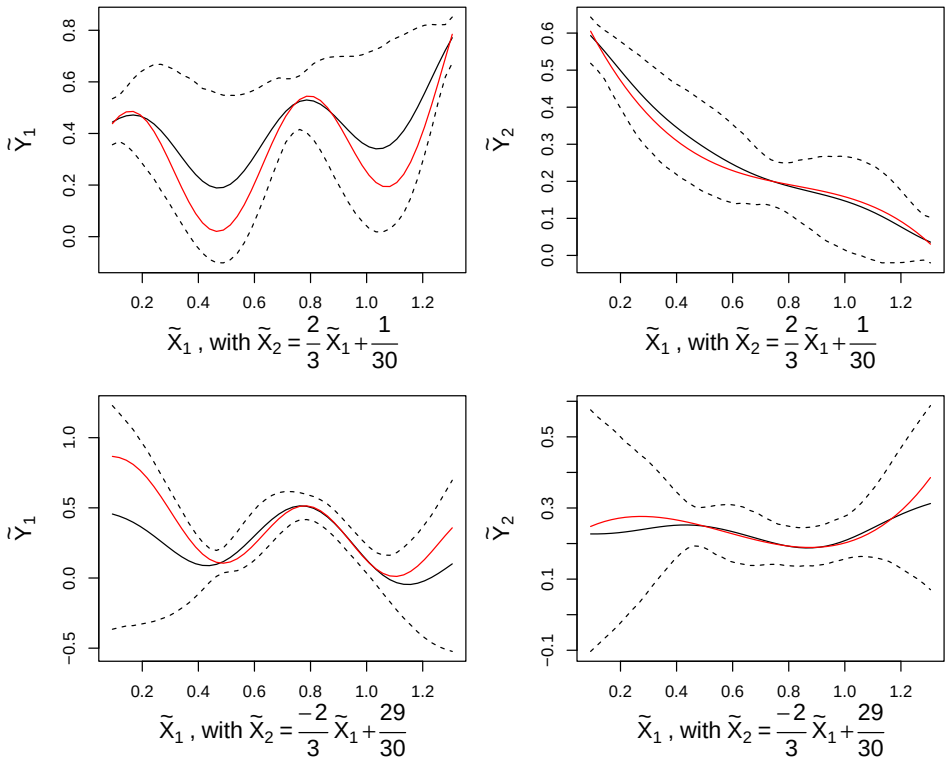

Figure B.24: Joint predictions of the output variables given the lines $X_2 = \frac{2}{3}X_1 + \frac{1}{30}$ (first row) and $X_2 = -\frac{2}{3}X_1 + \frac{29}{30}$ (second row). The predictions of the first output variable are given in column 1, and for the second output variable in column 2. The solid black lines provides the mean of the predictions, the dotted black lines provide 95% prediction intervals, and the red lines provide the underlying values of the output variables provided by the simulation function.

Appendix C

Informing and initialising priors for real data

C.1 Building informed priors for the real data

The probability statements from the fourth column of Table 4.10 are converted into parameters for probability distributions, following the same process established in Section 4.2.5. In order to do this, two further assumptions are required, as single point estimates that can be derived from the statements are not sufficient for describing a probability distribution with multiple parameters, such as the gamma distribution. The first of these assumptions that is made regards the ratio between the measurement error precision and the between-materials precision for the explanatory variables, that is, the ratio between τ_δ and $\tau_{\tilde{X}}$. It is expected that the variation of the subsamples would be *less* than the variation between the powders, that is, the precision of the subsamples would be *greater* than the precision between the powders, and so it is assumed that $P(\tau_\delta \geq \tau_{\tilde{X}}) = 0.99$ (this probability statement was enquired about in the elicitation, and the author has assumed the probability of 0.99, given the elicitation suggests it is expected that $\tau_\delta \geq \tau_{\tilde{X}}$). The second assumption that is made is that the shape parameters of the gamma prior distributions for τ_δ and $\tau_{\tilde{X}}$ is the same, which is equivalent to assuming that the measurement error precision behaves in the same way to the between-materials precision, but on different scales.

The process of producing the parameter estimates for τ_δ and $\tau_{\tilde{X}}$ is now demonstrated for the explanatory variable CBD. Consider the probability statement, corresponding to the elicited information for CBD,

$$P(0.35 \leq \tilde{X}_i \leq 0.55) = 0.95. \qquad \text{(C.1.0.1)}$$

Using the fact that, for any normal distribution $Y \sim \mathrm{N}(\mu, \sigma^2)$, where μ and σ^2 represent the mean and variance respectively, we have

$$P(\mu - 1.96\sigma \leq Y \leq \mu + 1.96\sigma) = 0.95,$$

which says that 95% of the probability density of the random variable Y lies in the interval $[\mu - 1.96\sigma, \mu + 1.96\sigma]$. Subtracting the lower bound from the upper bound gives the values 3.92σ, i.e., the width of the interval. Given that $\widetilde{X}_i$ is also normally distributed, and given the statement above in Equation C.1.0.1, it can be concluded that

$$P(0.35 \leq \widetilde{X}_i \leq 0.55) = 0.95 \implies 0.55 - 0.35 = 0.2 = 3.92\sigma_{\widetilde{X}},$$

where $\sigma_{\widetilde{X}}$ represents the standard deviation of the variable $\widetilde{X}_i$. This provides an estimate of the standard deviation of $\frac{0.2}{3.92} = 0.0510$, rounded to 3 significant figures. This can be rearranged to give an estimate of the precision, by squaring the value and then taking its reciprocal, giving 384.16. This value is taken to be the mean estimate of the precision $\tau_{\widetilde{X}}$. Consider now the probability statement

$$P(0.99\tilde{X}_i \leq X_{i,k} \leq 1.01\tilde{X}_i) = 0.95.$$

The value 0.45, the midpoint of the elicited range of true values, is assumed in this case for $\tilde{X}_i$ and thought of as the true value of CBD for an average powder. The above probability statement is then equivalent to

$$P(0.99 \times 0.45 \leq X_{i,k} \leq 1.01 \times 0.45) \iff P(0.4455 \leq X_{i,k} \leq 0.4545) = 0.95$$

and, following the same process as above, an estimate of the standard deviation of $\delta_{i,k}$, given by σ_δ, is

$$3.92\sigma_\delta = 0.009 \implies \sigma_\delta = 0.0230 \tag{C.1.0.2}$$

to 3 significant figures. The equivalent precision is therefore given by 189708.7, which is taken to be the mean estimate of the precision parameter τ_δ.

With Gamma distributions assumed for τ_δ and $\tau_{\tilde{X}}$, the mean estimates calculated above equate to

$$\frac{a_{\widetilde{X}}}{b_{\widetilde{X}}} = 384.16 \tag{C.1.0.3}$$

and

$$\frac{a_\delta}{b_\delta} = 189708.7, \tag{C.1.0.4}$$

where

$$\tau_{\widetilde{X}} \sim \text{Gamma}(a_{\widetilde{X}}, b_{\widetilde{X}}) \tag{C.1.0.5}$$

and

$$\tau_{\delta} \sim \text{Gamma}(a_{\delta}, b_{\delta}), \tag{C.1.0.6}$$

with the Gamma distributions parameterised by shape and rate. It is assumed that the shape parameters are equal, giving $a_{\tilde{X}} = a_{\delta} = a$. Choosing some value for a then provides the value for the rate of each distribution, using the mean estimates of the distribution provided above. What remains is to choose some value for a so that

$$P(\tau_{\delta} \geq \tau_{\tilde{X}}) = 0.99,$$

which is found in this case using simulation (that is, simulating samples of the distributions of τ_{δ} and $\tau_{\tilde{X}}$ and choosing a value of a such that a random sample of τ_{δ} is larger than $\tau_{\tilde{X}}$ 99% of the time). Following a trial and error process with the simulations, it is found that the value $a = 0.696$ satisfies the assumptions, which implies the informed probability distributions

$$\tau_{\widetilde{X}} \sim \text{Gamma}(0.696, 0.00181) \tag{C.1.0.7}$$

and

$$\tau_{\delta} \sim \text{Gamma}(0.696, 0.00000367), \tag{C.1.0.8}$$

with the rate parameters given to 3 significant figures. This process is performed analogously for SE and BFE; the measurement error precision τ_{η} is discussed after a comparison of the uninformed and informed priors below.

A comparison between the uninformative gamma prior distribution with shape 0.001 and rate 0.001, that has been used for both $\tau_{\widetilde{X}}$ and τ_{δ}, and the newly constructed informed prior distributions, is now given. Consider firstly a plot of the probability density function of Gamma$(0.001, 0.001)$, given in Figure C.1, which shows that the majority of the density of the distribution is found at values close to 0 (meaning a sample from the distribution is likely to be relatively close to 0). Furthermore, consider the plot in Figure C.2, which plots the probability that a random sample from Gamma$(0.001, 0.001)$ is greater than the value x (this is referred to as the survival function of a probability distribution). Note the range of the x-axis in this plot of [0,1], and that the probability of a random sample being greater than 0.2 (for example) is 0.0079, i.e., it is very unlikely for a random sample to be greater than 0.2. An analogous plot to Figure C.2 is then provided in Figure C.3, with the range of x now [0,1000]. Provided

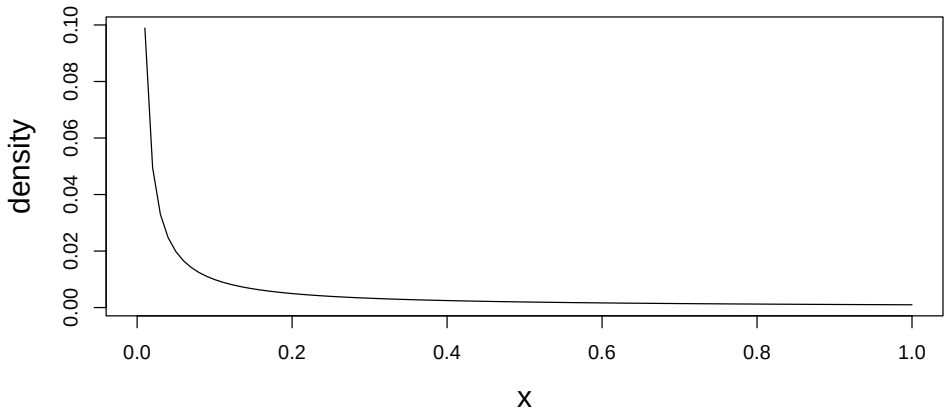

Figure C.1: The probability density function of some uninformative prior $\tau \sim \text{Gamma}(0.001, 0.001)$

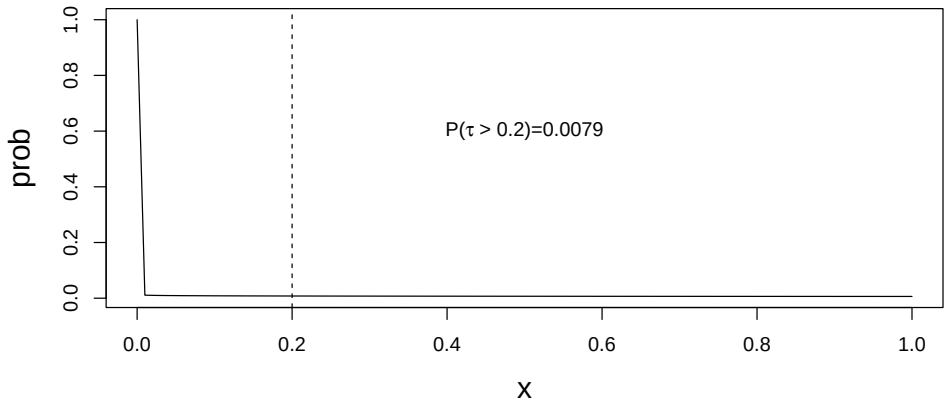

Figure C.2: A plot of the probability that a random sample from the uninformed prior $\tau \sim \text{Gamma}(0.001, 0.001)$ is greater than x, over the domain [0,1] (i.e., a plot of the survival function), with a vertical dotted line at $x = 0.2$.

on the plot is also the mean estimate of $\tau_{\widetilde{X}}$ from the elicitation, with the probability that a random sample from the uninformative prior is greater than the estimate being 0.00073. In other words, the probability that the uninformative gamma prior produces a sample greater than the mean estimate of between-materials precision for CBD is very small. A final plot

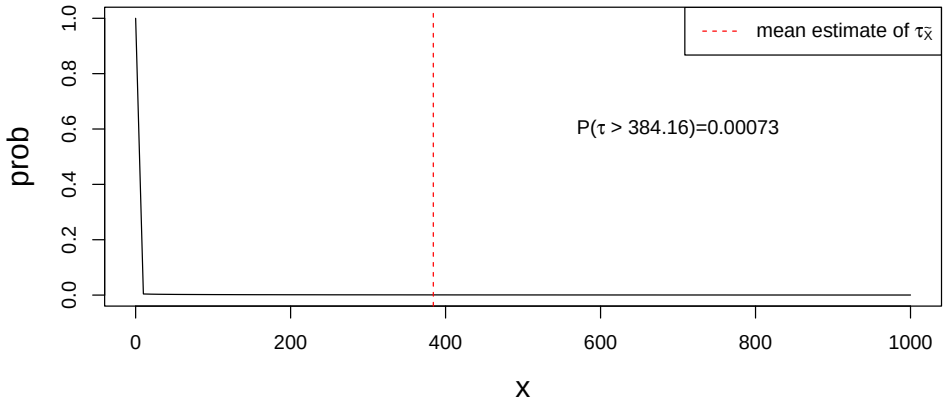

Figure C.3: A plot of the probability that a random sample from the uninformed prior $\tau \sim \text{Gamma}(0.001, 0.001)$ is greater than x, over the domain [0,1000] (i.e., a plot of the survival function), with a vertical line at the mean estimate of the between-materials precision for CBD.

relating to $\text{Gamma}(0.001, 0.001)$ is given in Figure C.4, which plots the same function, now over the domain $[0,10^6]$. In this case, the mean estimate of τ_δ is given, with the probability of a random sample from the uninformative prior distribution being larger than it given by 0, to 10 decimal places. The uninformative prior distribution has such little density in the right tail of its distribution, that the mean estimates of these precision parameters are such extreme samples from this prior, particularly for the measurement error precision τ_δ.

The plot in Figure C.5 is a plot of the survival function of the informed prior for $\tau_{\widetilde{X}}$, given in Equation C.1.0.7, over the domain [0,10000]. This demonstrates the suitability of this distribution over the uninformative prior distribution previously considered, with the probability of sampling larger precision values like the mean estimate of $\tau_{\widetilde{X}}$ being more likely to occur with this informed prior distribution. Similarly, the plot in Figure C.6 is a plot of the survival function of the informed prior for τ_δ, given in Equation C.1.0.8, over the domain $[0,10^6]$. Again, this demonstrates the suitability of this distribution over the uninformative prior distribution previously considered.

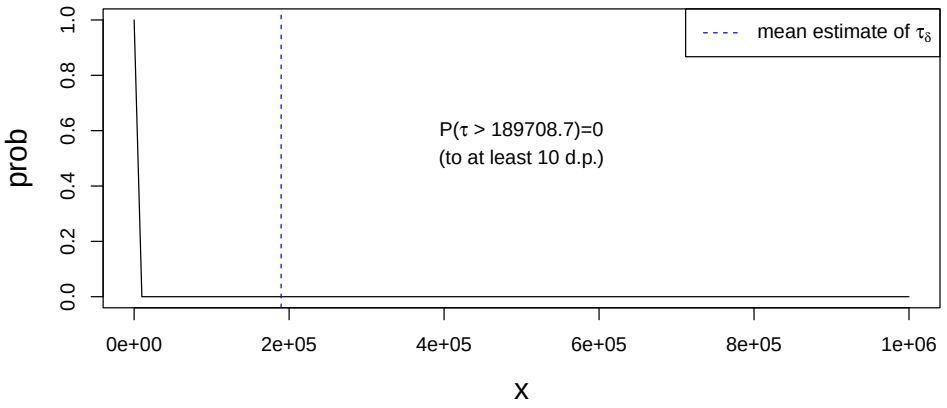

Figure C.4: A plot of the probability that a random sample from the uninformed prior $\tau \sim \text{Gamma}(0.001, 0.001)$ is greater than x, over the domain $[0,10^6]$ (i.e., a plot of the survival function), with a vertical line at the mean estimate of the measurement-error precision for CBD.

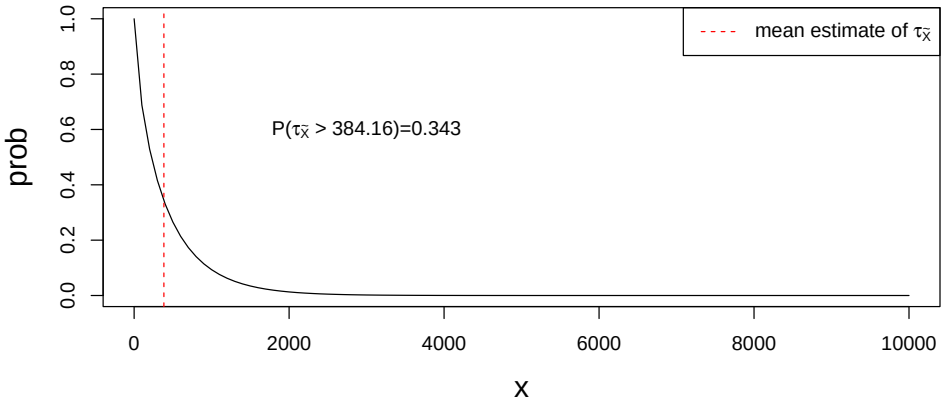

Figure C.5: A plot of the probability that a random sample from the informed prior distribution $\tau_{\tilde{X}} \sim \text{Gamma}(0.696, 0.00181)$ is greater than x, over the domain [0,10000] (i.e., a plot of 1 minus the cumulative distribution function)

This shows that the probability of sampling larger precision values like the mean estimate of τ_δ is more likely to occur in this informed prior distribution. Note that, the survival function of these informed prior distributions evaluated at their respective means gives the same probability because the shapes of the distributions are equal to one another.

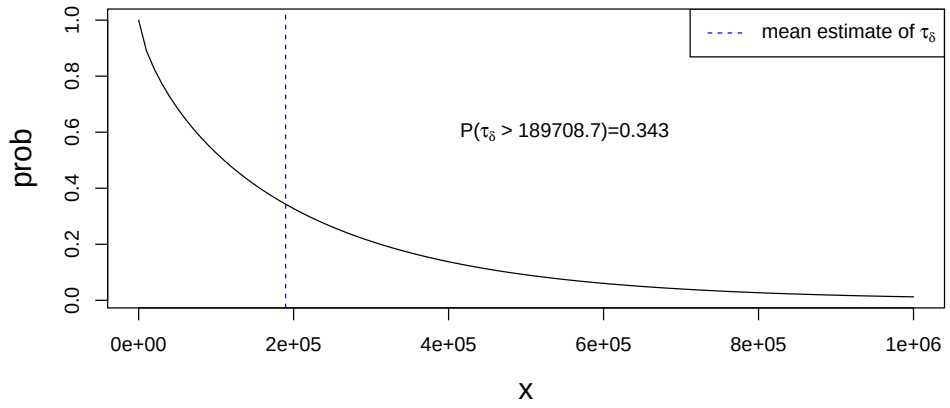

Figure C.6: A plot of the probability that a random sample from the informed prior distribution $\tau_\delta \sim \text{Gamma}(0.696, 0.00000367)$ is greater than x, over the domain $[0,10^6]$ (i.e., a plot of 1 minus the cumulative distribution function)

The means of the prior distribution of the true values are also informed by the elicitation, given the ranges of the true values for CBD, SE and BFE. That is, for a given powder i, the mean of the joint distribution of true values for CBD, SE and BFE (in that order below), is given by

$$\boldsymbol{\mu}_{\tilde{X}} = \begin{pmatrix} 0.450 \\ 0.650 \\ 0.889 \end{pmatrix},$$

to 3 significant figures.

The remaining precision parameters that require thought are the measurement error precision τ_η and the model error precision τ_ϵ (the univariate case is dealt with here, and further discussion of the multivariate case is given below). While the between-materials precision and measurement error precision for an explanatory variable is investigated jointly above, this is not as straightforward to do for the measurement error precision of the response variable and the model error precision. This is because the model

error precision measures the variability in the model error, whereas the between-materials precision measures the variability of the true values for the variable—these ideas are not equivalent.

A way of building a lower bound for the model error precision τ_ϵ is to consider a univariate regression model with just an intercept term, that is,

$$\tilde{Y}_i = \beta_0 + \epsilon_i,$$

where $\epsilon_i \sim \mathrm{N}(0, \tau_\epsilon)$. Being the simplest linear model to be fitted, a model containing any further information should provide at least the same quality of model fit, and so the standard deviation of the model error can only be decreased by providing the model with more information. The maximum possible range of tapped density values is given by the true value range provided by the elicitation of 0.494 to 0.725, so an upper bound of the model error standard deviation can be estimated by

$$\frac{0.725 - 0.494}{3.92} = 0.05892857,$$

giving an estimate of the lower bound of the model error precision τ_{ϵ_1} (for tapped density) of

$$\frac{1}{0.05892857^2} = 287.9706.$$

In order to use a gamma prior distribution for τ_{ϵ_1}, there must be some compromise for the density at values lower than this estimate; if this density were to be minimised, a large value of the shape parameter would have to be chosen, which would force all of the prior density to be close to this lower bound of model error precision. The shape and rate parameters of this gamma distribution needs to be able to satisfy the following (rough) probability statements:

- the probability $P(\tau_{\epsilon_1} < 287.9706)$ must be as small as possible,
- the probability $P(287.9706 < \tau_{\epsilon_1} < 350)$ (350 chosen arbitrarily, to correspond to relatively small levels of model error precision given tapped density values) must be not too small,
- the probability $P(\tau_{\epsilon_1} > 10000)$ (10000 chosen arbitrarily, to correspond to relatively large levels of model error precision) must be not too small.

Satisfying these criteria would allow for a prior distribution that makes it unlikely for unrealistic model error precisions to occur, makes it somewhat

likely that relatively poor model fits with (relatively) small model error precisions can occur, and makes it somewhat likely that relatively good model fits with (relatively) large model error precisions can occur. Finding suitable shape and rate parameters to satisfy these criteria is very difficult, and some compromise must be made, which falls on making density below the lower bound estimate being larger than preferred, so that the density at values just above the lower bound isn't too small. After trial and error, a seemingly suitable informed prior distribution for the model error precision is with shape $a_{\epsilon_1} = 0.2$ and rate $b_{\epsilon_1} = 2 \times 10^{-6}$. For this prior, $P(\tau_{\epsilon_1} \leq 287.9706) = 0.245$, $P(287.9706 \leq \tau_{\epsilon_1} \leq 350) = 0.010$ and $P(\tau_{\epsilon_1} \geq 10000) = 0.010$=0.504 (all to 3 decimal places).

The informed prior for the measurement error precision on the response variable is of interest. An estimate of the mean of the informed gamma prior is built analogously to the above example with CBD. In this case, a comparison with the between-materials precision is not possible in order to find a suitable value for the shape of the distribution. Instead, this value is assumed to be 0.6. Given the cases for the measurement error precision of the explanatory variables (see Table 4.11), appropriate shape parameters appear to range from 0.2 to 1, and so a value of 0.6 is chosen. The mean estimate of the measurement error precision from the elicitation, using the 2σ rule to estimate the standard deviation and rearranging to find the precision, is given by 103410.5, and so $a_{\eta_1} = 0.6$ and $b_{\eta_1} = \frac{0.6}{103410.5}$.

The prior distributions for the model error precision and measurement error precision for AoR are estimated analogously to those of tapped density. In the case for AoR, the lower bound estimate for the model error precision is 198.8303, leading to shape and rate parameters of $a_{\epsilon_2} = 0.2$ and $b_{\epsilon_2} = 3 \times 10^{-6}$. The mean estimate of the measurement error precision derived from the elicitation is 10154.84, leading to shape and parameter estimates of $a_{\eta_2} = 0.6$ and $b_{\eta_2} = \frac{0.6}{10154.84}$.

The above description of finding the informed prior distributions for the measurement error precision and between-materials precision have been carried out in the case of considering one explanatory variable and one response variable in a regression model. When considering multiple explanatory variables, it is the case that the true values $\tilde{\boldsymbol{X}}_i$ are multivariate normally distributed, with some mean vector and precision matrix $T_{\tilde{X}}$. In this case, a Wishart prior distribution is placed on the precision matrix $T_{\tilde{X}}$, and the informed Wishart prior is formed by a diagonal scale matrix whose diagonal elements are the mean estimates of between-materials precision from the elicitation. This applies also for the case where the replicate measurements are assumed to be jointly distributed, with the scale matrix S_δ being a diagonal

matrix whose diagonal elements are the mean estimates of the measurement error precision from the elicitation. For the model error precision matrix T_ϵ, the scale matrix for the Wishart prior is also a diagonal matrix whose diagonal elements are the means from the informed priors distributions for τ_{ϵ_1} and τ_{ϵ_2} (i.e., for tapped density, $S_{\epsilon_{1,1}} = \frac{a_{\epsilon,1}}{b_{\epsilon,1}} = \frac{0.2}{2\times 10^{-6}} = 10^5$. These are summarised in Table 4.11.

The priors for the precisions of those variables that do not have any prior information are also adjusted from the simulation section. Given the data is scaled onto the range [0,1] for each variable, this directly provides a range of true values for the variable, from which a mean estimate of the between-materials precision can be derived, given by $15.3664 = 3.92^2$. An estimate of the measurement error precision is hard to come by, so the assumption that the shape parameters are the same is carried over and strengthened further by assuming it is equal to 1. Using simulation, a value of the rate parameter b_δ is chosen so that $P(\tau_\delta > \tau_{\tilde{X}}) = 0.99$ is satisfied, given $a_\delta = a_{\tilde{X}} = 1$ and $b_{\tilde{X}} = \frac{1}{15.3664}$. A suitable value is $b_\delta = 0.000657$, to 3 significant figures.

The corresponding model coefficients for the explanatory variables also take informed prior distributions. For the model intercept terms, these are normally distributed, taking the midpoints of the elicited range of true values for the corresponding response variable as means, and precision values of 0.0001 (i.e., a standard deviation of 100). The small value for the precision is used as there is no much information available for the intercept term, but it should be an easy parameter to estimate. The slope terms are also normally distributed, with mean 0—the precision parameters are informed by the possible range of values the slope can take given the response variable range and the explanatory variable range. This is demonstrated using TD and CBD—the maximum value that the slope $\beta_{\text{CBD,TD}}$ can take is given by

$$0.494 + \beta_{\text{CBD,TD}}(0.55 - 0.35) = 0.725,$$

which implies the maximum value of $\beta_{\text{CBD,TD}} = \frac{0.725-0.494}{0.55-0.35} = 1.155$. A minimum value is therefore given by $\beta_{\text{CBD,TD}} = -1.155$, which provides a range of possible slope values of $[-1.155, 1.155]$. In this case, additional uncertainty is assumed for the standard deviation, assuming a 90% interval instead of a 95% interval—this is because the precision is fixed for the model coefficient prior, and so more uncertainty is placed on the fixed value. That is, for any normal distribution $Y \sim \text{N}(\mu, \sigma^2)$, where μ and σ^2 represent the mean and variance respectively, we have

$$P(\mu - 1.645\sigma \leq Y \leq \mu + 1.645\sigma) = 0.9,$$

and so the width of the interval is set equal to 3.29σ. This gives an estimate of the standard deviation of $\beta_{\text{CBD,TD}}$ of $\frac{2\times 1.155}{3.29} = 0.7021277$, which is rearranged to give the precision estimate 2.028466. Therefore, the informed prior distribution for $\beta_{\text{CBD,TD}}$ is a normal distribution with mean 0 and precision 2.028 to 3 decimal places. The analogous process is carried out for the variables SE and BFE, and similarly for some generic explanatory variable for which no prior information is available; in this case the fact that the data is scaled onto the range [0,1] is utilised. The prior distributions for the model coefficients are summarised in Table 4.12.

C.2 Initialisation

A formal process is defined here for choosing initial values for the MCMC simulation when considering the real data.

Firstly, a Latin hypercube sampling is simulated to provide quantiles for covering as much of the multivariate sampling space for the prior distributions as possible. The number of samples of the LHS is determined by the number of parallel chains, which is four in this case. The dimensionality of the LHS is dependent on the number of prior distributions to be initialised, which is dependent on the model to be fitted. In this work, the prior distributions are used to choose initial values that are overdispersed—overdispersed in the prior distribution is a suitable alternative to choosing values that are overdispersed with respect to the posterior distribution. It is implemented here as there are several models to be fitted, which is time-consuming to have to fit twice (the first time to provide a posterior distribution, the second time to choose overdispersed values from the posterior to use as initial values for the priors in the following model fit).

The notable detail of the initialisation is how the precision matrices are chosen to ensure they are overdispersed with respect to the posterior distribution. In this work, this is carried out by sampling a large number of precision matrices from the Wishart prior distribution, and evaluating the determinant of each matrix sample, then choosing overdispersed values with respect to the determinants. In this sense, the determinant is used to compare matrices with one another, with large determinants likely to be found with large diagonal elements and small off-diagonal elements, and small determinants likely to be found with small diagonal elements and (relatively) large off-diagonal elements. So, if the quantile simulated from the LHS is 0.34, then the matrix whose determinant is closest to the 34% quantile of the determinants is chosen as the initial matrix.

The remaining parameters to be initialised in the model are treated independently, i.e., using the respective marginal prior distribution to choose overdispersed values. As an example, for a given powder i, the true values of powder i for the explanatory variables are assumed to have a joint normal distribution, i.e.,

$$\tilde{\boldsymbol{X}}_i = (\tilde{X}_{1,i}, \ldots, \tilde{X}_{n_g,i})' \sim \mathrm{N}_{n_g}(\boldsymbol{\mu}_{\tilde{X}}, T_{\tilde{X}}).$$

Despite this, the true value for each explanatory variable is randomly selected using the LHS from the marginal distributions, for example for the first explanatory variable, $\tilde{X}_{1,i} \sim \mathrm{N}(\mu_{\tilde{X}_1}, S_{\tilde{X}_{1,1}})$, using the corresponding diagonal element of the scale matrix $S_{\tilde{X}}$ for the precision matrix $T_{\tilde{X}}$ as the precision.

The model intercept parameters are treated slightly differently, as their prior distributions are very vague. Choosing a starting value from this prior leads to the intercept terms for the model getting stuck at the starting value in the MCMC, and it is very likely that the posterior distributions for the intercept will be concentrated at values close to the prior mean (given the scaling of the data and the likely influence of other explanatory variables on the model). Instead, initial values are chosen using the prior distribution for the intercept with the same mean as in Table 4.12 and precision 15.3664 (corresponding to a standard deviation of $\frac{1}{3.92}$.

Appendix D

Model plots

D.1 EIV BR model for simulated data

In this section, plots are provided from model fits to simulated data.

The simple-linear-model simulation is firstly considered, in particular the true values of the model. If it is observed that these values are not estimated appropriately given the observed data, then the rest of the model clearly will not have been fitted well. The estimated density of posterior samples from the first parallel chain are plotted in each case. The plot in Figure D.1 shows the posterior densities of true values of randomly selected materials (groups), along with green vertical lines corresponding to the observed data and a red vertical line corresponding to the 'chosen' true value (i.e., the chosen true value of $\tilde{X}_3$ for this simulation is 0.03), with these values also having been scaled onto [0,1].

The model appears to have estimated the true values very well, with the posterior densities appearing to be normally distributed (this is expected). The range on the x-axis has been determined by the minimum and maximum of the posterior samples for the corresponding true value. The visualisation provides a good indication of the variance of the posterior samples in each case, given the limits on the x-axis, and it is clear that each observed data point (green vertical lines) is indicated on the plots. It seems possible that the variance in the posterior for each true value is slightly larger than expected, given there is fairly significant density away from the observed values in each case. The analogous plot to Figure D.1 for randomly selected true values on the response variable is found in Figure D.2, with the same observations made above for Figure D.1 being applicable.

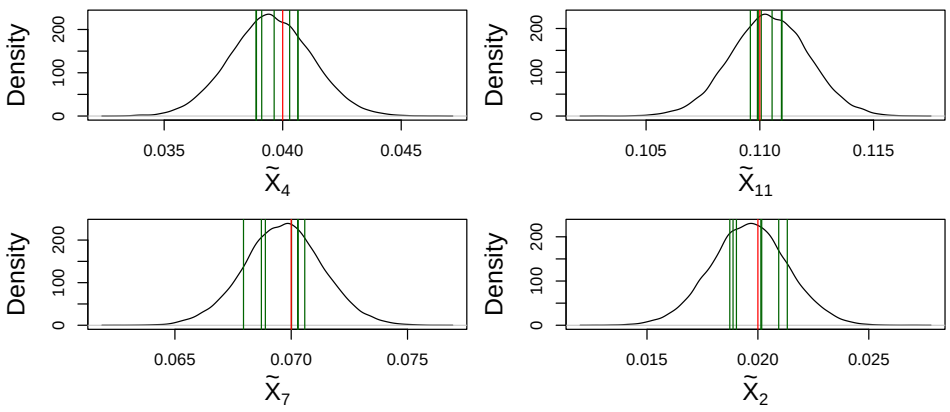

Figure D.1: Posterior densities for $\tilde{X}_4$ (top left), $\tilde{X}_{11}$ (top right), $\tilde{X}_7$ (bottom left), and $\tilde{X}_2$ (bottom right), along with the observed data from the corresponding material (group) from the simulation (vertical green lines) and the 'chosen' true values (vertical red lines).

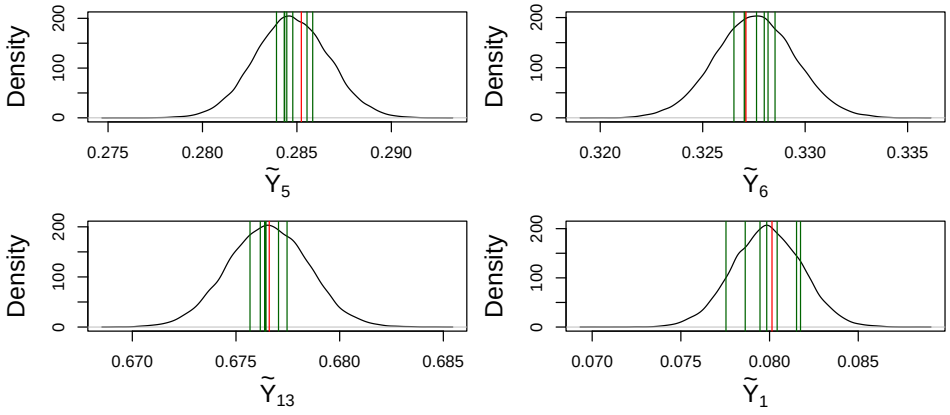

Figure D.2: Posterior densities for $\tilde{Y}_5$ (top left), $\tilde{Y}_6$ (top right), $\tilde{Y}_{13}$ (bottom left), and $\tilde{Y}_1$ (bottom right), along with the corresponding observations of each true value from the simulation (vertical green lines) and the 'chosen' true values (vertical red lines).

D.2 EIV BR model for powder flow data

In this section, some plots are provided relating to the EIV BR model fit to the real data with all seven input variables considered simultaneously (i.e., a seven-covariate EIV BR model). The plots in Figure D.3 demonstrate examples of the model recovering appropriate marginal posterior densities for the true values of each group for the variables tapped consolidation, CBD, TD, AoR respectively. These plots indicate the true values are recovered well, with posterior means lying close to the line $\tilde{Y} = Y$ of $\tilde{X} = X$ (i.e., true values of the response or explanatory variable close to observed data of the response or explanatory variable).

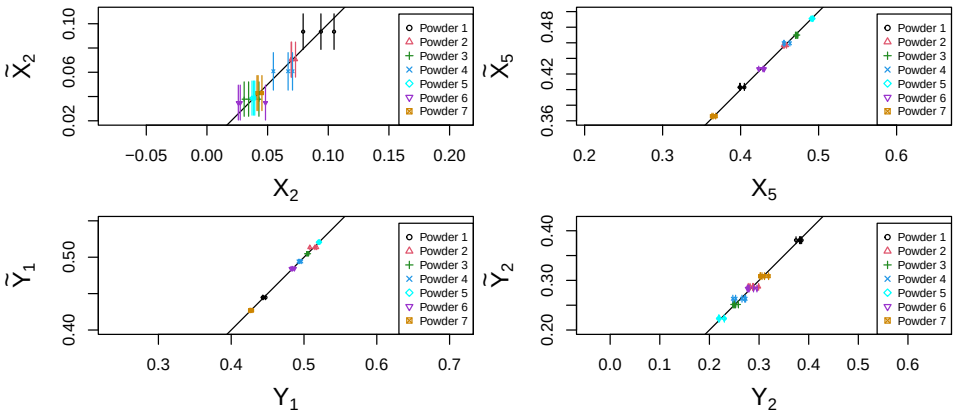

Figure D.3: Posterior densities, in the form of 95% credible intervals, plotted against the respective observed data for each group $i = 1, \ldots, 7$, for the variables tapped consolidation, CBD, TD and AoR (row-wise). Output is based on the fitted model including all seven explanatory variables.

Along with the comparisons between the posterior densities of true values and the respective observed data, the fitted values of the model are also inspected to see whether the relationship between the true values has been appropriately estimated. These fitted values plots are given for each response variable, i.e., these are marginal predictions of the response variables given the estimated relationship with the seven explanatory variables. These are given in Figure D.4. The fitted values plots indicate the model is estimating the relationship effectively—the mean fitted values align with the respective mean true value for both response variables for all groups. It is also clear from the ellipses that there is more uncertainty in the posterior distribution for the predictions of the response variable than there is is the estimates of

the true values, given the vertical radius of the ellipses is larger than the horizontal radius.

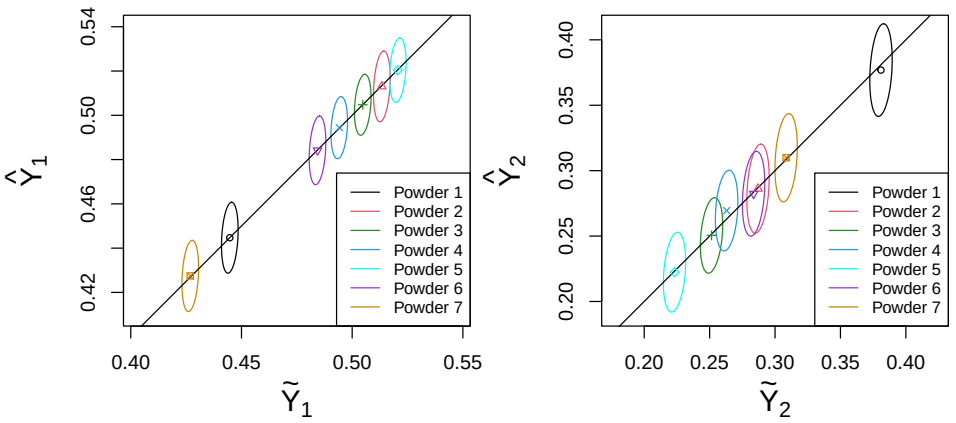

Figure D.4: Joint 95% credible ellipses for the joint distribution of fitted values and true values of the response variable TD. Output is based on the fitted model including all seven explanatory variables.

Appendix E

Backward model posterior convergence issues

The issues with convergence to the posterior distribution for the backward model, specifically the cubic simulation, is discussed below. Note that this issue of lack of convergence is prevalent in all simulations where a multimodal posterior distribution is expected.

Note that each backward model is run with two parallel chains, with both parallel chains starting at the same initial value for X_b^* (a random value between 0 and 1, which is the range of the uniform prior). The posterior mode is found for each parallel chain, giving two modes for each backward model. The posterior modes of those models that did not converge in the cubic simulation with $Y^* = 0.404$ are provided in the left plot of Figure E.1, and the equivalent for those models that did appear to converge are provided in the right plot of Figure E.1. These plots are noteworthy for many reasons—the first reason is that, in those models that did not converge, the reasons appears to be that the two parallel chains for a given model are converging to different modes. It should be noted that more of the posterior density should be investigated in each of these cases, to confirm whether or not the given chain is finding density only in the region immediately surrounding the posterior mode. A second noteworthy outcome is that, overall, the model is recovering the two 'real' values of $x^* = 0.0512$ and $x^* = 0.122$ that were highlighted above, and these are being recovered as high-density values in these chains. Thirdly, the models which have converged according to the PSRF estimates only appear to be converging because both parallel chains are identifying the same posterior mode.

To confirm whether the parallel chains are finding any posterior density

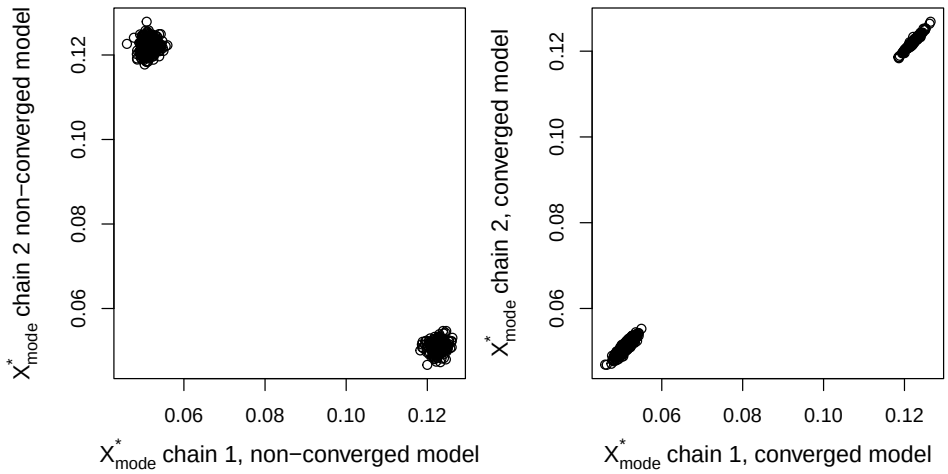

Figure E.1: Two plots comparing the posterior modes of the two parallel chains for each backward model run in the cubic backward model simulation, with $Y^* = 0.404$. The left plot corresponds to those backward model runs that did not converge, and the right plot corresponds to those run that did converge.

at both of the possible posterior modes, the minimum and maximum values of each parallel chain are examined. The four plots in Figure E.2 represent the differences between the maximum and minimum values within each parallel chain; the top row of plots represents the model that have apparently converged, and the bottom row represents those that have not converged, and the first column represents the differences from the first chains, and the second column represents the differences from the second chains. Unsurprisingly, given the models that have not converged (bottom row of Figure E.2, the differences between the maximum and minimum values are relatively small, particularly compared with the difference between the two 'real' values of $x^* = 0.0512$ and $x^* = 0.122$, given by 0.0709 to 3 significant figures. As suspected above, these same differences for the models that have 'converged' are also relatively small, and indicate that these 'converged' models are only finding one of the posterior modes, and zero density in the region surrounding the other posterior mode. However, the solitary backward model that has converged which sticks out in the top row of plots appears to have performed as desired, with the difference between the maximum and minimum posterior sample of X^*s being greater than the difference between the 'real' values of x^*. This backward model run corresponds to $b = 461$,

and the posterior density of X^*_{461} for this converged backward model is provided in Figure E.3. The posterior density of X^*_{461} represents the desired

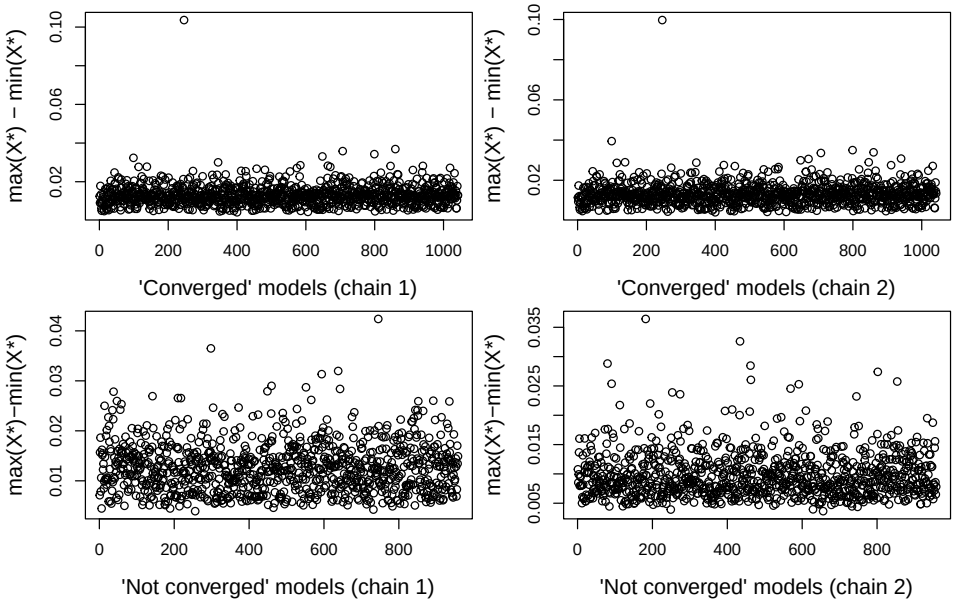

Figure E.2: Four plots considering the differences between the maximum and minimum values of the posterior samples of X^*_b from each parallel chain of each backward model run. The top row shows those differences for models that have converged, and the bottom row for those that have not converged. The first column corresponds to the first parallel chain from these models, and the second column the second parallel chain.

kind of outcome, which occurs only once from 2000 backward model runs, in which the model identifies the two possible posterior modes of X^* that could produce the corresponding desired output value Y^*.

The issue highlighted above with the model struggling to cope with a multimodal posterior distribution is common in Bayesian regression modelling. There are possible methods to deal with this, such as implementing a tempered transition sampler (the sampler itself is discussed in Gelman et al. (2013), and the recommendation for using it can be found in Plummer et al. (2003)), specifically designed to jump between distant modes of some multimodal posterior distribution, which are available to use in JAGS, but were not implemented in this work.

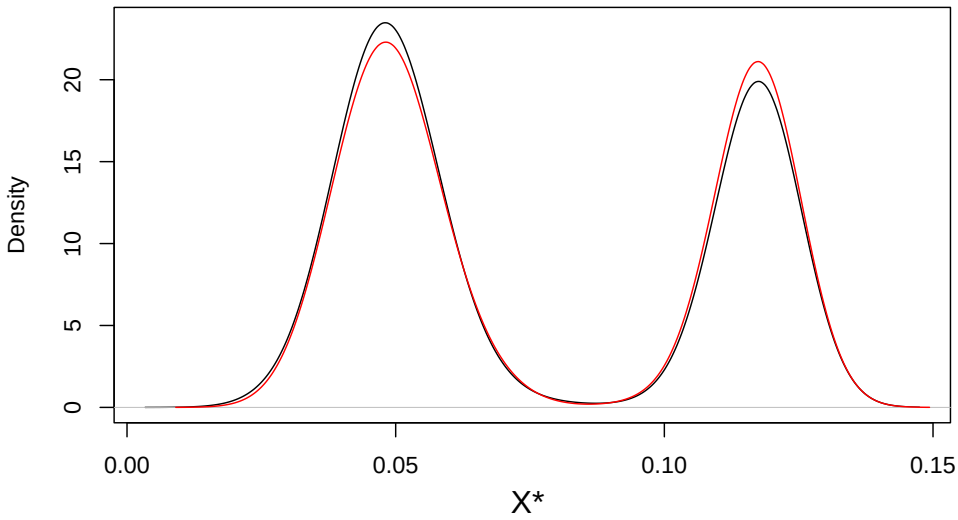

Figure E.3: The posterior density of X^*_{461} (i.e., the posterior density of X^* given joint posterior sample 461 from the randomly selected joint posterior samples from the forward model). The posterior density from the first parallel chain of this model corresponds to the solid black line, and the posterior density from the second parallel chain corresponds to the solid red line.

Appendix F

Model code

The content of this appendix demonstrates some examples of code that produce some of the models examined in Chapters 4 and 5.

Listing F.1: Simple linear EIV BR forward model

```
model{

  tau_delta ~ dgamma(a_delta, b_delta)

  tau_eta ~ dgamma(a_eta, b_eta)

  tau_epsilon ~ dgamma(a_epsilon, b_epsilon)

  tau_Xt ~ dgamma(a_Xt, b_Xt)

  for(i in 1:n_m){

    mu[i] <- inprod(Xt[i,1:2], beta[1:2])

    Yt_resid[i] <- Yt[i] - mu[i]

    Yt[i] ~ dnorm(mu[i], tau_epsilon)

    loglik[i] <- logdensity.norm(Yt[i], mu[i], tau_epsilon)

    Xt[i, 1] <- 1
```

```
    Xt[i, 2] ~ dnorm(mu_X, tau_Xt)
    for(j in 1:n_r){
      Y[i, j] ~ dnorm(Yt[i], tau_eta)
    }
    for(k in 1:n_c){
      X[i, k] ~ dnorm(Xt[i, 2], tau_delta)
    }
  }
    for(d_X in 1:2){
      beta[d_X] ~ dnorm(mu_beta[d_X], tau_beta[d_X])
      tau_beta[d_X] ~ dgamma(a_tau_beta[d_X], b_tau_beta[d_X])
    }
}
```

Listing F.2: Single-input single-output reparameterised EIV GP forward model

```
model{
  l <- k_l * sigma_epsilon
  k_l ~ dunif(k_l_U[1], k_l_U[2])
  sigma_epsilon ~ dgamma(sigma_epsilon_shape,
    sigma_epsilon_shape/sigma_epsilon_mean)
  sigma_k <- k_sigma_k * sigma_epsilon
```

```
k_sigma_k ~ dunif(k_sigma_k_U[1], k_sigma_k_U[2])

alpha ~ dnorm(0, 0.01)

tau_delta ~ dgamma(a_delta, b_delta)

tau_Xt ~ dgamma(a_Xt, b_Xt)

tau_eta ~ dgamma(a_eta, b_eta)

for(i_1 in 1:n_m){

  mu[i_1] <- alpha

  S[i_1, i_1] <- pow(sigma_epsilon,2) + pow(sigma_k,2)

  for(i_2 in (i_1+1):n_m){

    S[i_1, i_2] <- pow(sigma_k,2) *
      exp( -0.5 * pow( ( Xt[i_1] - Xt[i_2] ) / 1 , 2 ) )

    S[i_2, i_1] <- S[i_1, i_2]

  }

}

S.inv <- inverse(S)

Yt ~ dmnorm(mu, S.inv)

for(i in 1:n_m){

  Xt[i] ~ dnorm(mu_X, tau_Xt)

  for(k in 1:n_c){

    X[i,k] ~ dnorm(Xt[i], tau_delta)
```

```
    }
    for(j in 1:n_r){
      Y[i,j] ~ dnorm(Yt[i], tau_eta)
    }
  }
}
```

Listing F.3: Simple linear EIV BR backward model

```
model{
  X_star ~ dunif(-rho_X + centre_X, rho_X + centre_X)
  mu_star <- inprod(c(1, X_star), beta) # beta fixed
  Y_star ~ dnorm(mu_star, tau_epsilon) # tau_epsilon fixed
}
```

Listing F.4: Single-input single-output reparameterised EIV GP backward model

```
model{
  # Xt, Yt, sigma_epsilon, sigma_k, l, alpha fixed
  Y_star ~ dnorm(m_star, pow(S_star, -1))
  m_star <- m_X_star + t(k_X_star_X_train) %*%
    inverse(S_X_train) %*% (Yt - m_X_train)
  S_star <- S_X_star - t(k_X_star_X_train) %*%
    inverse(S_X_train) %*% k_X_star_X_train
```

```
  m_X_star <- alpha

  S_X_star <- pow(sigma_epsilon , 2) + pow(sigma_k ,2)

  for(i_1 in 1:n_m){

    m_X_train[i_1] <- alpha

    k_X_star_X_train[i_1] <- pow(sigma_k ,2) *
        exp( -0.5 * pow( ( X_star - Xt[i_1] ) / l , 2 ) )

    S_X_train[i_1 ,i_1] <- pow(sigma_epsilon , 2) + pow(sigma_k ,2)

   for(i_2 in (i_1 +1):n_m){

       S_X_train[i_1 ,i_2] <- pow(sigma_k ,2) *
        exp( -0.5 * pow( ( Xt[i_1] - Xt[i_2] ) / l , 2 ) )

       S_X_train[i_2 ,i_1] <- S_X_train[i_1 ,i_2]

     }

   }

   X_star ~ dunif(centre_X - rho_X, centre_X + rho_X)

}
```

Bibliography

Alvarez, M. A., Rosasco, L. & Lawrence, N. D. (2011), 'Kernels for vector-valued functions: A review', *arXiv preprint arXiv:1106.6251* .

Amin, E., Verrelst, J., Rivera-Caicedo, J. P., Pipia, L., Ruiz-Verdú, A. & Moreno, J. (2021), 'Prototyping sentinel-2 green lai and brown lai products for cropland monitoring', *Remote Sensing of Environment* **255**, 112168.

Batabyal, A., Sagar, S., Zhang, J., Dube, T., Yang, X. & Zhang, J. (2022), 'Gaussian process-based model to optimize additively manufactured powder microstructures from phase field modeling', *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg* **8**(1).

Baturynska, I. (2018), 'Statistical analysis of dimensional accuracy in additive manufacturing considering stl model properties', *The International Journal of Advanced Manufacturing Technology* **97**(5), 2835–2849.

Betancourt, M. (2017), 'A conceptual introduction to hamiltonian monte carlo', *arXiv preprint arXiv:1701.02434* .

Bilionis, I. & Zabaras, N. (2012), 'Multi-output local gaussian process regression: Applications to uncertainty quantification', *Journal of Computational Physics* **231**(17), 5718–5746.

Bilionis, I., Zabaras, N., Konomi, B. A. & Lin, G. (2013), 'Multi-output separable gaussian process: Towards an efficient, fully bayesian paradigm for uncertainty quantification', *Journal of Computational Physics* **241**, 212–239.

Bois, F. Y. & Maszle, D. R. (1997), 'Mcsim: a monte carlo simulation program', *Journal of Statistical Software* **2**, 1–60.

Brooks, S. P. & Gelman, A. (1998), 'General methods for monitoring convergence of iterative simulations', *Journal of computational and graphical statistics* **7**(4), 434–455.

Brown, P. (1982), 'Multivariate calibration', *Journal of the Royal Statistical Society: Series B (Methodological)* **44**(3), 287–308.

Browne, W. J., Draper, D. et al. (2006), 'A comparison of bayesian and likelihood-based methods for fitting multilevel models', *Bayesian analysis* **1**(3), 473–514.

Cervone, D. & Pillai, N. S. (2015), 'Gaussian process regression with location errors', *arXiv preprint arXiv:1506.08256* .

Chen, R., Reutzel, E. W., Khanzadeh, M. & Imani, F. (2022), 'Heterogeneous quality characterization and modeling of thin wall structure in additive manufacturing', *Additive Manufacturing Letters* **3**, 100042.

Conti, S. & O'Hagan, A. (2010), 'Bayesian emulation of complex multi-output and dynamic computer models', *Journal of statistical planning and inference* **140**(3), 640–651.

Cressie, N. & Kornak, J. (2003), 'Spatial statistics in the presence of location error with an application to remote sensing of the environment', *Statistical science* pp. 436–456.

Czelusniak, T. & Amorim, F. L. (2020), 'Selective laser sintering of carbon fiber–reinforced pa12: Gaussian process modeling and stochastic optimization of process variables', *The International Journal of Advanced Manufacturing Technology* **110**(7), 2049–2066.

Daumé III, H. (2008), 'hbc: Hierarchical bayes compiler', *Pre-release version 0.7, URL http://www. cs. utah. edu/˜ hal/HBC* .

de Castro, M., Bolfarine, H. & Galea, M. (2013), 'Bayesian inference in measurement error models for replicated data', *Environmetrics* **24**(1), 22–30.

Dellaportas, P. & Stephens, D. A. (1995), 'Bayesian analysis of errors-in-variables regression models', *Biometrics* **51**(3), 1085–1095.
URL: *http://www.jstor.org/stable/2533007*

Denham, R. J., Falk, M. G. & Mengersen, K. L. (2011), 'The bayesian conditional independence model for measurement error: applications in ecology', *Environmental and Ecological Statistics* **18**(2), 239–255.

Depaoli, S., Clifton, J. P. & Cobb, P. R. (2016), 'Just another gibbs sampler (jags) flexible software for mcmc implementation', *Journal of Educational and Behavioral Statistics* **41**(6), 628–649.

Epifani, I., MacEachern, S. N. & Peruggia, M. (2008), 'Case-deletion importance sampling estimators: Central limit theorems and related results', *Electronic Journal of Statistics* **2**, 774–806.

Fang, X., Li, B., Alkhatib, H., Zeng, W. & Yao, Y. (2017), 'Bayesian inference for the Errors-In-Variables model', *STUDIA GEOPHYSICA ET GEODAETICA* **61**(1), 35–52.

Flaxman, S., Gelman, A., Neill, D., Smola, A., Vehtari, A. & Wilson, A. G. (2015), 'Fast hierarchical gaussian processes', *Manuscript in preparation* .

Frazier, P. I. & Wang, J. (2016), Bayesian optimization for materials design, *in* 'Information science for materials discovery and design', Springer, pp. 45–75.

Freeman Technology (2020), 'Powder flow testing with the ft4 powder rheometer', `https://www.freemantech.co.uk/powder-testing/ft4-powder-rheometer-powder-flow-tester`. [Online; accessed 12-January-2020].

Frigola, R., Lindsten, F., Schön, T. B. & Rasmussen, C. E. (2013), 'Bayesian inference and learning in gaussian process state-space models with particle mcmc', *arXiv preprint arXiv:1306.2861* .

Fuller, W. A. (1987), *Measurement Error Models*, John Wiley & Sons, Inc., New York.

Garthwaite, P. H., Kadane, J. B. & O'Hagan, A. (2005), 'Statistical methods for eliciting probability distributions', *Journal of the American statistical Association* **100**(470), 680–701.

Gelfand, A. E., Dey, D. K. & Chang, H. (1992), Model determination using predictive distributions with implementation via sampling-based methods, Technical report, STANFORD UNIV CA DEPT OF STATISTICS.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian Data Analysis, Third Edition*, CRC Press, Taylor & Francis Group, Florida.

Gelman, A., Lee, D. & Guo, J. (2015), 'Stan: A probabilistic programming language for bayesian inference and optimization', *Journal of Educational and Behavioral Statistics* **40**(5), 530–543.

Gelman, A. & Rubin, D. B. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical science* **7**(4), 457–472.

Gelman, A. et al. (2006), 'Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper)', *Bayesian analysis* **1**(3), 515–534.

Goodridge, R., Tuck, C. & Hague, R. (2012), 'Laser sintering of polyamides and other polymers', *Progress in Materials science* **57**(2), 229–267.

Ha, S., Han, H., Kwon, D., Kim, N., Kim, H., Hwang, C., Shin, H. & Park, K. (2015), Systematic dimensional calibration process for 3d printed parts in selective laser sintering (sls), *in* 'International Design Engineering Technical Conferences and Computers and Information in Engineering Conference', Vol. 57045, American Society of Mechanical Engineers, p. V01AT02A036.

Ha, S., Ransikarbum, K., Han, H., Kwon, D., Kim, H. & Kim, N. (2018), 'A dimensional compensation algorithm for vertical bending deformation of 3d printed parts in selective laser sintering', *Rapid Prototyping Journal* .

Hecht, M., Weirich, S. & Zitzmann, S. (2021), 'Comparing the mcmc efficiency of jags and stan for the multi-level intercept-only model in the covariance-and mean-based and classic parametrization', *Psych* **3**(4), 751–779.

Hoadley, B. (1970), 'A bayesian look at inverse linear regression', *Journal of the American Statistical Association* **65**(329), 356–369.

Hopkinson, N., Hague, R. & Dickens, P. (2006), *Rapid manufacturing: an industrial revolution for the digital age*, John Wiley & Sons.

Hunter, W. G. & Lamboy, W. F. (1981*a*), '[a bayesian analysis of the linear calibration problem]: Response', *Technometrics* **23**(4), 344–350.

Hunter, W. G. & Lamboy, W. F. (1981*b*), 'A bayesian analysis of the linear calibration problem', *Technometrics* **23**(4), 323–328.

Ionides, E. L. (2008), 'Truncated importance sampling', *Journal of Computational and Graphical Statistics* **17**(2), 295–311.

Jain, P. K., Pandey, P. M. & Rao, P. (2008), 'Experimental investigations for improving part strength in selective laser sintering', *Virtual and Physical Prototyping* **3**(3), 177–188.

Jin, S., Iquebal, A., Bukkapatnam, S., Gaynor, A. & Ding, Y. (2020), 'A gaussian process model-guided surface polishing process in additive manufacturing', *Journal of Manufacturing Science and Engineering* **142**(1), 011003.

Kaintura, A., Spina, D., Couckuyt, I., Knockaert, L., Bogaerts, W. & Dhaene, T. (2017), 'A kriging and stochastic collocation ensemble for uncertainty quantification in engineering applications', *Engineering with Computers* **33**(4), 935–949.

Kelly, B. C. (2007), 'Some aspects of measurement error in linear regression of astronomical data', *The Astrophysical Journal* **665**(2), 1489.

Kennedy, M. C. & O'Hagan, A. (2001), 'Bayesian calibration of computer models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(3), 425–464.

Lalchand, V. & Rasmussen, C. E. (2020), Approximate inference for fully bayesian gaussian process regression, *in* 'Symposium on Advances in Approximate Bayesian Inference', PMLR, pp. 1–12.

Lappo, K., Jackson, B., Wood, K., Bourell, D. & Beaman, J. (2003), Discrete multiple material selective laser sintering (m2sls): experimental study of part processing, *in* 'Proceedings of the Solid Freeform Fabrication Symposium', Vol. 109, The University of Texas, p. 119.

Li, J., Jin, R. & Hang, Z. Y. (2018), 'Integration of physically-based and data-driven approaches for thermal field prediction in additive manufacturing', *Materials & Design* **139**, 473–485.

Liu, H., Cai, J. & Ong, Y.-S. (2018), 'Remarks on multi-output gaussian process regression', *Knowledge-Based Systems* **144**, 102–121.

Liu, J., Ye, J., Momin, F., Zhang, X. & Li, A. (2022), 'Nonparametric bayesian framework for material and process optimization with nanocomposite fused filament fabrication', *Additive Manufacturing* **54**, 102765.

Lu, C., Jia, X., Lee, J. & Shi, J. (2022), 'Knowledge transfer using bayesian learning for predicting the process-property relationship of inconel alloys

obtained by laser powder bed fusion', *Virtual and Physical Prototyping* pp. 1–19.

Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009), 'The bugs project: Evolution, critique and future directions', *Statistics in medicine* **28**(25), 3049–3067.

Maculotti, G., Genta, G., Quagliotti, D., Galetto, M. & Hansen, H. N. (2022), 'Gaussian process regression-based detection and correction of disturbances in surface topography measurements', *Quality and Reliability Engineering International* **38**(3), 1501–1518.

Majewski, C., Zarringhalam, H. & Hopkinson, N. (2008), 'Effect of the degree of particle melt on mechanical properties in selective laser-sintered nylon-12 parts', *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* **222**(9), 1055–1064.

McHutchon, A. & Rasmussen, C. (2011), 'Gaussian process training with input noise', *Advances in Neural Information Processing Systems* **24**.

Monnahan, C. C., Thorson, J. T. & Branch, T. A. (2017), 'Faster estimation of bayesian models in ecology using hamiltonian monte carlo', *Methods in Ecology and Evolution* **8**(3), 339–348.

Muff, S., Riebler, A., Held, L., Rue, H. & Saner, P. (2015), 'Bayesian analysis of measurement error models using integrated nested laplace approximations', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**(2), 231–252.

Mukesh Kumar, P. & Kavitha, R. (2021), 'Prediction of nanofluid viscosity using multilayer perceptron and gaussian process regression', *Journal of Thermal Analysis and Calorimetry* **144**(4), 1151–1160.

Neal, R. M. (2011), MCMC Using Hamiltonian Dynamics, *in* Brooks, S and Gelman, A and Jones, GL and Meng, XL, ed., 'HANDBOOK OF MARKOV CHAIN MONTE CARLO', Chapman & Hall-CRC Handbooks of Modern Statistical Methods, CRC PRESS-TAYLOR & FRANCIS GROUP, 6000 BROKEN SOUND PARKWAY NW, STE 300, BOCA RATON, FL 33487-2742 USA, pp. 113–162.

O'Hagan, A. & Oakley, J. E. (2004), 'Probability is perfect, but we can't elicit it perfectly', *Reliability Engineering & System Safety* **85**(1-3), 239–248.

O'Hagan, T. (2005), 'Elicitation', *Significance* **2**(2), 84–86.

Oyebamiji, O. K., Wilkinson, D. J., Jayathilake, P. G., Curtis, T. P., Rushton, S., Li, B. & Gupta, P. (2017), 'Gaussian process emulation of an individual-based model simulation of microbial communities', *Journal of Computational Science* **22**, 69–84.

O'Hagan, A. (2019), 'Expert knowledge elicitation: subjective but scientific', *The American Statistician* **73**(sup1), 69–81.

Peruggia, M. (1997), 'On the variability of case-deletion importance sampling weights in the bayesian linear model', *Journal of the American Statistical Association* **92**(437), 199–207.

Plummer, M., Stukalov, A., Denwood, M. & Plummer, M. M. (2016), 'Package 'rjags", *Vienna, Austria* .

Plummer, M. et al. (2003), Jags: A program for analysis of bayesian graphical models using gibbs sampling, *in* 'Proceedings of the 3rd international workshop on distributed statistical computing', Vol. 124, Vienna, Austria., pp. 1–10.

Raghunath, N. & Pandey, P. M. (2007), 'Improving accuracy through shrinkage modelling by using Taguchi method in selective laser sintering', *INTERNATIONAL JOURNAL OF MACHINE TOOLS & MANUFACTURE* **47**(6), 985–995.

Rosa, J. L., Robin, A., Silva, M., Baldan, C. A. & Peres, M. P. (2009), 'Electrodeposition of copper on titanium wires: Taguchi experimental design approach', *Journal of materials processing technology* **209**(3), 1181–1188.

Rozliman, N. A., Ibrahim, A. I. N. & Yunus, R. M. (2017), Bayesian Approach to Errors-in-Variables in Regression Models, *in* AbuBakar, SA and Yunus, RM and Mohamed, I, ed., '3RD ISM INTERNATIONAL STATISTICAL CONFERENCE 2016 (ISM III): BRINGING PROFESSIONALISM AND PRESTIGE IN STATISTICS', Vol. 1842 of *AIP Conference Proceedings*, Malaysia Inst Stat. 3rd ISM International Statistical Conference (ISM III), Univ Malaya, Inst Math Sci, Kuala Lumpur, MALAYSIA, AUG 09-11, 2016.

Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. (2016), 'Probabilistic programming in python using pymc3', *PeerJ Computer Science* **2**, e55.

Sharpe, C., Seepersad, C. C., Watts, S. & Tortorelli, D. (2018), Design of mechanical metamaterials via constrained bayesian optimization, *in* 'International Design Engineering Technical Conferences and Computers and Information in Engineering Conference', Vol. 51753, American Society of Mechanical Engineers, p. V02AT03A029.

Song, J.-H., Choi, K.-H., Dai, R., Choi, J.-O., Ahn, S.-H. & Wang, Y. (2018), 'Controlled kinetic monte carlo simulation of laser improved nano particle deposition process', *Powder Technology* **325**, 651–658.

Stavropoulos, Panagiotis & Foteinopoulos, Panagis (2018), 'Modelling of additive manufacturing processes: a review and classification', *Manufacturing Rev.* **5**, 2.
URL: *https://doi.org/10.1051/mfreview/2017014*

Tapia, G., Elwany, A. H. & Sang, H. (2016), 'Prediction of porosity in metal-based additive manufacturing using spatial gaussian process models', *Additive Manufacturing* **12**, 282–290.

Tapia, G., Khairallah, S., Matthews, M., King, W. E. & Elwany, A. (2018), 'Gaussian process-based surrogate modeling framework for process planning in laser powder-bed fusion additive manufacturing of 316l stainless steel', *The International Journal of Advanced Manufacturing Technology* **94**(9), 3591–3603.

Teh, Y. W., Seeger, M. & Jordan, M. I. (2005), Semiparametric latent factor models, *in* 'International Workshop on Artificial Intelligence and Statistics', PMLR, pp. 333–340.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T. & Gelman, A. (2023), 'loo: Efficient leave-one-out cross-validation and waic for bayesian models'. R package version 2.6.0.
URL: *https://mc-stan.org/loo/*

Vehtari, A., Gelman, A. & Gabry, J. (2017), 'Practical bayesian model evaluation using leave-one-out cross-validation and waic', *Statistics and computing* **27**(5), 1413–1432.

Vehtari, A., Mononen, T., Tolvanen, V., Sivula, T. & Winther, O. (2016), 'Bayesian leave-one-out cross-validation approximations for gaussian latent variable models', *The Journal of Machine Learning Research* **17**(1), 3581–3618.

Vehtari, A., Simpson, D., Gelman, A., Yao, Y. & Gabry, J. (2015), 'Pareto smoothed importance sampling', *arXiv preprint arXiv:1507.02646* .

Vetterli, M. (2019), Powder optimization for laser sintering: an insight in powder intrinsic and extrinsic properties, PhD thesis, ETH Zurich.

Vigneshwaran, K. & Venkateshwaran, N. (2019), 'Statistical analysis of mechanical properties of wood-pla composites prepared via additive manufacturing', *International Journal of Polymer Analysis and Characterization* .

Wegner, A., Harder, R., Witt, G. & Drummer, D. (2015), 'Determination of optimal processing conditions for the production of polyamide 11 parts using the laser sintering process', *International Journal of Recent Contributions from Engineering, Science & IT (iJES)* **3**(1), 5–12.

Wegner, A. & Witt, G. (2012), 'Correlation of process parameters and part properties in laser sintering using response surface modeling', *Physics Procedia* **39**, 480–490.

Williams, C. K. & Rasmussen, C. E. (2006), *Gaussian processes for machine learning*, Vol. 2, MIT press Cambridge, MA.

Wohlers, T., Campbell, I., Diegel, O., Kowen, J. & Huff, R. (2019), 'Wohlers report 2019. fort collins: Wohlers associates'.

Wong, K. V. & Hernandez, A. (2012), 'A review of additive manufacturing', *International scholarly research notices* **2012**.

Xie, R., Darvishzadeh, R., Skidmore, A. K., Heurich, M., Holzwarth, S., Gara, T. W. & Reusen, I. (2021), 'Mapping leaf area index in a mixed temperate forest using fenix airborne hyperspectral data and gaussian processes regression', *International Journal of Applied Earth Observation and Geoinformation* **95**, 102242.

Yang, Y., McGregor, D. J., Tawfick, S., King, W. P. & Shao, C. (2022), 'Hierarchical data models improve the accuracy of feature level predictions for additively manufactured parts', *Additive Manufacturing* **51**, 102621.

Ye, J., Mahmoudi, M., Karayagiz, K., Johnson, L., Seede, R., Karaman, I., Arroyave, R. & Elwany, A. (2022), 'Bayesian calibration of multiple coupled simulation models for metal additive manufacturing: A bayesian network approach', *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg* **8**(1).

Zhou, S., Pati, D., Wang, T., Yang, Y. & Carroll, R. J. (2019), 'Gaussian processes with errors in variables: Theory and computation', *arXiv preprint arXiv:1910.06235* .