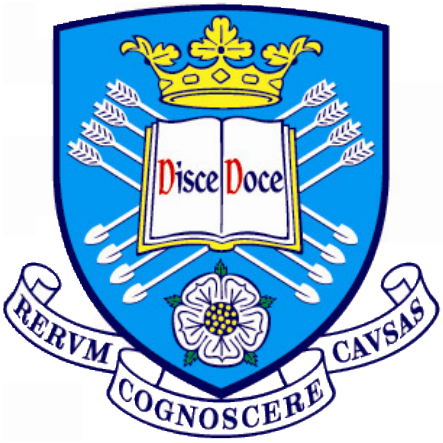
A Novel Approach to Fatigue: Grey-box Modelling for Probabilistic Damage Assessment



A Thesis submitted to the University of Sheffield
for the degree of Doctor of Philosophy in the Faculty of Engineering

by

S. J. Gibson

Department of Mechanical Engineering

University of Sheffield

October 2024

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Lizzy Cross, for her guidance throughout this work. I genuinely could not have wished for better supervision, mentorship, and friendship along the way. I am also deeply grateful to Dr. Tim Rogers for his insightful ideas, constructive feedback, and unwavering availability whenever I have sought his help.

The DRG has been a warm and welcoming environment to work in, and a special mention goes to Dan and Matt for their friendship and encouragement, particularly during the final stages of this work.

Lastly, a huge thanks to my family for their long-standing support and, above all to my partner, Maria. The decision to pursue a PhD involves significant sacrifice and you have had to bear this as much as me.

Publications

Journal Papers

Distributions of fatigue damage from data-driven strain prediction using Gaussian process regression. **SJ Gibson**, TJ Rogers, EJ Cross, Structural Health Monitoring

Improving aircraft performance using machine learning: A review. S Le Clainche, E Ferrer, **S Gibson**, E Cross, A Parente, R Vinuesa, Aerospace Science and Technology

A spectrum of physics-informed Gaussian processes for regression in engineering. EJ Cross, TJ Rogers, DJ Pitchforth, **SJ Gibson**, S Zhang, MR Jones

Journal Papers in Preparation

Physics informed machine learning for fatigue analysis: a kernel design approach. **SJ Gibson**, TJ Rogers, EJ Cross.

Book Chapters

Physics-Informed Machine Learning for Structural Health Monitoring, EJ Cross, **SJ Gibson**, MR Jones, DJ Pitchforth, S Zhang, TJ Rogers, Structural Health Monitoring Based on Data Science Techniques

Conference Papers

Data-driven strain prediction models and fatigue damage accumulation, **SJ Gibson**, TJ Rogers, EJ Cross, Proceedings of the 29th international conference on noise and vibration

Integrating Physical Knowledge into Gaussian Process Regression Models for Probabilistic Fatigue Assessment, **SJ Gibson**, TJ Rogers, EJ Cross, 2022 European Workshop on Structural Health Monitoring

Damage-informed cost functions for improving data-driven fatigue prediction. **SJ Gibson**, EJ Cross, 2024 European Workshop on Structural Health Monitoring.

Abstract

There is a growing interest in monitoring the loads that engineering structures withstand in order to better predict the fatigue damage accrual that has been accumulated. This is done with two goals: improving the safety of structures and enabling a greater useful life to be achieved. However, when structures operate in harsh environments, such as aircraft or offshore wind turbines, maintaining sensing networks for measuring stress and strain at crucial locations is difficult. As a result, virtual loads monitoring, or inferential sensing, in which machine learning methods are used to predict the stress at critical locations is becoming increasingly popular.

In the first part of this thesis, Gaussian process (GP) regression is used to develop a probabilistic approach for fatigue. The choice of a GP for a virtual sensor is not uncommon, however, developing a probabilistic view of fatigue by propagating the model uncertainty throughout the fatigue assessment procedure is novel. By doing this, a more robust assessment of the damage state of the structure is achieved. Furthermore, a discussion is facilitated around the causes and consequences of uncertainty in data-driven models with respect to fatigue assessment.

Treating fatigue analysis probabilistically is considered to be one way of reducing the conservatism that is common as a result of many uncertainties in the assessment procedure. In this work, it is the uncertainty from a loading perspective that is considered, but this thesis will also discuss how this could fit in with other existing probabilistic methods.

Following this, grey-box modelling - the introduction of our knowledge of physics into data-driven models – is considered. This is a new area of research currently attracting significant interest and different methods of inputting this physical knowledge into

the model are presented. By considering different loading scenarios (both dynamic and quasi-static), it is shown that domain-specific knowledge can both improve the accuracy, and also reduce the uncertainty, of model predictions, with the impact on probabilistic fatigue prediction being significant.

Contents

Chapter 1

Introduction

1.1 The challenge: Using data to predict fatigue damage

The use of data is revolutionising common practices across many industries. In this thesis, we will be looking at how we can use data to predict fatigue damage in engineering structures. Although it may seem contradictory, accurately predicting fatigue damage can both increase structural safety and enable a greater lifespan to be attained from the structure.

This work will also pose the question: how far can data take us in this challenge? And how can we use our knowledge of the physics of the process within a data-driven framework to improve this prediction?

1.2 Fatigue Failure

Fatigue is one of the most frequent causes of the failure of engineering structures [1–3], common across all engineering sectors. It can cause the failure of structures at stress levels that are far below their yield strengths and even small failures, such as of a bolt in a machine, will often be caused by fatigue. The majority of the time, this failure will not be of significant concern (we can replace the bolt) and, ideally,

redundant load paths can be designed into components to limit the consequences of fatigue failure [4]. Unfortunately, this is frequently impractical owing to space or weight restrictions. When this is the case and the potential cost of failure is large, we naturally want to avoid fatigue accrual at all costs.

In the 19th century, August Wohler was the first person to carry out significant research into the phenomenon of fatigue after realising that repeated applied stresses can cause materials to fail at relatively low-stress levels [5]. Since then, the topic has attracted a great deal of interest and substantial advancements have been made, with some of the most notable contributions coming from Palmgren (1924), Miner (1945, [6]), Paris (1961, [7] [8]), and Matshuisi & Endo (1968, [9]). Yet, despite considerable resources being invested in the subject of fatigue, there continue to be many engineering failures caused by fatigue.

This leads to an obvious question: When fatigue behaviour of most engineering materials is well understood (at least in laboratory conditions) [2], why is fatigue failure still prevalent? The most obvious answer to this is the fact that the failure can occur significantly below yield stress. Typical engineering overload failure analysis consists of predicting or measuring stress and comparing this to the ultimate tensile, or yield, strengths of the material. Generally, this would be with a single applied stress under consideration at any time. On the other hand, for a comprehensive fatigue assessment to be carried out, we need to have data on the full load spectrum that we think the structure is likely to see and information about stress at crucial locations on a structure [10, 11].

1.2.1 How does fatigue failure occur, and can we avoid it?

Even if we do have accurate information on the load spectrum of the structure or component, how easy is it to design against fatigue?

Fatigue cracks propagate from defects in materials that can exist on a microscopic level (such as voids, grain boundaries) or at a macroscopic level (such as at notches, scratches, surface roughness) [5, 12]. It is, therefore, very difficult to completely design against fatigue failure as avoiding such microscopic defects from which cracks form is not possible. Sometimes the improved fatigue resistance of materials under compressive stresses is harnessed by using methods such as shot peening to induce

compressive residual stresses at the material surface [13], where most cracks grow from. Yet, even with such precautions, it is generally not possible to completely avoid fatigue damage without over-designing components, adding significant bulk and weight.

1.3 Assessing fatigue damage

There are three main types of fatigue damage assessment: stress-life (S-N), strain-life(ϵ-N), and a linear elastic fracture mechanics (LEFM) approach [14, 15]. This work will focus on stress-life fatigue analysis, suitable for problems where there may be zero crack tolerance (such as aircraft landing gear that follows a safe-life design practice [16]) or for structures where regular physical inspections may be difficult. The strain-life method is suitable for low-cycle fatigue problems, in which larger stresses are seen and plastic deformation occurs as a result. The LEFM approach is used, when there are cracks known to be present in a structure, to assess the amount of time for a crack to reach a critical size.

1.3.1 Stress-life fatigue

Material stress-life, or S-N, curves are the fundamental building block of the S-N method. These are empirically found by testing the number of times a material specimen can withstand an applied stress at a specific amplitude in laboratory tests. These tests to develop S-N curves are performed at a constant repeated applied stress amplitude with zero mean stress. In reality, this is rarely the case for the applied loading on a structure. Instead, a cycle counting method is required, to convert a stress response $\sigma(t)$ into a set of stress ranges and corresponding mean values. The most widely used approach for this is the Rainflow cycle count [9] which is typically regarded as the most accurate [17, 18]. Figure 1.1 demonstrates how the Rainflow count will convert a stress-time history ('Load Reversals') into a set of stress ranges, mean values and number of cycles ('Rainflow Matrix Histogram').

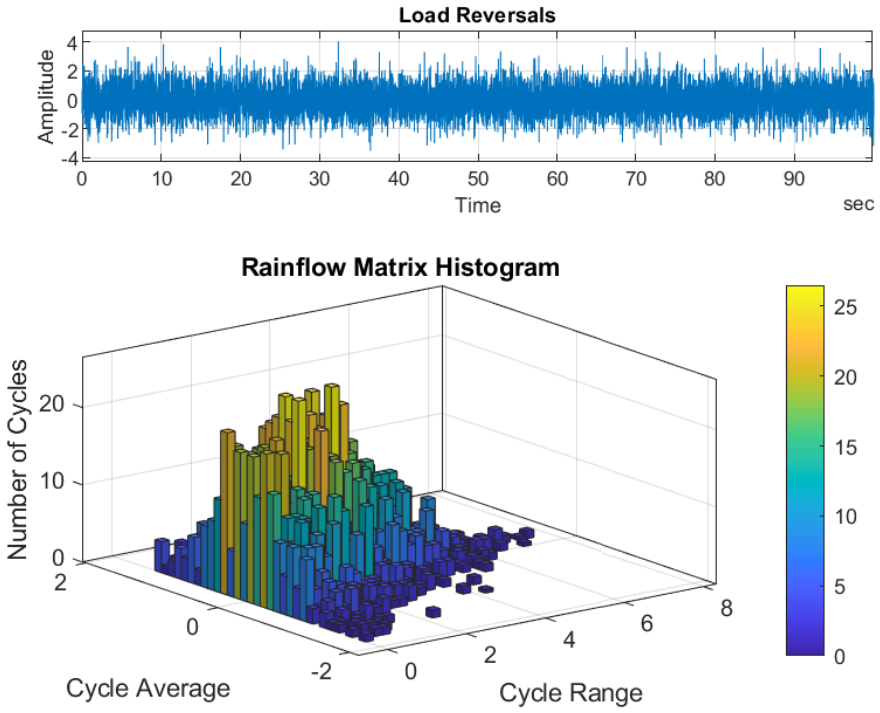


Figure 1.1: The rainflow cycle counting algorithm is used to take a variable-amplitude stress response and convert into a set of stress ranges/amplitudes and corresponding cycle mean values.

Using the obtained cycles and the material S-N curve (and a mean-stress correction), the Palmgren-Miner rule can be used to obtain the fatigue damage accrued. The Palmgren-Miner rule states that the total damage is the sum of the damage from the individual stress cycles.

Generally, when discussing 'damage' in the context of S-N fatigue, we are not referring necessarily to a detectable defect such as a crack or impact damage. Instead, we are generally considering the proportion of the expected life of the component that has been 'used'. The S-N methodology, including mean-stress correction, is discussed in more detail in Chapter 4, alongside some of its limitations. Suffice to say, it is the S-N methodology that is the underlying fatigue methodology used throughout this thesis.

1.4 SHM: A brief overview

To be able to predict fatigue damage from data, we need to know the loads that the structure is subjected to. To do this, technology developed from the field of structural health monitoring is used. Structural health monitoring (SHM) began to develop in the early 1980s, facilitated by reductions in the cost of computing power and increased availability of sensing equipment. SHM can broadly be defined as the implementation of an online damage detection strategy for engineering infrastructure [19]. Prior to the development of SHM, Non-Destructive Testing (NDT) would typically require local inspections to inspect structures for damage and often, a requirement for a structure to be taken out of service to do so. Thus, the continuous nature of SHM was the main driver of the development of the field, enabling the monitoring of structures without disrupting their usage. The fundamental aims of SHM are typically described by Rytter's hierarchy for damage detection [20]:

- Detection: Identifying if a structure is damaged
- Localisation: Identifying the location of damage on a structure
- Assessment: Assessing the extent of damage
- Prediction: Assessing the remaining life of the component

The popularity of SHM has developed alongside the widespread use of data across all engineering (and many non-engineering) disciplines, however, it is important to note that the collection of data does not, in itself, constitute SHM [21]. It is only by extracting useful features and information about the condition of the structure that we are performing actually SHM.

Despite the potential benefits, SHM as a domain has remained largely a research topic, struggling to break through into widespread industrial use. Two domains in which SHM has broken through into more common usage is within rotorcraft and rotating machinery (for the latter, the practice is generally referred to as *condition monitoring*). For the former, the constraints inherent to their operation provide a 'stable vibration from which changes in measured parameters can be attributed to component deterioration' [22]. In the case of the latter, the high rate of tool wear and controlled manufacturing environments make a cost-benefit analysis easier to carry out, in addition to the fact that there are no immediate significant costs (in terms of safety or economic terms) in the event of failure of the component.

1.4.1 Loads Monitoring

Much of SHM research focusses on the initial points of Rytter's hierarchy, with only a small subset focussing on prognosis. One of the largest unknowns in terms of prognosis is knowing what loads a structure has undergone in its lifetime and current operation. The first step, therefore, for researchers interested in fatigue damage prognosis is one of *loads monitoring* or prediction. Ideally, this knowledge would come from permanently installed strain gauges distributed across the structure. However, there are often limits to such monitoring campaigns. Firstly, it is unlikely that strain can be measured in all areas of interest for practical reasons (geometry, cable requirements, data storage). Alternatively, budgetary constraints may mean that sensing equipment is not able to be installed at the commissioning stage. High-value structures, such as offshore wind turbines or oil platforms, frequently operate in harsh environments making maintaining such sensor networks particularly difficult.

For example, aircraft are subjected to a wide spectrum of loads, operating in harsh environments [23] with phenomena such as high-velocity debris that can severely impact the integrity of components [24]. In the event of damage, sensors are often difficult or infeasible to replace. The result of this is that there is generally a knowledge gap of the real loads a structure has seen in service, which is of crucial importance when trying to extend the life of these structures. To overcome this, an assumed load spectrum will typically be used for predicting fatigue loads. However, one of the main motivations for this work is that these assumed loading spectra will frequently differ differ from the actual loads seen by the structure in operation [25, 26]. For example, the in-service life of wind turbines can frequently be increased between 25% and 50% of the structures original design life [27], indicating that the loading spectrum originally used was overly conservative.

As a result of this lack of knowledge, there is a growing area of interest within the SHM community in building models to predict stress and strain values at different locations on a structure using permanently installed sensors in accessible but remote locations. This is sometimes known as virtual loads monitoring where one attempts to use measurements such as acceleration (with more robust sensors available) to build regression models for predicting loads at critical locations. Sometimes it is possible to gain loading information from sensors that are installed for purposes other than measuring structural response, such as wind speed measurements [28, 29].

1.5 Machine learning for SHM

The key technology for the the data-driven regression models that are used for virtual loads monitoring, and many other tasks in SHM, is machine learning (ML). The use of machine learning has been key in the development of the field of SHM. There are two main categories of problems that ML is used to solve in SHM:

- Classification is the process of identifying and grouping data into different categories. For example, the most fundamental task of SHM is deciding if a structure is damaged based on the data that you receive from sensors. This is a classification problem.
- Regression is the process of identifying the relationship between variables. For example, if we know that the strain values at two locations on a structure are correlated, we can use regression to find the correlation. Ideally, we would also have information about how strong the correlation is via a probabilistic prediction.

The other major distinction to be made is the difference between *supervised* and *unsupervised* learning. In supervised learning, we provide the model with the information that we are hoping to predict. Conversely, in unsupervised learning, we are allowing the model to find interesting patterns in the data without providing labels *a priori*.

In many ways, structural health monitoring provides an ideal example of useful applications of machine learning in the real world. Increasing computational power and advances in sensing technology means that there is now an abundance of data available from measurement campaigns on structures and with SHM, the potential benefits in terms of safety, economics, and sustainability are significant. However, the widespread adoption of machine learning within engineering is not without challenges. One of the main challenges to the widespread use of ML 'in-the-field' is scepticism about 'new' technology to replace established methods. It is the authors' view that we can minimise this scepticism by focusing on the use of ML to supplement existing practices, rather than to replace them.

For example, we will see in some of this work that an ML-based approach can be used to gather information about a structure that could not have easily been achieved

with a physics-based modelling approach. It has been stated that machine learning theory offers a framework for steps 1-3 in Rytter's hierarchy, but necessarily not step 4 [30]. In this work, we will be focussing exclusively on the use of machine learning for Step 4: making predictions about the remaining usable life of the structure.

1.6 A probabilistic approach

This thesis will develop technology for using a probabilistic framework for the fatigue prediction problem. The motivation for this comes from both a machine learning and fatigue damage perspective.

1.6.1 A machine learning perspective

An engineering structure will withstand a wide range of loading phenomena over its lifetime, displaying 'long-tail' characteristics where a number of loading scenarios are common (e.g. take-off or landing of aircraft in benign conditions) but also with many possible phenomena with a low occurrence rate (e.g. performing a complex flight manoeuvre while withstanding wind gusts from the aircraft's port-side).

From a data-driven perspective, this makes it very difficult for a model to have 'seen' all of the possible loading situations in its training phase. This is partly a computational problem: training such a model would require very large amounts of training data and for complex structures this quickly becomes difficult to include all loading situations when training a model. However, there are also practical limitations to this problem: Throughout the structure's operational life, numerous phenomena might only happen a small number of times. As a result, obtaining real-world training data for these phenomena is also often difficult or impossible, especially in the case where monitoring is only possible for a fixed time.

The result of this is that a model will often be required to predict situations that it has not 'seen' in its training phase. Many of these will be similar to training scenarios and a reasonable prediction can be made, nonetheless. However, sometimes this will not be the case and, in such situations, it is desirable to know that the model is not confident in its prediction. For this reason, the authors advocate embracing methods that are able to assess confidence, i.e. those that are probabilistic.

The methods which are the focus of this thesis allow one to, in some way, account for epistemic uncertainties (such as those discussed above) and also aleatoric uncertainties arising from e.g. measurement noise (discussed further in Chapter 5).

1.6.2 A fatigue perspective: Uncertainties in the S-N approach

Machine learning and monitoring data aside, it is widely accepted that there are many sources of uncertainty in the S-N fatigue framework [16, 31, 32]. The most extensive overview of these uncertainties that the author has seen can be found in Hoole [33].

These include:

S-N Curves commonly shown and used are developed by averaging the results from many tests using a line of best fit. In reality, a large amount of scatter is present [34] in the number of cycles to failure because of inhomogeneities within materials. This scatter has led to the development of probabilistic S-N (P-S-N) curves [35]. Such curves give access to a confidence level of the material failing at a given stress range. The choice of 'which' S-N curve to use would typically depend on the use of safety factors elsewhere in the process, but may also depend on the availability of data from manufacturers, etc.

Load estimates: As discussed above, a lack of data [36], difficulty in maintaining sensors, complex loading, and modelling error can all contribute to uncertainty in the loads that the structure has 'seen'.

Uncertainties in the Miner/Palmgren rule also result in uncertainty in the resulting fatigue assessment. For example, it is known that the same load spectrum with stress cycles applied in different loading patterns will cause differing damage [37]. Furthermore, stress cycles below the endurance limit are not considered to cause fatigue damage, however, it has been shown that stress cycles below this value can still cause failure in structures [38].

Design uncertainties related to geometry and design of the component impact the fatigue life of a component but are difficult to predict. For example, welded joints present particular difficulty to predict due to the inhomogeneity of parent

and filler materials in addition to the not insignificant impact of the rapid heating and cooling nearby material [39].

The Rainflow cycle count is considered the most accurate cycle counting method [17, 18], but is known to be unreliable if the stress response under consideration is not sufficiently long [40].

Many of these sources of uncertainty have been acknowledged for a long time, yet, in the absence of a superior model (it can be noted that many 'correction' models of the Miner rule have been proposed over the years [41, 42]), the impact of this uncertainty is a reliance on large safety factors [16, 31].

Structures are, as a result, typically over-designed at the start of their design life and retired with a large remaining useful life. In recent years, there has been a trend towards treating fatigue assessment as a probabilistic problem to reduce the reliance on safety factors [16, 43–45][1]. It is useful to remember that safety factors are not only applied to design against fatigue failure but indeed across the entirety of structural design and analysis. Attempts within the fatigue community to *reduce* safety factors are not, therefore, attempts to reduce all forms of over-design (which would be impossible to achieve safely), but to simply reduce the wasted life of structures.

Treating fatigue as a probabilistic approach is not novel. Indeed, research has been carried out for decades and as early as 1971, there was an American Society for Testing and Materials (ASTM) symposium dedicated to the topic [47]. Early approaches to the problem included the development of probabilistic S-N (P-S-N) curves (mentioned above) to overcome the variability in the number of cycles to failure in laboratory tests. The problem of excessive safety factors was recognised as early as this symposium, with probabilistic safety factors being one of the approaches suggested [48].

Most researchers only attempt to account for one of the sources of uncertainty discussed above. Indeed, this thesis will focus solely on capturing the uncertainty from unknown loads in operation. It is however important to understand how generally these sources of uncertainty can and are treated in the whole. The most common method of doing so, for any kind of failure is the stress-strength interference method, also known as the *load-capacity interference* method. This method separates

[1] A safety factor can be defined as *the ratio between a measure of the maximal load not leading to failure and a corresponding measure of the applied load* [46]

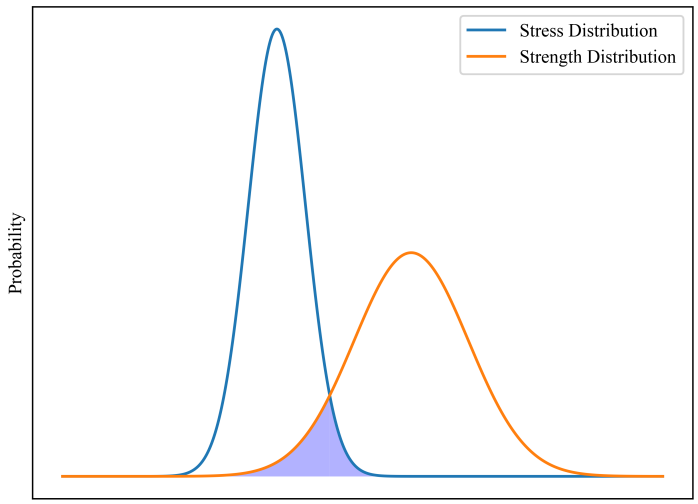


Figure 1.2: Stress-strength interference method: The overlap of the probability density functions represents a possibility of failure

the uncertainty into two distributions: that of the the the assumed loading that the structure has seen (the 'stress' distribution) and that of the ability of the material to withstand that level of loading (the 'strength' distribution). This is the simplest and most commonly applied way of assessing the probability of failure and is suitable for a wide range of situations [49].

The graphical representation shown in Figure 1.2 is common and demonstrates that there is a risk of failure if the probability density functions (PDF) of the two random variables overlap. The use of a safety factor effectively moves apart the two PDFs. Naturally, this approach can also be applied to any form of failure criterion (i.e. is not limited to fatigue analysis) and thus can fit in as part of a wider strategy more easily.

The most common method of developing a probabilistic distribution for the 'stress' component of the problem is to propagate the sources of uncertainty through a physics-based model. Typically, this is achieved using Monte Carlo Sampling (MCS) methods. Naturally, to do this we need to have a statistical understanding of the sources of uncertainty. Significant work has been conducted on developing these distributions [50]. Running physics-based models for complex structures is a computationally demanding task (assuming it is feasible for the problem at hand). As

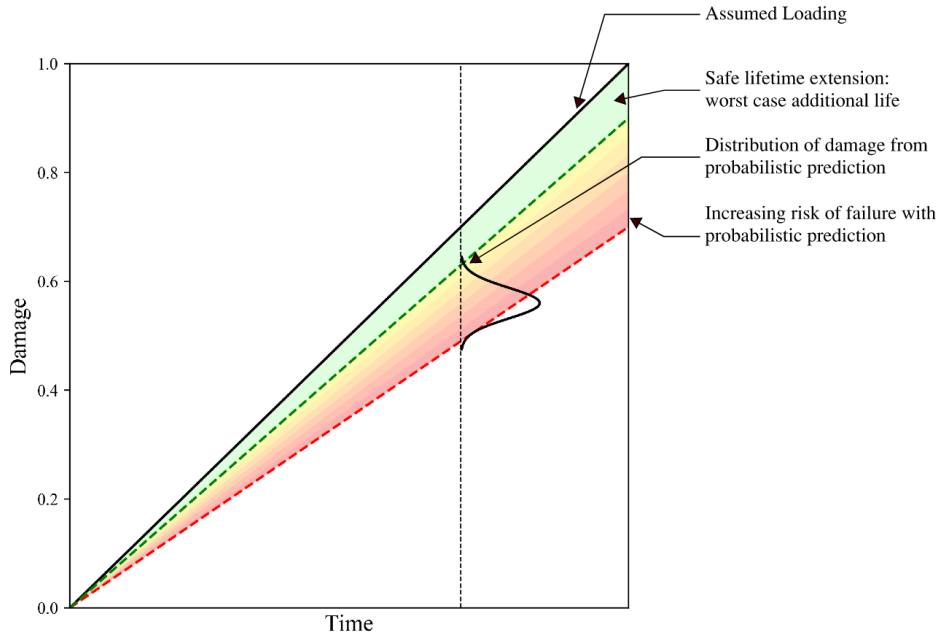


Figure 1.3: Treating fatigue as a probabilistic problem means that we would have a distribution of damage for a given point in time, rather than a single, deterministic, prediction

a result, there has been considerable effort to make this task more efficient, either by using data-driven surrogate models or by using more efficient sampling methods, such as Importance Sampling, to make the choice of model evaluations to be performed more selectively. These will be discussed in more detail in the next chapter.

The 'stress' distribution can be developed by combining the uncertainties related to the fatigue resistance of the component and material itself as well as inherent uncertainties in the S-N framework (such as those related to the Miner rule and Rainflow counting, etc.).

There is an (understandable) reluctance to treating failure as something that *could* happen as an engineer. Naturally, maintaining human safety is of paramount importance in this challenge and there is an aversion to 'accept' uncertainty in any sector of engineering where there is any threat to life. Yet, in some scenarios (principally, when the potential cost-of-failure is economic and not life-threatening), there is a pathway toward a risk-based cost-benefit analysis being used. For example, a recent NASA report states that '[estimating likelihoods of failure is essential to make] credible reliability and risk assessments' [51].

While intuitively simple, it is useful to ask: what is risk in this context? According to Hughes et. al [52], it is the combination of the cost/utility of a given consequence and the likelihood of that consequence occurring. This 'consequence' can take many forms and not all should be considered equal.

1.7 The challenges of adopting a probabilistic approach

There are many challenges to adopting such a probabilistic approach. On a pragmatic level, one of the main challenges lies in the availability of data for assigning distributions for probabilistic parameters in fatigue [50]. One of the challenges of adopting machine learning for engineering applications is the paradigm shift required to understand and interpret results: by using a model probabilistically, this change in mentality is only increased. It is for this reason that a not insignificant amount of this work will be dedicated to understanding the uncertainty of models. Furthermore, in later chapters, interpretability is (on top of improved results) a motivation for introducing physics into models.

There is also a reluctance to the use of data-driven modelling since many engineers are not trained to a high level in statistics or, often, the programming and linear algebra that are required for the implementation and interpretation of machine learning models. This thesis will attempt to break down some of these concepts accordingly and will discuss the challenge of interpreting the results of models.

Widespread adoption of probabilistic approaches within fatigue will, indeed, require a paradigm shift in many areas of engineering. While this is an uninviting prospect to many, it is important to remember that *'All models are wrong, but some are useful'*. The author believes, however, that it is better to have an understanding of *how* wrong the model is, rather than over-design until 'certainty' is achieved [2].

If one considers measured data as a gold standard for understanding a process, it is useful to remember that even this is simply a set of observations (which are biased) of a process. In the opinion of the author, having tools to access this uncertainty is

[2] it is important to highlight, once again, that this does not mean that human-safety is negotiable. Instead, it is a recognition that in a safety-factor driven, deterministic mindset, the failure of structures *does* still occur

desirable. Reluctance to embrace risk/probabilistic thinking is not insurmountable. It can be noted that even two of the most conservative engineering domains, the Aerospace and Nuclear sectors, have adopted forms of probabilistic risk assessment strategies [53–55]

While the extensive installation of sensing equipment is facilitating SHM, the sheer amount of data of a system does indeed represent one of the challenges. Even a rudimentary system may potentially have thousands of sensors installed, which results in huge quantities of data, even at relatively low sampling rates. This is especially true for the prediction of fatigue, for which it is necessary to collect data at particularly high sampling rates [56].

1.8 Contribution of this work

In this work, a novel probabilistic approach is presented for predicting the fatigue accumulation of structures. Firstly, by fully utilising the probabilistic nature of data-driven models (in this case, Gaussian process regression) for strain and stress prediction, a probabilistic view of fatigue damage is achieved. By embracing uncertainty in the modelling approach, it is the authors opinion that there is a pathway to both increase structural safety and also enable a greater lifespan to be attained from the structure. Throughout this thesis, the focus shall mainly be on aerospace applications. However, the methodology can straightforwardly be applied to other domains such as offshore, rotating machinery, and many other civil and mechanical applications. This is the first work to quantify the uncertainty on data-driven load prediction models and propagate through the fatigue assessment. In doing so, a discussion around how one can interpret the uncertainty of models is facilitated.

Using this framework, we will discuss how fundamental physical knowledge can be integrated into models. This is some of the first work that considers how physical knowledge can be used in a machine learning environment to predict fatigue damage. In this work, this is achieved by inserting knowledge of the dynamics of the problem into the structure of the model. In doing so, a clear improvement in the predictive capability is achieved, both from the perspective of strain prediction modelling and also probabilistic damage prediction. By utilising the novel probabilistic methodology discussed above, the effectiveness of the physics-informed machine learning modelling

approach is demonstrated, as is the value of the uncertainty propagation framework. Furthermore, by integrating physical knowledge in the structure of the model, physical meaning can be developed while working within a data-driven framework. This work will also discuss how this physical meaning can enable a greater trust in model prediction.

While the focus of this work is fatigue analysis, it can also be considered as a form of extended case study to demonstrate the advantages of uncertainty propagation for engineering applications. Furthermore, it can be considered a results-driven case study for the potential of integrating physics into data-driven models, focussing on the impact of model performance on structural damage, rather than overall performance as quantified by a generic error metric.

Chapter 2

In service prediction of fatigue damage accumulation: current approaches

In this chapter, an overview of the background theory and literature relevant to this thesis is presented. The overall themes that will be covered can be seen in Figure 2.1 on the following page. The focus here shall be on the in-service prediction of fatigue accrual and the strengths and weaknesses of physics-based and data-driven approaches to the problem. Clearly this is a large topic and, as a result, there are many areas of research that are outwith the scope of this work: Many of these are complimentary to this work, particularly the work focussing on probabilistic fatigue due to uncertainties in materials or, more generally, due to problems within the S-N process itself. The scope of the literature that will, or won't, be included in this review is shown in Figure 2.1.

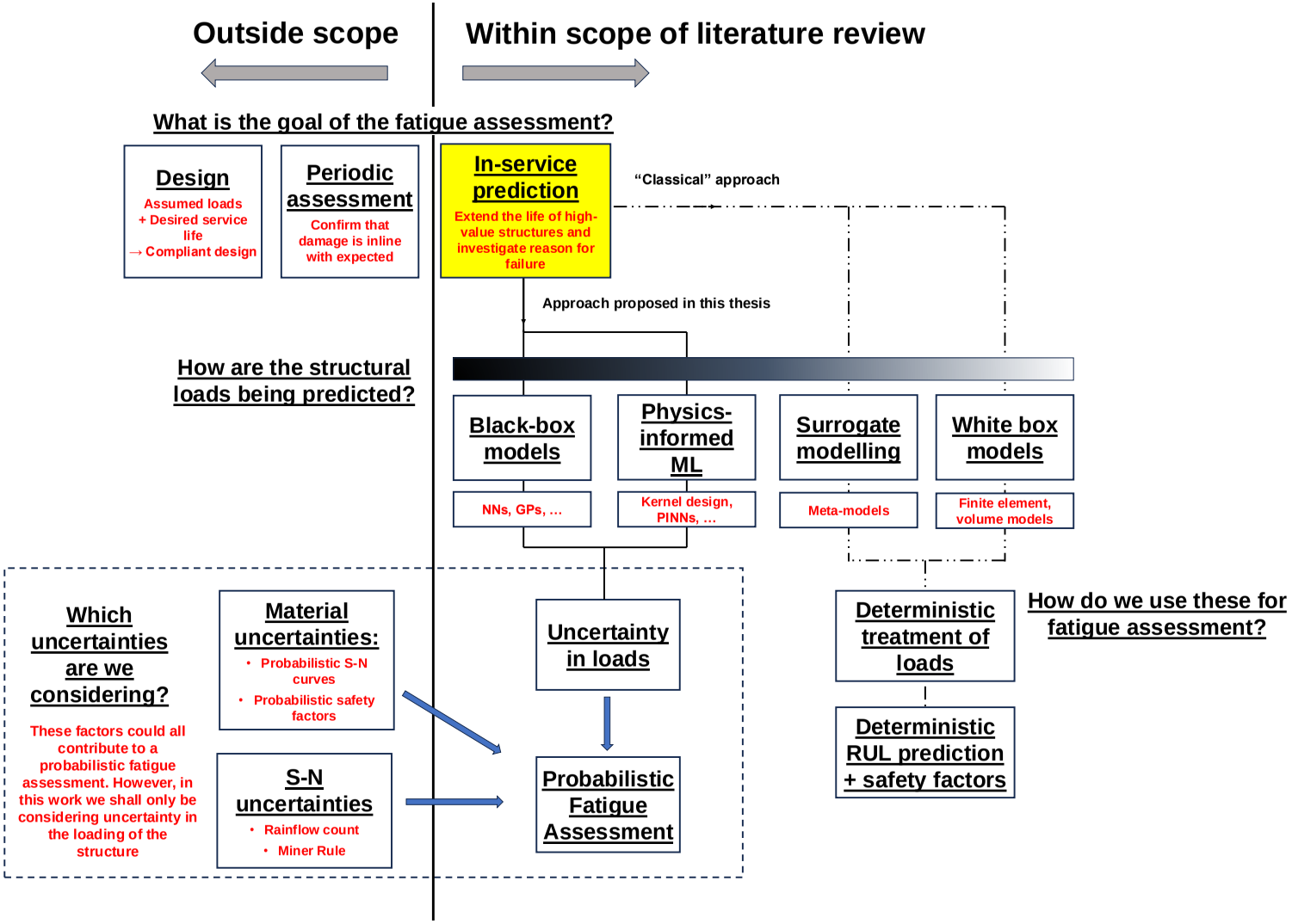


Figure 2.1: An overview of the scope of this literature review. In this work, the focus is on in-service prediction of damage using the safe-life fatigue methodology. A classical approach to this may be to use physics-based modelling to predict the loads and often, utilise a deterministic prediction with large safety factors. In this work, the focus is on data-driven modelling of structural loads in the time domain and utilising a probabilistic framework for safe-life fatigue.

2.1 Motivation: In-service fatigue damage prediction for safe-life applications

There are (broadly) three major reasons to be interested in fatigue analysis. Firstly, at the design stage, one would have an assumed load spectrum and perform fatigue analysis to ensure that the component is capable of withstand this loading. Later, during the operational life of the structure, one may look to perform inspections in order to ensure that the component is not experiencing more fatigue damage than expected. When the component has a crack tolerance, this may involve performing a crack-propagation assessment in order to ensure that it does not reach a critical size during it's service life and performing repair to assist with this.

In the third case, we have in-service prediction in the case where one cannot perform inspection or the avoidance of cracks is of paramount importance. There are two main reasons that this may be the case; firstly, the component may be located on a structure which cannot be taken out of service due to practical reasons (much of the field of SHM is concerned with mitigating this); Secondly, if the component is designed within the *safe-life* fatigue methodology, there *should* be no cracks present in the structure to detect (of course, this is not to say that performing periodic inspections, or SHM, on safe-life components is not recommended or worthwhile).

What is the motivation for predicting fatigue damage in this case? Firstly, human safety is of prime importance: in-service fatigue prediction in this case may be one final safety check. However, another motivation is economic: the potential to extend the lifetime of high-value structures. Structures such as wind turbines frequently only have a design life of around 20 years, but are frequently determined to have the potential to operate beyond this (often, between 25% and 50% [27, 57] of additional life can be achieved). Another motivation may be the detection of cracks, or even failure, in similar structures in a population: In this case, evidently either the structural model, or the load spectrum used at the design stage, is inaccurate. One factor that makes fatigue particularly difficult to model is the fact that the magnitude of error can propagate: a 10% error in loading estimate can result in fatigue damage error of 100% or more [58].

2.2 How to predict structural loads

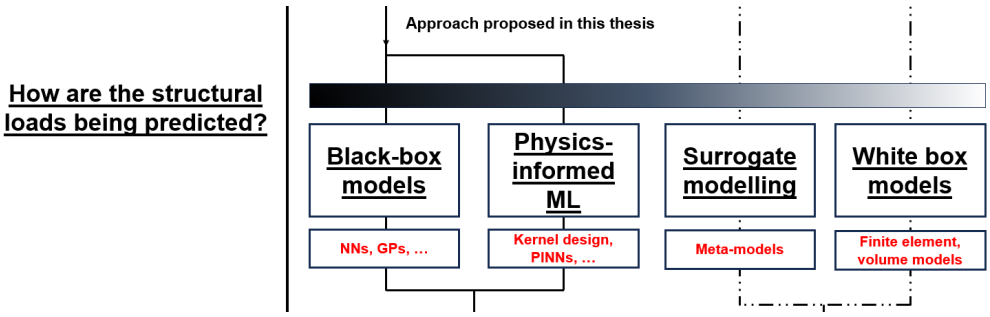


Figure 2.2: The methods of predicting loads considered in this literature review

2.2.1 Physics based modelling, a deterministic view

This raises the question: how do we predict these structural loads. We will first look at the literature from the right hand side in Figure 2.2, i.e. physics-based models becoming enhanced by the use of data. Physics-based modelling methods, such as finite element analysis (FEA), are the most widely used method for estimating structural loads [59, 60] and there are many examples of FEA being used for fatigue assessment in the literature [61, 62]. However, these are generally intended for use at the design stage, rather than to utilise monitoring data to predict in-service loads. Typically, an FE-based modelling approach for predicting fatigue would consist of evaluating the model across loads defined by loading spectrum assumed at the design stage. However, one reason we may wish to perform in-service loads prediction arises from the (not uncommon) situation scenario in which the loading spectrum assumed at the design stage is not representative of the loads the structure sees in service [63].

Sometimes, in-service monitoring data is available to refine this loading spectrum. One example of this could be using weigh-in-motion data to update the load spectrum for conducting fatigue assessment on road bridges [64, 65]. However, this (traffic loading on a bridge) is a case which lends itself to easy refinement of the loading spectrum by using low fidelity data sources such as camera feeds or weight sensors. Naturally, achieving this refined loading spectrum for more complex structures, or under more complex loading, would be more challenging.

Predicting the fatigue accumulation on wind turbines is an example of a more

complex environment this has been achieved with some success. Numerous papers have described how data can be used to update the loading spectrum and subsequently achieve greater accuracy of results while using a 'white-box' modelling approach. The Damage Equivalent Load (DEL) concept simplifies a variable amplitude loading history into a single equivalent (from a damage-perspective) load, avoiding the requirement to predict the full strain-time history, and is frequently used in this literature. Using this method, statistics on loading conditions that the structure is subjected to, such as mean wind speed, are used to update the loading spectrum of the structure. For example, Ziegler et al. [28] use this data in combination with aero-hydro-elastic simulations to update finite element models for extrapolation of fatigue damage prediction to locations across a monopole wind turbine where measured data is not available. Naturally, the predictive nature of this data is limited due to the use of statistics that are only available in 10-minute intervals. Even relatively major changes in either the operation of the turbine or the loading conditions may not be captured .

By refining the spectrum of loads a structure is subjected to, this goes some-way towards solving one of the major challenges with all modelling methods. Another challenge with an FEA approach, naturally, is related to the accuracy of the model itself, for which it may be difficult to validate across the loading spectrum. One way in which data can be used to assist in this process is via *updated finite element modelling* [66–68] in which model parameters are adapted to match the response seen in measured data. By doing this, a greater trust in the accuracy of the model can also be attained.

Perhaps the greatest criticism of the FE method for the purpose of fatigue analysis is that it requires solving of a very large number of simultaneous equations [69] and, therefore, has a very large computational cost of running model evaluations. While structural overload analysis may have a limited number of evaluations for different failure modes, fatigue analysis requires stress prediction over a wide loading spectrum that the structure is expected to see. Chapters 6 onwards in this thesis are investigating problems with dynamic loading, for which the computational cost of time-domain FE modelling is especially demanding [70].

As a result of the high computational cost of evaluating physics-based models, various methods have been developed to optimise the model evaluations needed using data-driven methods. Methods such as Importance Sampling [71] and Latin

Hypercube sampling [72] are used. While this does go some way towards reducing the computational demands of an FE approach, it comes at the cost of some accuracy.

2.3 Surrogate Modelling

The computational cost of performing repeated model evaluations using physics-based models was discussed above. As a result, various methods of selecting the model evaluations that can most effectively reduce this computational expense have been proposed. However, even for these efficient methods (such as Latin Hypercube Sampling or Importance Sampling), the computational expense of this approach can be large, especially for fatigue assessment which requires the evaluation of long time-series of data.

There is, therefore, an incentive to use *surrogate modelling*. Also known as meta-modelling, this is the use of a machine learning model to represent a computationally expensive model (typically, this is referring to a FE model, but this could apply to any expensive function or model). Using a data-driven learner, such as Gaussian processes [73, 74], to act as a 'surrogate' to the computationally expensive model can significantly reduce its computational cost.

From another perspective, further motivation for using a surrogate modelling approach comes from the potential to collect data for a machine learning model that would otherwise be difficult to obtain. While obtaining operational test data at the extreme ends of the loading spectrum for a pure data-driven modelling approach is challenging, this is not a problem with a surrogate model approach, as data should be readily available across the domain of loading scenarios (obviously, this requires an assumed loading spectrum that is representative of the actual loading on the structure, something that is not always the case, and rarely is in complex environments). While this does represent a major opportunity of surrogate modelling, there is an associated downside: As with all modelling methods, as complexity increases, our uncertainty in the results of the model also grows. One of the advantages of a surrogate modelling approach is that the 'long-tail' events, i.e. those that are unlikely to happen more than a few times over the structures lifetime, can be modelled and training data can be provided for the machine learning model (by contrast, this is very difficult in the purely-data-driven approaches presented next). However, the flip-side of this is that

the likelihood of having validated data in this region is also unlikely. Another of the downsides of surrogate modelling it that of stacking uncertainty: the addition of uncertainty using the data-driven learner as well as that of the physical model. This is true of all surrogate modelling but, especially problematic in this region, when uncertainty of the original model is also of question. Furthermore, it is in these extreme regions that nonlinear behaviours are most likely to be exhibited, which may not be captured by a black-box model (naturally, depending on the training data and sparsity of training points).

Compared to a purely-data-driven approach, surrogate modelling also presents the opportunity to acquire data in physical locations that would otherwise be difficult. For example, the presence of stress concentrations are important for fatigue design. Such points are very difficult to model using a data-driven learning (or extrapolation to such places would result in the uncertainty stacking that we've discussed above). However, by using a surrogate modelling approach, a data-driven solution could still be possible. For example [75] presents a Kriging-based surrogate model for S-N fatigue on the welds of a railway bogie beam including . A purely-data-driven approach would be impractical here as acquiring the stress profile in such detail for training the model would not be possible.

Again, the use of efficient computational methods could be combined with surrogate modelling to further reduce computational time. Successful examples in the literature include that of Echard et al. [76], who use a Kriging surrogate model combined with Importance Sampling to assess failure reliability probabilistically. In [77], Teixeira et al. focus on the *design of experiments*, reducing the number of calls that are required to be made to the physics-based model with a Kriging surrogate again being used. However, in the authors opinion, this has the potential to further exaggerate the uncertainty stacking phenomena discussed above.

In general, surrogate modelling is a practical approach to utilising data-driven methods to model structural loads and fatigue that is appropriate in many environments. Generally, the author would argue, this is most the case when the input loads are well known and the potential loading spectrum is relatively narrow. Using surrogate modelling has the potential to significantly improve computational times of a physics-based approach and, as a result, will appear attractive in many situations and industries that are otherwise reluctant to embrace a data-driven modelling approach. However, many of the other downsides of finite element modelling are still

present. In particular, in the authors opinion, it is the lack of validation that is most important here. Again, for good results, a good knowledge of the input loads into the structure should be known, which is one of the fundamental challenges of finite element modelling. Furthermore, the impact of adding uncertainty of the data-driven model to the uncertainty of the physics-based model should be considered.

2.4 Data-driven methods

We will now begin to consider the problem from the data-driven perspective, starting with a black-box modelling approach and subsequently considering how physics can be incorporated into such models (with a focus on Gaussian process regression, as this is the modelling selection made throughout this work).

There is a rapidly growing literature of examples of data-driven approaches being used for loads prediction. For example, artificial neural networks [78–81], Kalman filters [82, 83] and Gaussian process regression [84–86] have all been used for this purpose. Generally, fatigue is mentioned as a motivation for performing loads prediction, but these citations often do not continue the loads prediction through to the fatigue assessment. Those that do, generally limit this to a deterministic view, without any consideration of model uncertainty.

As with the physics-based approaches described above, the topic of fatigue in wind turbines is, again, a frequent topic of study, lending itself to a data-driven framework as well. Again, these generally utilise the damage equivalent load (DEL) concept. For example, Ziegler et al. [87] present a k-nearest neighbour approach for predicting the DEL on wind turbines as an alternative to the finite-element method proposed by the same author in [28] (cited above). In this case, greater accuracy is achieved with the data-driven approach than was achieved with the physics-based approach. Avendano-Valencia et al. [88] also use the DEL approach, using GP regression to provide an estimate of the fatigue accumulation on a turbine (discussed more in the probabilistic section below). The results of these approaches show the potential for predicting fatigue using a data-driven approach, but share the same fundamental downside described above for the DEL approach: by using mean statistics over a specified time interval and not predicting the full stress-time history, even relatively major changes in either the operation of the turbine or the loading conditions may

not be captured .

The aerospace sector has seen interest in loads and fatigue prediction from a data-driven approach. For example, the fatigue accumulation of landing gear is of interest [33]. One example of Gaussian process regression being used for virtual loads monitoring can be found in [89], where the loads on the landing gear of an aircraft are predicted with a good level of accuracy. In [90], Cross uses a Multilayer perceptron (MLP) for predicting the loads on the same landing gear data. However, in both cases, these stop short of using this for fatigue prediction. In Chapters 4 and 5, a study will be performed into the fatigue accumulation on a Tucano military aircraft. The same data has been used previously for studies using an artificial neural network (ANN) based parametric fatigue monitoring system [80, 81], and in Ref. [86], the loads on the aircraft wing are predicted and fatigue analysis is carried out using the predicted strain values. In both cases, strain is the target of the model and subsequent fatigue assessment is carried out. While a good level of accuracy is generally achieved in these citations, no thorough consideration is given to the uncertainty of the model (although in [81], an element of aleatoric uncertainty is included by indicating the difference in fatigue prediction that would arise from an error simulated on the strain gauges of 5%).

Another area in which there has been interest in data-driven fatigue damage prediction is within the health and usage monitoring systems (HUMS) within the rotorcraft domain. There are numerous ML-based approaches to loads monitoring and fatigue analysis: In Ref. [91], the mechanical loads of a helicopter are predicted using an MLP, although the results are not propagated through the fatigue assessment. In Ref. [92], an extreme learning machine (a subclass of feed-forward NN) is used for load prediction and the results are fed through the S-N fatigue analysis procedure to predict damage accumulation.

There are, therefore, many citations demonstrating the ability to predict structural loads using a data-driven approach. However, generally these are limited to this and do not consider the consequence of the prediction from a damage perspective. In the coming chapters, it will be demonstrated that prediction accuracy at crucial points in the time-history are more important from a fatigue damage perspective than the overall prediction quality (as quantified using a metric such as normalised mean squared error). Those approaches that do indeed conduct a fatigue assessment generally limit this to a deterministic approach, without giving any consideration

to the uncertainty of the model. Again, this is something that is challenged in the coming chapters.

2.5 Grey-box modelling

In a machine learning model, we generally have a known input and visible output, but don't often know how we got from one to another. This lack of knowledge about the mechanics of the model has resulted in machine learning models commonly being referred to as *black-box* models. A physics based model, by contrast, is grounded in rules of physics and often considered to have a more interpretable output as a result. In contrast to *black-box* models, such models are commonly referred to as *white-box* models. The concept of physics-informed machine learning was briefly introduced in the previous chapter and is often otherwise referred to as *grey-box* or *hyrid* modelling, sitting somewhere between the two and blending the use of data and physical knowledge in a model. As alluded to in the introduction, there can be various motivations for this modelling choice: achieving computational efficiency while improving accuracy, improving the interpret-ability of data-driven models [93] or improving trust in a data-driven model. As this is a new and emerging field, there is not a great deal of literature related to grey-box modelling for fatigue analysis, so this section will aim to provide a more general introduction into the field. A more thorough introduction to grey-box modelling for SHM applications can be found in Ref. [94]. Note, that while not discussed further here, a surrogate model trained using a physics-based model could be considered a form of grey-box model.

Perhaps the easiest, most common and intuitive way of inputting knowledge of physics into a machine learning model is via *semi-physical* modelling or *input augmentation*. This involves manipulating inputs prior to their use in a model to indicate physical processes [80, 86, 95]. This methodology is used in Chapters 4 and 5 to manipulate inputs to be representative of the forces acting on an aircraft. An example of this is squaring the measured airspeed of an aircraft before using it as an input to the model [80, 81] (in the case of these citations, a neural network is the type of model used) to give an indication of dynamic pressure acting on the aircraft. Similarly, In Ref. [95], Azzam develops a *mathematical network* for fatigue monitoring of helicopters, in which physics-inspired merging functions are used as the first layer of the network.

An alternative way of integrating physics into models is via *residual modelling* in which ML methods such as Gaussian process regression [96, 97] are used to predict the discrepancy between a physics-based or empirical prediction and the measured data. The potential benefit of this is demonstrated in [96] where a relatively simple empirical physically-representative model (Morrison's equation) is used as a mean function within a GP regression model for predicting wave height on an offshore structure. Although not performed in [96], the wave height naturally has a direct effect on fatigue loading [98]. By accurately predicting the wave height and forces, the loading spectrum used in any of the above modelling methods could be refined (and of particular use for physics-based modelling).

Neural networks are among the most commonly used forms of machine learner and, unsurprisingly, one of the most frequently used forms of grey-box model is *Physics informed Neural Networks* (PINNs) [99]. If the differential equations that govern a problem are known, this knowledge can be included as an extra term in the loss function that we use to train the data-driven model. While there aren't examples of PINNs used for loads monitoring and S-N fatigue in the literature, there are examples of their use for for LEFM fatigue assessment of additively manufactured alloys [100], developing new S-N models for notched metallic components [101], corrosion fatigue [102, 103] and developing physically consistent probabilistic S-N curves [104]).

The last form of grey-box model that will be considered here are kernel based methods. Gaussian process (GP) regression is the machine learning method used throughout this work, and much of the flexibility of a GP arises from the fact that we can choose an appropriate covariance function for the model, specifying the functional form that the model posterior will take. While typically, data-driven models will use an 'ready made' kernel, it is also possible to design kernels to suit our needs. A basic form of this is the periodic kernel: as periodic responses are known to exist in many areas, such as robotics [105], this fact can be harnessed to improve the response prediction significantly [105]. In [106], this principle is used to develop a kernel for single-degree-of-freedom (SDOF) oscillating systems under zero-mean Gaussian white noise forcing. This method is applied to the problem of fatigue assessment in the later chapters of this work.

2.6 Probabilistic Damage Modelling

In the first chapter, it was acknowledged that there are uncertainties in the safe-life fatigue methodology related (but not limited) to S-N curves, load estimates, uncertainties in the Palmgren-Miner rule, design uncertainties and the rainflow cycle count. Naturally, with such a large number of uncertainties widely acknowledged, there has been a relatively large amount of literature looking into these, starting as early as the 1970's [47]. However, the only one of these that this work is concerned with is the uncertainties within the loads prediction, with the other sources of uncertainty out of the scope of this work.

While in the data-driven loads monitoring section above, many examples are presented demonstrating the capability of data-driven methods to predict structural loads, generally these stop at this point: despite fatigue being frequently cited as a motivation for virtual sensing, the fatigue assessment is generally not performed using the predicted loads. For those that do, this is generally treated in a deterministic manner and little consideration is given to how the strain prediction impacts the fatigue assessment.

However, probabilistic methods are becoming increasingly popular. Gaussian process regression is one of the most popular ML methods for virtual sensing problems, in part, due to its inherently probabilistic nature and subsequent capacity to indicate uncertainty. Naturally, either a data-driven or physics-based modelling approach can lend itself to a probabilistic methodology. However, the typical steps for achieving in a physics-based context would be [107]:

1. Building a model for the structure that you want to model. Most commonly, this is a finite element model.
2. Generating random variables for the inputs to this model. This could be the input forces into the model, or varying modelling parameters.
3. Evaluating the model many times.
4. Statistically analysing the results, using the random inputs generated in step two.
5. Performing model convergence studies.

Earlier, the large computational cost of evaluating many physical models was discussed. This is especially true when extending such a physics-based approach to a probabilistic framework, as many more evaluations have to be performed for developing Monte Carlo simulation (or a similar approach).

There have, indeed, been numerous data-driven methods that allow for a probabilistic fatigue assessment. For example, the approach by Avendano-Valencia et al. [88], mentioned above, predicts damage equivalent loads using Gaussian process regression. By predicting a damage equivalent load, the need for taking a probabilistic view of the full strain-time history, as is done in this work, is avoided. The method proposed by Echard et al. [76], who use a Kriging surrogate model combined with Importance Sampling to assess failure reliability also used the probabilistic nature of Kriging models. Using the LEFM methodology, Nie et al. [108] utilise an ANN with Monte Carlo simulation to establish the fatigue reliability of steel bridges. Such approaches can fit into probabilistic risk-based decision frameworks for structural health monitoring [52], which can be used to make risk-informed decisions for safety-critical components. By adopting a probabilistic approach to fatigue, one has the potential to better control the risk level and plan maintenance more effectively [33, 44].

2.7 Summary

In this chapter, the scope of this work was defined while reviewing the existing approaches to the problem. The focus of this work is on data-driven virtual sensing, and in particular, propagating the uncertainty of the prediction through the fatigue assessment. While there is extensive literature on predicting fatigue, most notably using finite element analysis, the propagation of uncertainty using such a method is computationally costly. By adopting a data-driven approach, the computational cost of adopting a probabilistic methodology is significantly reduced. Furthermore, much of this work focusses on predicting damage in environments where understanding the spectrum of loads the structure is subjected to is difficult, a situation which again lends itself to a data-driven methodology. A brief introduction to grey-box modelling was also covered, which is the focus of the later chapters of this work. In the next chapter, more details will be given of the methodology developed in this work, and the datasets used for demonstration.

Chapter 3

An introduction to Gaussian process regression and the data used in this work

Throughout this work, Gaussian process regression is employed to predict stress and strain on structures before propagation through to the fatigue assessment. In this chapter, an overview of the requisite theory will be presented. Following this, the two datasets that will be used in this thesis to demonstrate the proposed methodology will be introduced.

3.1 Gaussian process regression

Gaussian process regression is a powerful machine learning technique, popular within the structural health monitoring field [109–111]. A frequent description of a GP is a *distribution over functions*, without a defined functional form, but rather a distribution over functions that are coherent with the data that the process is conditioned by.

A Gaussian process model is specified by the choice of a mean and a covariance function, $k(\mathbf{x}, \mathbf{x}')$, where $\mathbf{x}$ and $\mathbf{x}'$ are distinct inputs to the GP. Generally, a zero mean prior is assumed (although if we have prior knowledge of our process, incorporating this into the mean function is a relatively easy method of importing this knowledge into the model). The squared exponential is the most commonly used covariance

functions [112]:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2}||\mathbf{x} - \mathbf{x}'||^2\right) \tag{3.1}$$

The choice of covariance function results in a number of free-parameters. These parameters are referred to as *hyperparameters*. For the squared-exponential kernel used here, these are σ_f^2 and ℓ, the signal variance and lengthscale hyperparameters, respectively. It is these hyperparameter values that are adjusted during the model optimisation stage in order for the model target to fit the input data it sees (the role each hyperparemeter plays in the model is demonstrated shortly).

Using equation (3.1), we can build a matrix relating the similarity of the data points under consideration. This is called a covariance matrix $K(\cdot, \cdot)$. Predictions can now can be specified as a function of $K(X, X)$, the covariance matrix of the training input values, $K(X_*, X)$, the covariance matrix of the test (prediction) and training input values, $K(X_*, X_*)$, the covariance matrix of the test input values, and $\mathbf{y}$, the observed training output/target values.

During the training stage, the hyperparameters in equation (3.1) (or other covariance function of choice) are optimised with respect to a chosen metric. In this work, this is achieved by minimising the negative log marginal likelihood:

$$\log p\left(\mathbf{y}|X\right) = -\frac{1}{2}\mathbf{y}^\top \left(K + \sigma_n^2 I\right)^{-1}\mathbf{y} - \frac{1}{2}\log\left|K + \sigma_n^2 I\right| - \frac{n}{2}\log 2\pi \tag{3.2}$$

where I is the identity matrix and n is the number of training points used. The noise hyperparameter σ_n^2 represents the variance of noise ϵ. This noise hyperparameter is required as real world situations are unlikely to follow a function $y = f(x)$, but rather a noisy representation of this $y = f(x) + \epsilon$. Throughout this work, a particle swarm algorithm is used to search for optimal hyperparameters. After optimising the hyperparameters, the GP posterior mean, $\mathbb{E}\left[\mathbf{f}_*\right]$, and covariance, $\mathbb{V}\left[\mathbf{f}_*\right]$, functions can now be defined by equations (3.3) and (3.4), respectively [112]:

$$\mathbb{E}\left[\mathbf{f}_*\right] = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1}\mathbf{y} \tag{3.3}$$

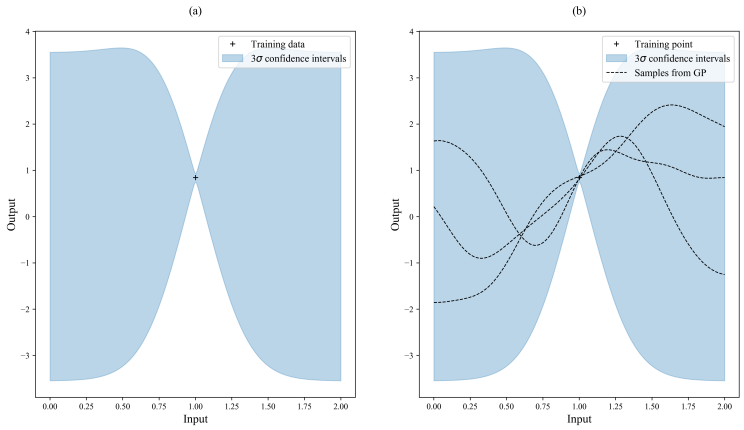


Figure 3.1: (a) 3σ confidence intervals of (generic) Gaussian posterior (b) functions drawn from the posterior distribution

$$\mathbb{V}[\mathbf{f}_*] = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \tag{3.4}$$

3.1.1 Sampling from the GP

A feature of GP regression is that the distribution of any finite number of points in the output space are jointly Gaussian. This results in an infinite number of functions that can fit, or be drawn from, the posterior distribution. Figure 3.1(a) illustrates an arbitrary GP posterior distribution that has been conditioned on a single data point at $x = 0$, while Figure 3.1(b) shows a series of functions that this non-parametric model can fit, which are generally referred to as 'draws' from the posterior distribution. Each function fits to the training/conditioning point: away from the training data the behaviour of the functions are characterised by the prior covariance function.

From a fatigue damage perspective, where one will typically be predicting strain before using this prediction in fatigue assessment, these draws are typically not utilised by the work in the literature (including in the references discussed in the previous chapter). In this work, samples are taken from the posterior strain distribution of the GP model propagated through the fatigue assessment. The novel concept (starting in Chapter 5) of this work arises from the notion that there is no specific $f(x)$ that characterises the data, but infinite $f(x_i)$ instead. By calling upon a large

number of these functions, a probabilistic prediction for fatigue can be built from the GP posterior distribution. Specifically, by generating a large number of draws from the posterior and using the stress-life fatigue methodology to assess damage, we will be able to generate a distribution for damage accumulation given a GP strain prediction.

Samples/draws are generated from the posterior by calculating ([112]):

$$\mathbf{f}_{sample} = \mathbb{E}[\mathbf{f}_*] + L\mathbf{u} \tag{3.5}$$

where $\mathbb{E}[\mathbf{f}_*]$ is the the GP mean prediction calculated using equation (3.3), $\mathbf{u}$ follows an independent Gaussian distribution, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, I)$. L is the lower triangular matrix of the Cholesky decomposition of the posterior covariance, such that $\mathbb{V}[\mathbf{f}_*] = LL^T$, where $\mathbb{V}[\mathbf{f}_*]$ is the posterior covariance, equation (3.4). This process is used to generate the draws in Figure 3.1.

It should be noted that this process is not equivalent to generating samples by adding the mean prediction, $\mathbb{E}[\mathbf{f}_*]$, to Gaussian noise scaled over the magnitude of the confidence interval. Such an approach would have sampled values independent from each other and would result in effectively generating small fatigue cycles within the confidence interval, as shown in Figure 3.2.

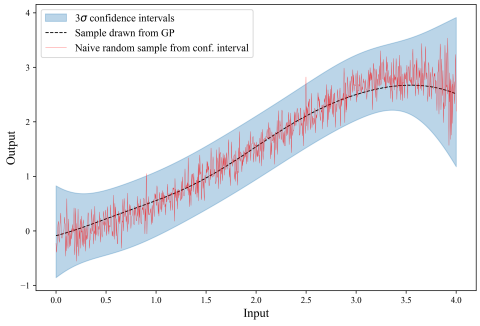


Figure 3.2: Sample taken from GP posterior compared to an independently sampled noise process

The role of kernels

The ability to choose a covariance function, or *kernel*, is fundamental to the flexibility of Gaussian process regression over a range of scenarios. It is the selection of the kernel, along with the optimisation of its hyperparameters, that determines the ability of the model to fit the process under consideration. In Figure 3.3, we can see three GP posteriors conditioned on the same training data using 'off-the-shelf' kernel functions: The squared-exponential, or radial basis function, kernel is the most commonly used despite being known to often over-smooth through the data. The Matern kernel is often used when the data is less smooth, with hyperparameters impacting the roughness of the prediction (is is said that the Matérn class of covariance functions often relects real processes better than the SE [113]). The Cosine kernel is an example of domain-specific knowledge being input into the model at the kernel selection stage: this could be used for modelling a process if the data is known to follow a sinusoidal pattern, but with unknown amplitude or frequency.

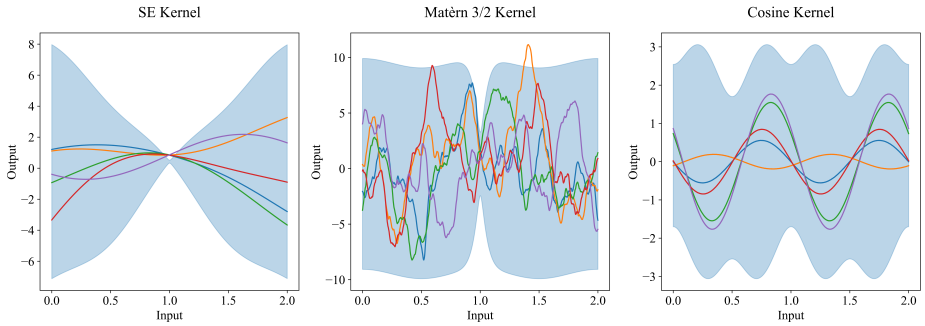


Figure 3.3: The covariance function (and its optimised hyperparameters) define the functional form of the GP posterior. The squared exponential kernel is the most commonly used, but there are a wide range of kernels that can be selected based on prior knowledge of the data. Furthermore, kernels can be combined to provide further flexibility [114]

In combination with the choice of kernel, it is the model hyperparameters that determine the form that the model posterior takes. In Figure 3.4 three GP posteriors can be seen that are conditioned on the same training data, but with varying hyperparameters to demonstrate their respective roles. The signal hyperparameter, σ_f, reflects the strength of the correlation between data points. The lengthscale, l, impacts how rapidly data points become uncorrelated from each other and, as a result, how rapidly the model can respond, with a shorter lengthscale having shorter fluctuations in prediction. The noise hyperparameter σ_n is a parameter of the GP

model, rather than the covariance function, but this specificies the variance of the data being modelled. In practice, an increase in noise hyperparameter results in a larger posterior variance, even close to the training data.

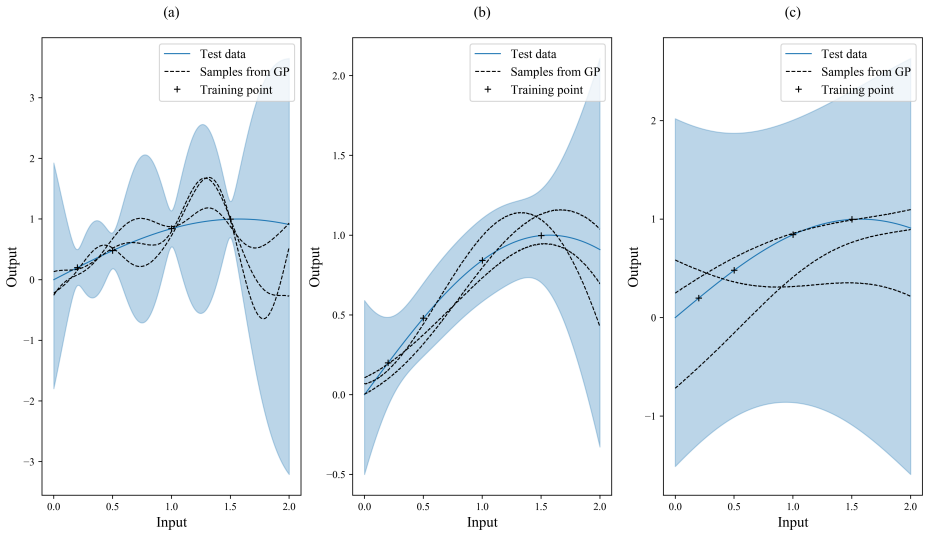
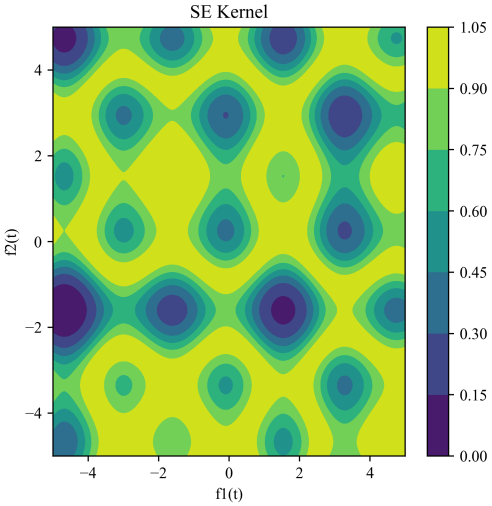
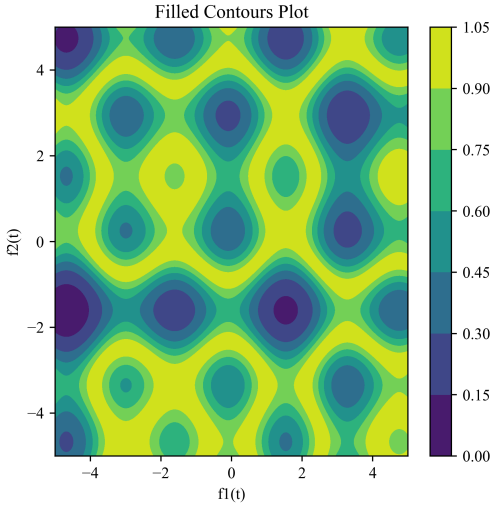


Figure 3.4: The posterior of 3 GPs, each using a squared-exponential covariance function with varying hyperparameters $\{\sigma_f, l, \sigma_n\}$. (a) $\{1.4, 0.3.0.01\}$ (b) $\{3, 1.1.0.1\}$ (c) $\{0.9, 1.1.0.9\}$

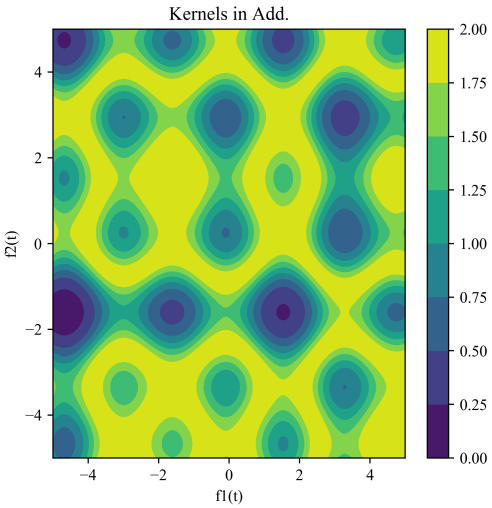
Kernels can also be combined through any linear operator, most commonly addition or multiplication. Figures 3.5a to 3.5d, demonstrate the effect of adding and multiplying covariance functions. Adding kernels together averages their respective impact on the combined covariance. Meanwhile, multiplying kernels exaggerates the impact of both kernels having a high or low value, resulting in steeper gradients between them .



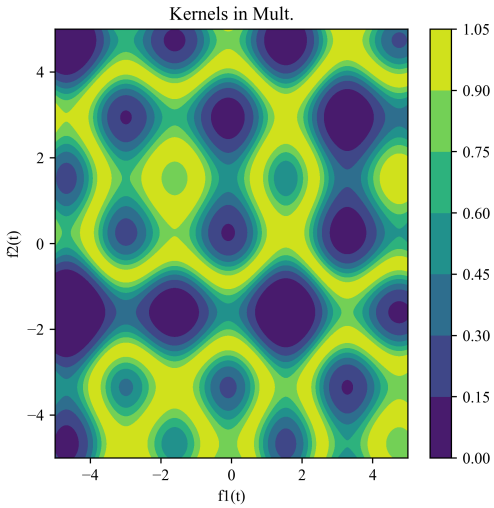
(a) SE Kernel



(b) Matern Kernel



(c) Kernels in addition (a + b)



(d) Kernels in multiplication (a x b)

Figure 3.5: A demonstration of addition and multiplication of kernels. In (a) and (b), a squared exponential and Matérn kernel, respectively, are trained on the same datasets. In (c), we can see a visual representation of kernels (a) and (b) used in addition, where the effect of averaging their respective values occurs. In (d), we can see kernels (a) and (b) used in multiplication, where the impact of high or low values in each kernel is exaggerated, with steep gradients seen between regions.

3.2 Data

For demonstrating the proposed methodology, two datasets will be used in this work. Firstly, data from a Tucano military aircraft; Secondly, data acquired on a model aircraft in the Laboratory for Validation and Verification in Sheffield. Here, the two datasets will be introduced and the data used for demonstrating some of the GP theory introduced above.

3.2.1 Tucano

Data are available, with the thanks of Steve Reed at the Defence Science and Technology Laboratory (DSTL), from a monitoring campaign on a Tucano short aircraft. The monitoring campaign included 101 flights from 2 separate aircraft, with 84 of these being from Aircraft A (and 17 from Aircraft B).

Figure 3.6: [115] A monitoring campaign was undertaken on a Tucano aircraft. This data will form the focus of Chapters 4 and 5.

The data will be briefly detailed here, with available data channels and sampling rates available etc. Much of this information is based on Reed [116], where a much more thorough description of the data is available.

3.2.2 Available channels

Table 3.1 provides a selected overview of the Tucano data with a description of the available measurements and their respective units and sample rates. Figure 3.7 shows

the correlations between five of the inputs presented in Table 3.1.

Clearly, there are strong correlations between the four of the five inputs (the acceleration measurements across the aircraft and the strain on the wing). By contrast, the strain on the tailplane is not correlated well with any of the inputs. This is not surprising as the other data channels are taken from relatively similar locations on the aircraft. The strong correlations between acceleration and strain are a result of the largely static behaviour in the data. Some hysteresis, where the response of the structure is not only a function of its current state but also its past states, can be seen due to the less prominent dynamic behaviour which makes this correlation less strong. For the initial investigations using this data, in chapters 4 and 5, the data is downsampled to 16 Hz and the focus will be on the static component of the data (this is described further in the following chapter). In Chapter 7, the dynamic component of the stress response of the Tucano data will be interrogated more thoroughly.

Conducting this data visualisation step is crucial before constructing a model, as it enables the user to select appropriate inputs, where univariate relationships exist. Including additional inputs that are not appropriate can significantly decrease performance, as further discussed in the uncertainty section below. In the era of 'big-data' solutions, this is important to highlight: the inclusion of data 'just because you have it' is not only *not* useful, but can be harmful to results (in a supervised learning environment).

No.	Channel	Units	Sampling Frequency
1	Fuel	kg	16 Hz
2	Airspeed	kts	16 Hz
3	Altitude	ft	16 Hz
4	Normal acceleration at c.o.g	g	32 Hz
5	Normal acceleration at tail	g	16 Hz
6	Normal acceleration at p.s. wing	$\mu\epsilon$	16 Hz
7	Strain: inner-wing bending (p.s.)	$\mu\epsilon$	128 Hz
8	Strain: mid-wing bending (p.s.)	$\mu\epsilon$	16 Hz
9	Strain: outer-wing bending (p.s.)	$\mu\epsilon$	16 Hz
10	Strain: inner-wing bending (s.b.)	$\mu\epsilon$	128 Hz
11	Strain: tailplane root bending (p.s)	$\mu\epsilon$	128 Hz
12	Strain: tailplane root bending (s.b.)	$\mu\epsilon$	128 Hz
13	Strain: fin root bending (s.b.)	$\mu\epsilon$	256 Hz
14	Strain: main gear - vertical (p.s)	$\mu\epsilon$	128 Hz
15	Strain: main gear - drag direction (p.s)	$\mu\epsilon$	64 Hz
16	Strain: main gear - side direction (p.s)	$\mu\epsilon$	128 Hz
17	Lateral acceleration at c.o.g.	g	16 Hz
18	Rudder position	deg	16 Hz
19	Elevator position	deg	16 Hz
20	Aileron position	deg	16 Hz
21	Weight-on-wheels	1/0	16 Hz
22	Flap position	deg	16 Hz
23	Airbrake position	1/0	16 Hz

Table 3.1: Available data channels for Tucano dataset. (s.b.: star-board, p.s.: port-side, c.o.g.: centre-of-gravity

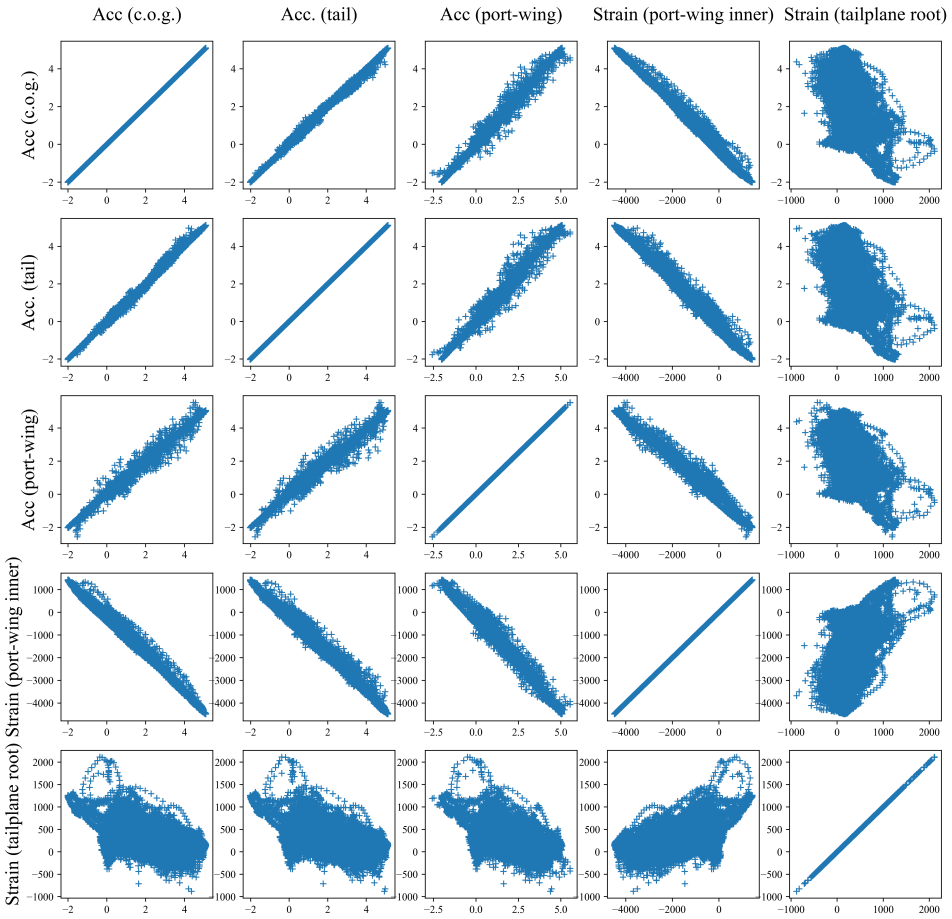


Figure 3.7: Correlation plots for some of the variables shown in Table 3.1

3.3 GARTEUR aircraft

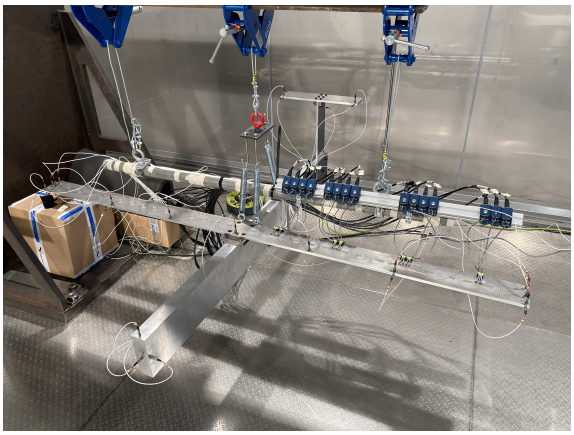


Figure 3.8: The GARTEUR SM-AG19 benchmark in the Laboratory for Validation and Verification, Sheffield

The second dataset used within this work is that of the GARTEUR SM-AG19 aircraft. The SM-AG19 aircraft was originally developed by the Structures and Materials Action Group of The Group for Aeronautical Research and Technology in Europe (GARTEUR) [117–119]. The structure was designed as a benchmark study for modal analysis, with testing performed in laboratories across Europe .

The SM-AG19 was replicated and testing carried out at The Laboratory for Validation and Verification (LVV) in Sheffield. The dimensions of the structure and information of the location of all the accelerometers can be found in [120]. Minor adjustments were made to the original structure: the 'winglets' were removed to encourage more bending strain and lower the torsional component. Tests were performed at a range of temperatures in the laboratory environmental chamber and thus, the original elastomer bungees from the original test were replaced with steel springs that enable a faster response to the temperature conditions. The accelerometers from the original experiment were maintained in their original positions (with the exception of those on the removed winglets). Strain gauges were installed across the starboard wing, providing high-fidelity strain information across wing. The position of the installed strain gauges can be seen in Figure 3.9.

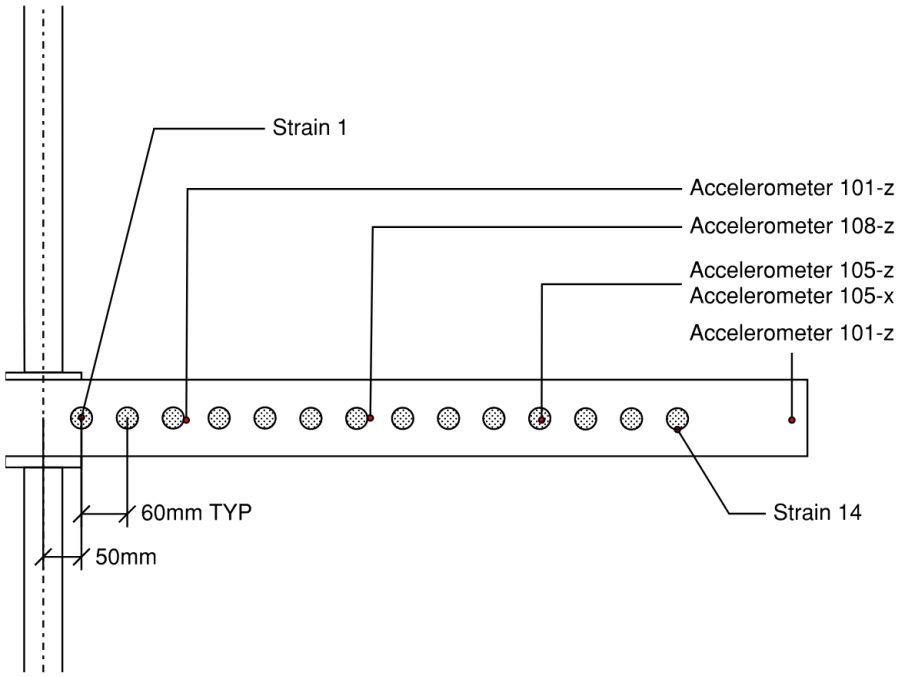
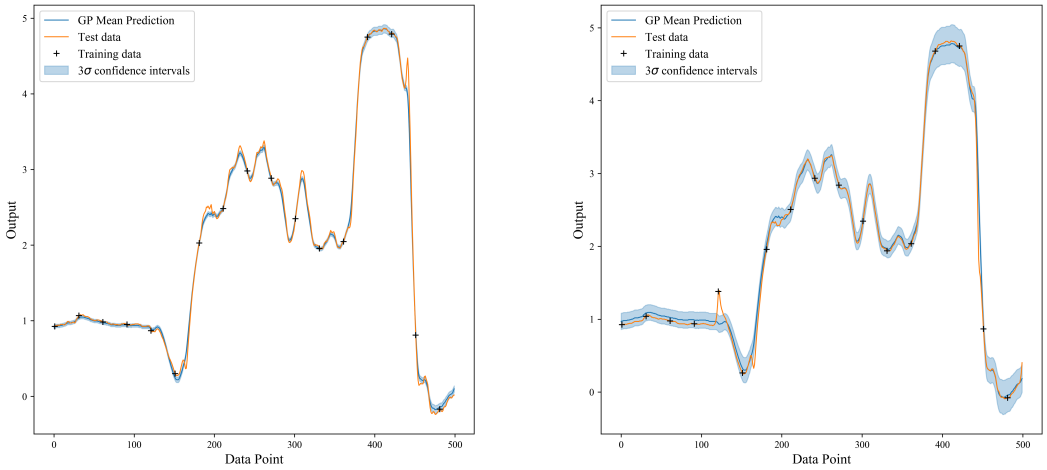


Figure 3.9: Positions of strain gauges on the port-side wing of the GARTEUR aircraft.



(a) (b)

Figure 3.10: Posterior GP for acceleration at the aircraft (a) tailplane (b) port-side wing. We can see that the stronger correlation shown in the correlation plots corresponds to a smaller posterior variance.

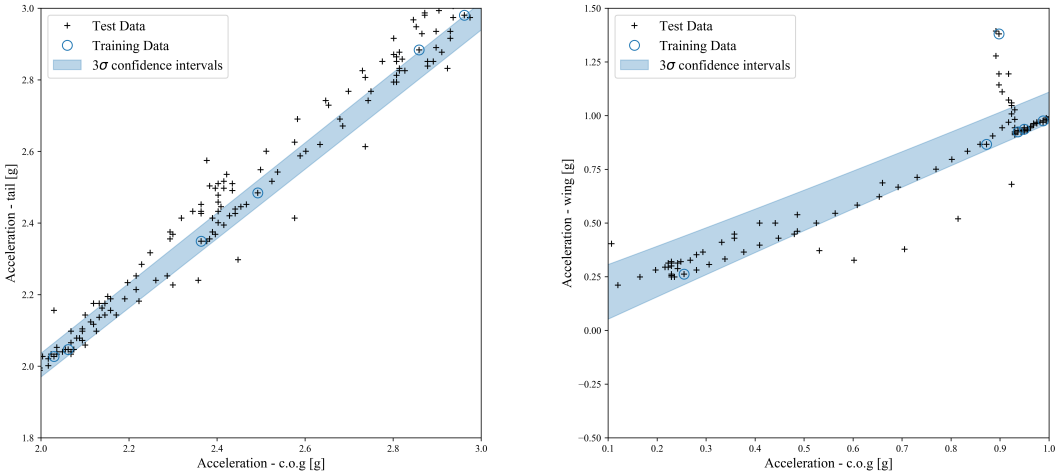
3.4 GP regression demonstration with Tucano data

The Tucano data will now be used to demonstrate the functioning of a GP regression model on a smaller dataset before using throughout the work. In Figure 3.7, we can see that the acceleration at the aircraft centre-of-gravity (first column/row) correlates well with the acceleration at the tail (second column/row) and slightly less well with the acceleration on the wing (central colum / row).

Two models are set up, with the acceleration at the center-of-gravity used as the input in both cases and the tail and wing accelerations used as the respective model targets. The data is downsampled to 16 Hz and the same training and testing data indices used for the two models. The posterior distributions for the two models can be seen in Figure 3.10. The stronger correlation, and reduced variance, shown in Figure 3.7 translates into a smaller posterior variance and improved model confidence. One thing that can be seen in Figures 3.10(a) and (b) is that the training points lie within the posterior confidence intervals, but the variance of the testing data is not accurately captured.

It is difficult to visualise why this is the case by looking at the data in the time domain. However, we can alternatively view the posterior distribution in the input-space, as shown in Figure 3.11. By doing this, it is easier to see that the posterior covariance in (a) is coherent with the training data, despite much of the testing data not fitting the distribution selected. Conversely, the wider variance of training data in (b) results in a larger noise variance in the model, and thus, a much greater number of the testing points are captured.

This can lead us to another conclusion: more data will not necessarily make a model more confident. While increasing data should (if well chosen) reduce the distance between the training and testing data, the increasing data can also provide the model with more information about the true variance of the model. The two models here have the same number of data points, but despite the lower variance of the data for model (a), the GP posterior is overconfident as a result the training data used. Note that the models presented here are for demonstration of the theory and introduction to the data. In both cases, the number of training points is limited and only one input is used to the model. It is unsurprising that the model does not perform perfectly in either case.



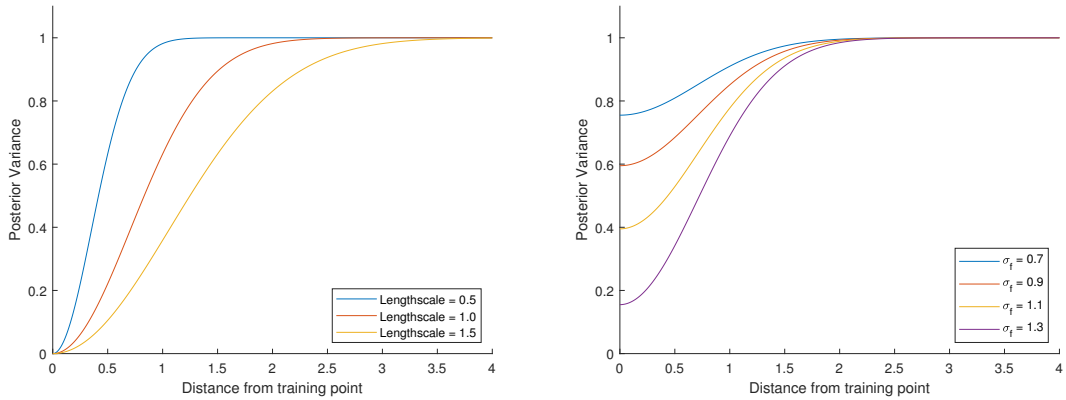
(a) (b)

Figure 3.11: The posterior distribution of the acceleration predictions shown in Figure 3.10 in the input-space, rather than time series prediction.

3.5 Understanding confidence in GPR models

Whilst exploiting the probabilistic nature of the GP model, or any probabilistic model, is advocated, care must be taken in how uncertainty assessments are interpreted. What actually *is* the uncertainty representative of? The uncertainty in the GP prediction in this case is based on a measure of the distance between input training and testing points. When using a typical, data-driven, kernel such as the squared exponential (Equation 3.1), we can see that when the testing data is not close to the training data, the subtractive term in Equation 3.4 tends towards zero, thereby maximising the variance. In this case, the variance depends only on $K(X_*, X_*)$ as $\mathbf{x}$ and $\mathbf{x}'$ are equal and therefore the posterior variance is dictated by the signal variance σ_f. The inputs to the GP and choice of covariance function are specified by the user.

In Figure 3.12, a squared exponential kernel is used to demonstrate how its hyperparameters, lengthscale and signal variance, determine the rate at which the posterior covariance (and thus, uncertainty) grows with increasing distance between testing and training points.



(a) Varying lengthscale (b) Varying signal hyperparameter

Figure 3.12: Posterior variance with SE kernel with distance from training points

So long as the strain prediction quality is acceptable, the selection of the covariance function should not be of large consequence in a purely data-driven model. However, care must be taken in selection of the model inputs. For example, if an irrelevant input were included which exhibited different behaviour between training and testing

sets, this would increase the uncertainty on the fatigue damage assessment in an unrepresentative manner. As with all data-driven methods, prediction quality and reliability of the confidence bounds will decrease in extrapolation. If the relationships between inputs and targets change outside of the training dataset, the inference here will not necesarily be valid.

3.5.1 Input augmentation

One of the aims in this work is to demonstrate the impact of the utilisation of physics in machine learning methods. *Input augmentation* or *semi-physical modelling* is often one of the most simple ways of achieving this. Rather than a modelling choice, this is something that generally happens at the data processing state and involves the manipulation of model inputs in advance of use within a data-driven model to reflect the known physics. For example, drag forces are proportional to velocity squared, so making this adjustment to the indicated airspeed to the input prior to use can better reflect the loads on a structure (see chapters 4 and 5). A demonstration of the impact of input augmentation is presented at the start of the following chapter.

Chapter 4

Tucano aircraft case study: a preliminary study

In this chapter, the Tucano dataset will be used to demonstrate the performance of GP regression for predicting fatigue damage. This will be limited to a deterministic study, without propagating the model uncertainty, novel to this work, which will be the focus of the next chapter. The concept of semi-physical modelling will be used to demonstrate that we can improve our model prediction by effectively pre-processing data.

4.1 Fatigue analysis

In Chapter 1, an overview of the main concepts of S-N fatigue was presented. Here, the particular process followed in this work will be described further, before using the S-N framework for predicting damage accumulation on the Tucano aircraft.

The S-N curve is the foundation upon which the stress-life methodology is built. Constant-amplitude cyclic stresses are applied in laboratory conditions to test how many cycles a specimen can withstand before failing at a given stress level. The result is a relationship between stress range and the number of cycles to failure that is generally linear on a log-log scale: the most common analytical expression for the S-N relation is the Basquin model [121]:

$$\sigma_a = aN_f^m \tag{4.1}$$

where σ_a is the stress amplitude, a is a numerical constant, N_f is the number of cycles to failure at stress σ_a and m is the exponent which gives the slope of the relationship (which is linear on a log-log curve).

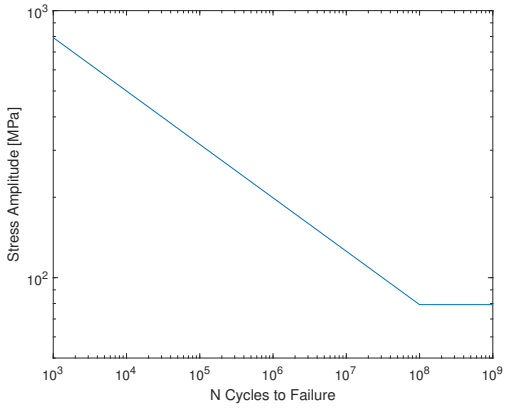


Figure 4.1: The S-N curve describes the number of cycles to failure that a material can withstand at a given stress range.

For a component in operation, prior to using the S-N curve to predict fatigue damage, a cycle counting method has to be used to convert a stress response into a set of fully-reversed stress ranges. The most commonly used is the Rainflow count, as is used in this work, following the ASTM 1049 standard [122].

A cycle counting method should also provide the mean stress for each cycle in addition to the stress amplitudes within the response. This also has an impact on its subsequent damage: a tensile mean stress will result in a faster failure and vice versa for compressive mean stresses [123]. As S-N curves are generally developed under zero-mean applied stresses [124], a correction has to be applied for stress ranges with a non-zero mean stress. Methods such as the Goodman [125] correction can, thus, be used to compensate for the altered damage from the non-zero mean (other examples are the Gerber [126] or Soderberg [127] methods). This can be done on a cycle-by-cycle basis following the cycle count.

Having calculated a mean-adjusted cycle range, the damage from each cycle can be calculated using the S-N curve. Palmgren-Miner rule is simply that the total damage from the complete stress response is the sum of the damage caused by each cycle [6]:

$$D = \sum \frac{n_i}{N_i} \tag{4.2}$$

where n_i and N_i are the number of cycles in the stress response at a given stress range (taken from the rainflow cycle count) and the number of cycles to failure the material is capable of withstanding at that stress level, respectively. D is the summation of damage from each of the individual stress cycles identified within the stress response, with the material *expected* to fail when $D = 1$.

For the work in this chapter and Chapter 5, fatigue analysis is performed using the S-N curve described by Reed [80] for an L-65 aluminium alloy, including the Goodman mean stress correction [128]:

$$N = \left[\frac{A\sigma_\infty \left(\sigma_u - \sigma_m\right)}{\sigma_a \left(\sigma_u - \sigma_{mi}\right) - \sigma_\infty \left(\sigma_u - \sigma_m\right)} \right] \tag{4.3}$$

where N is the number of cycles to failure at a given stress amplitude, σ_u is the ultimate tensile strength of the material, σ_m and σ_{mi} are the cycle mean stress and S-N curve mean stress, respectively. The numerical constant A can be found in [128].

4.2 Model inputs and targets

In this and the following chapter, the strain on the port-side wing shall be used as the target of the model. The wing is selected as the target as this is the most important aircraft component from a fatigue perspective owing to its requirement to provide the aircraft lift and support fuel loads while remaining light [116]. Strain measurements are available at three points across the wing, but the inner strain measurement is chosen as the target throughout the work as it experiences the greatest stress and thus, fatigue damage accumulation.

Candidate inputs to the GP model are acceleration measurements in combination with some standard in-service measurements. The acceleration measurements used are located on the port-side wing and at the aircraft centre of gravity (channels 4 and 6 in Table 3.1). The standard aircraft measurements used are airspeed, and altitude. Note that no strain measurements from other locations on the aircraft are

used as model inputs. The motivation for this is that accelerometers are much more robust than strain gauges, which would be difficult to maintain beyond an initial fixed-duration study.

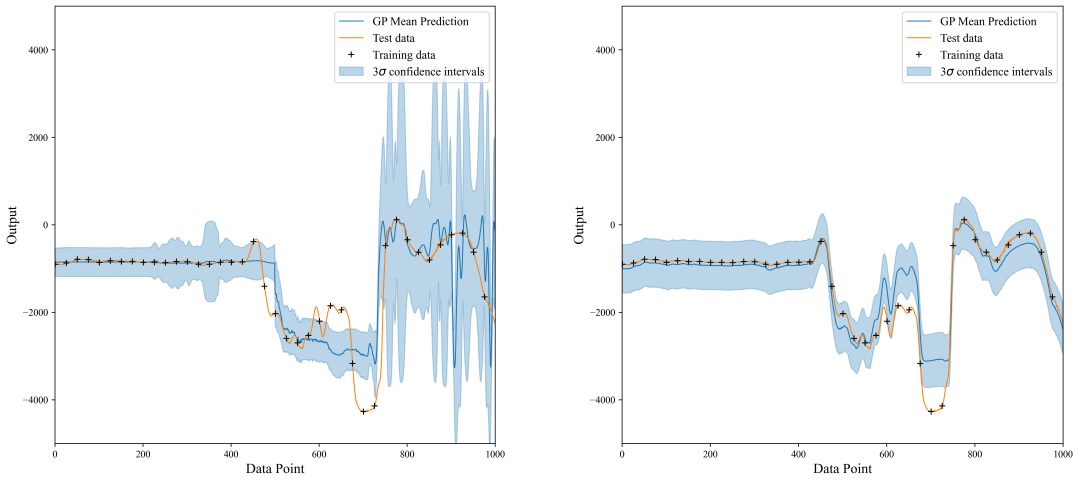
The work in the following two chapters will be focussing on the 84 flights available from Aircraft A. Many of the inputs shown in Table 3.1 are only available at 16 Hz, while the strain data is available at 128 Hz. Therefore, in this and the following chapter, the complete dataset is downsampled to 16 Hz before use. As this does result in some loss of dynamic strain behaviour for some of the flights, the analysis shall initially only be assessing the quasi-static behaviour shown at this downsampled frequency. Later on, in Chapter 7, the Tucano data will be used again to demonstrate interpolative performance with a physics-informed ML approach.

At this downsampled rate, each of the flights has between 20,000 and 90,000 data points. The GP posterior sampling operation requires a matrix inversion of complexity $\mathcal{O}(n^3)$ and therefore, for computational reasons, the study here is limited to 10,000 data points from each flight for demonstration. The most damaging 10,000 data points in each flight are selected (based on fatigue analysis performed on the measured data before prediction) to ensure that non-negligible damage accumulation occurs in the time history under consideration.

4.2.1 Input Augmentation

The concept of *input augmentation*, or *semi-physical* modelling, was introduced in the previous chapter. Following, [81], such adjustments were made to some of the inputs before use within the model. To briefly assess the potential impact that this input augmentation step has, in Figure 4.2, two GP posteriors are shown. The model is simple, with only two inputs used (airspeed and normal acceleration at the aircraft c.o.g.). The target of the model in both cases is the strain on the aircraft wing. In (a), an SE kernel is used with no special treatment of the input data (except for standardising before use in the model). In (b), the same inputs and training indices are used with an SE kernel, however, before use in the model, the airspeed is squared to indicate the dynamic pressure, and the acceleration measurements were multiplied to indicate the forces acting on the aircraft. This latter step is performed, following [81]. However, in reality, the input data is standardised prior to its use in the GPR model throughout this work and thus, this step could be omitted. This is not the

case when squaring the airspeed, as the significance of this is maintained after the data is standardised.



(a) (b)

Figure 4.2: (a) Strain prediction model using inputs of normal acceleration at the plane centre-of-gravity and indicated airspeed (MSE = 41.0). (b) The same inputs as in (a) but with the inputs manipulated to give an indication of physical processes acting on the aircraft (MSE = 13.4).

While one can see that the strain prediction is still not perfect in (b) (note that the model is still only using two inputs), we can see a significant improvement in prediction. This is quantified by a reduction in the model NMSE (see Equation 4.4) from 41.0 to 13.4. Importantly, the underlying nature of the strain is captured much more accurately: the fluctuations in strain are captured, albeit with errors in their amplitudes. By contrast, in (a), the GP posterior is smoothing through many of the turning points. Again, while again not perfect, we can see that the model confidence is more representative in model (b), especially around the strain minima, around data point 700. Further information about possible domain-specific knowledge for the Tucano data can be found in [81].

4.3 GP Model

The model was trained using 2000 data points, randomly selected from five training flights, and then tested on all 84 flights from Aircraft A (including the five training flights).

Normalised mean square error (NMSE), defined by equation (4.4), is frequently used as a metric for assessing the prediction accuracy of regression models:

$$\text{NMSE} = \frac{100}{n\sigma_y^2}\sum(y_p - y_i)^2 \tag{4.4}$$

where y_p and y_i are the values of the prediction and the measured data at data point i, σ_y is the standard deviation of the measured values and n is the number of data points.

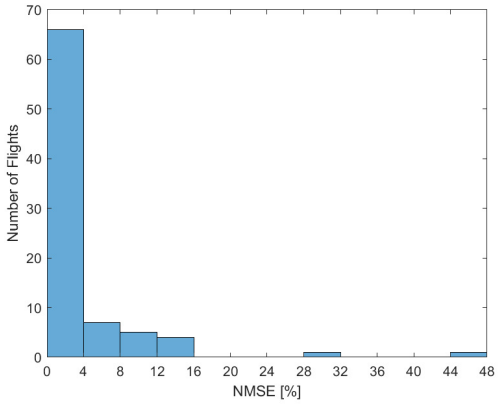


Figure 4.3: Histogram of NMSE values for the 84 flights

The strain prediction shows a very good level of accuracy across the majority of flights, with 82 of the 84 flights having an NMSE value less than 15.7% with most of of these being below 5%. The range of errors is displayed in a histogram shown in Figure 4.3. There are two flights with NMSE values greater than this (31.1% and 46.9%). The worst of these (Flight 12) can be seen in Figure 4.4 below.

Here, we can see that the prediction is generally good over the 10000 data points, but there is a region in which the prediction struggles between 3000 and 5000 data points. This results in the poor NMSE seen in Figure 4.3. Furthermore, these two flights

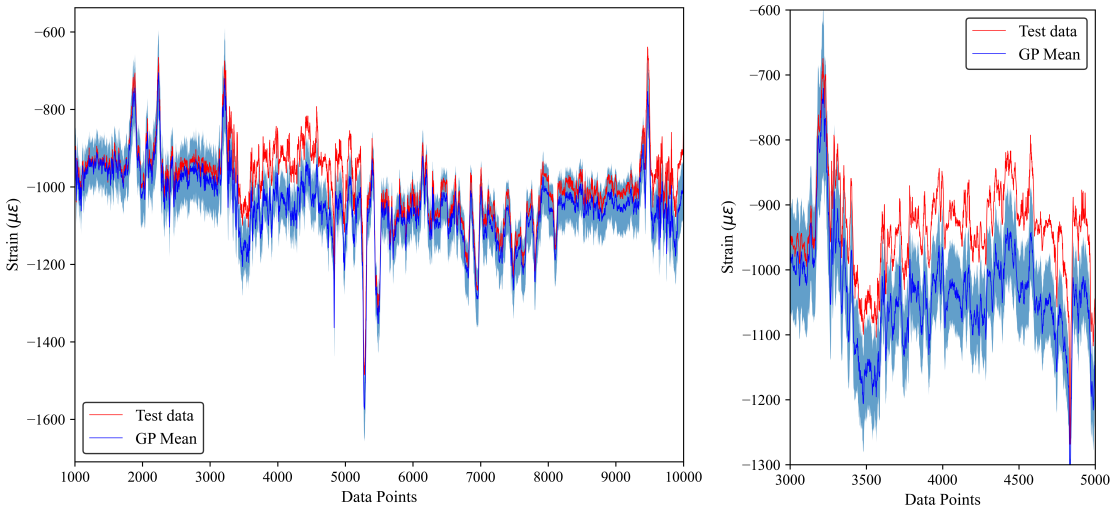


Figure 4.4: (a) Strain prediction model for the worst performing flight in terms of NMSE (flight 12). (b) An enhanced view of (a). We can see that between data points 3000 and 5000, the prediction is poor, resulting in the high NMSE.

(Flight 12 and Flight 23, which also had a large NMSE of 31%) showed relatively low stress amplitudes which are difficult to capture due to their benign nature and this amplifies the modelling error. The low damage accumulated in these flights are highlighted in the following fatigue assessment.

By contrast, a typical example of a good flight is seen in Figure 4.5. The majority of flights have NMSE values less than 4%, so an example from within this range is selected. Flight 73 has an NMSE of 3.4%: Only 1000 (of 10000) data points are shown in Figure 4.5, but this is representative of the full flight, with the posterior mean matching the measured data closely and remaining within the confidence intervals throughout the prediction.

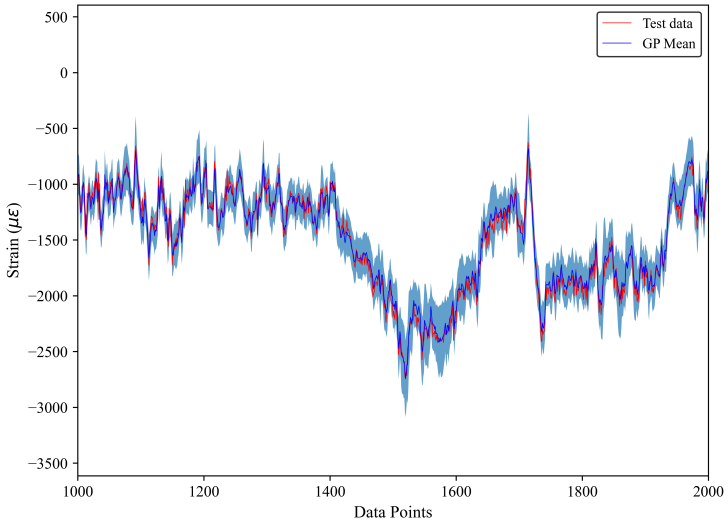


Figure 4.5: Strain prediction model for a typical flight (Flight 73). The majority of flights have an NMSE value of less than 4%. In this case, the NMSE is 3.4%.

For each prediction, the GP posterior mean is now used for predicting fatigue damage using the S-N approach and compared to the damage from the measured strain data, which serves as a benchmark. The result of this can be seen in Figure 4.6. For the majority of flights, a good level of accuracy is achieved, with a close correlation between predicted and measured damage. The flights discussed through this, and the following chapter are indicated.

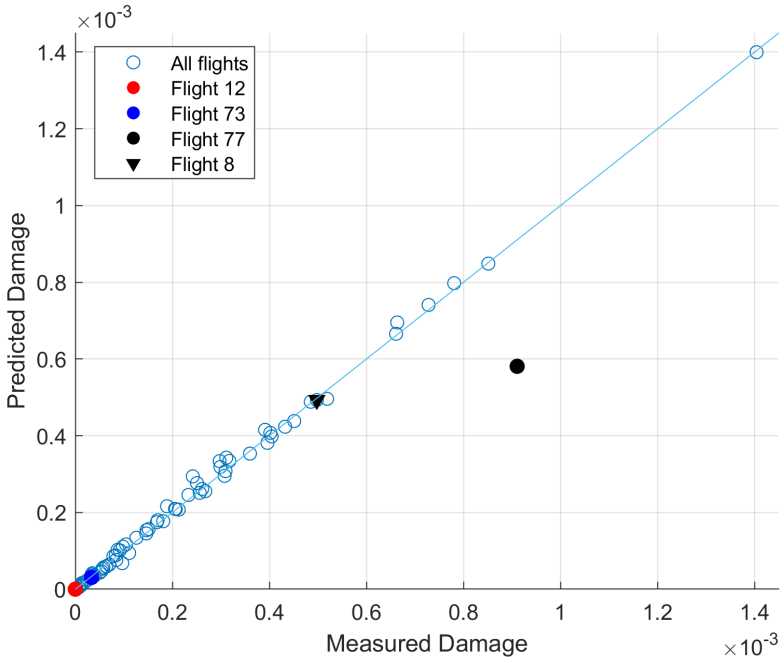


Figure 4.6: Prediction of fatigue damage for the 84 flights using the GP posterior mean

We will look briefly at two further flights, identified as flights 77 and 8 in Figure 4.6. First, we will look at Flight 8, for which the GP posterior is shown in Figure 4.7. From both the model posterior and the fatigue damage prediction, we can see that this is an example of a good prediction from the test set of flights. The posterior mean prediction follows the measured damage very closely, verified by a low NMSE value of 0.9%. We can see from Figure 4.6 that the error of the GP mean at predicting fatigue damage compared to the measured data is very low, underestimating the measured damage by only 1.8%.

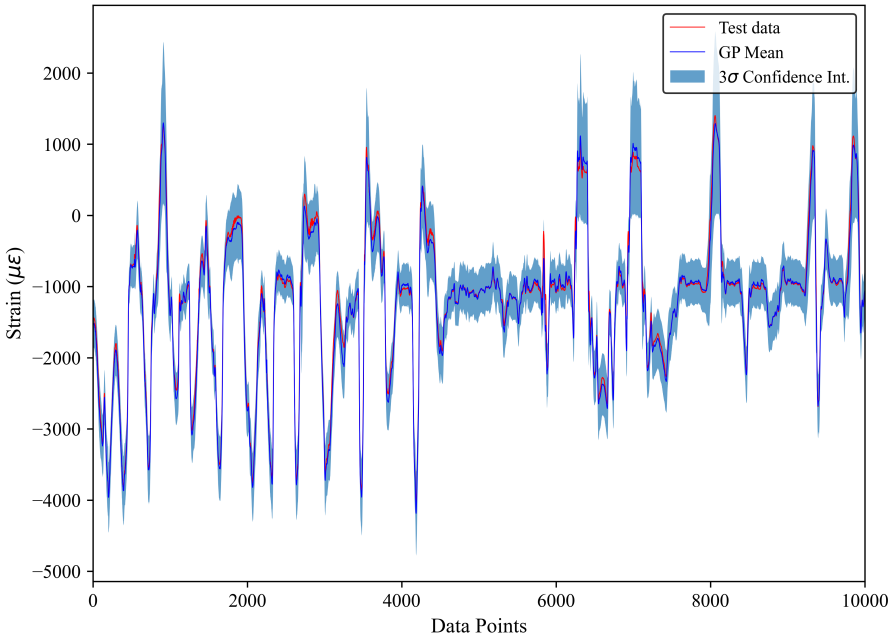


Figure 4.7: GP Posterior for Flight 8: we can see that the GP mean predicts the measured data very closely.

By contrast, Flight 77 is the worst performing flight of the 84 in terms of fatigue damage prediction (Figure 4.6), underestimating the measured damage state by 36.2%, and is the clear outlier of the test set. While this flight is the worst for fatigue damage prediction, it is worth noting that the NMSE is only 9.8%, which can be considered relatively low. Clearly, small errors in strain prediction can have a much larger impact on damage prediction. We see that NMSE is not necessarily a good indicator of whether a model will be able to predict fatigue damage accrual well (despite having the worse performance in terms of fatigue damage prediction, the NMSE of Flight 77 is not the worst of the test set). Generally, it is the accuracy of at the local maxima and minima of the largest stress cycles that is most important here, which is not reflected in an error metric which looks at the average error across the stress response.

The GP posterior for Flight 77 is shown in Figure 4.8. Comparing the data to Flight 8 (Figure 4.7), we can see that this flight has many repeated stress cycles of similar magnitude, whereas Flight 8 has a smaller number of large stress cycles with many smaller cycles.

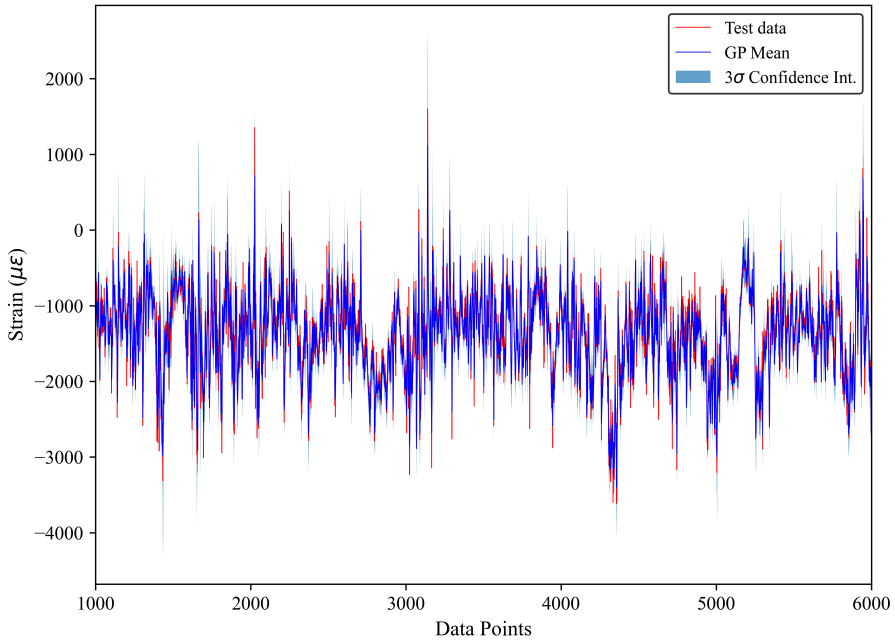


Figure 4.8: Flight 77: GP posterior, data points 1000-6000

As it is difficult to see the data clearly, two sections of the same prediction are presented again in Figures 4.9 (a) and (b) below. While the prediction is generally good (hence, a NMSE value of less than 10%), the GP mean does underestimate the measured value at numerous points over the stress time-history. Even in the small number of test points shown here, this occurs numerous times. As this is one of the more damaging examples of the 84 flight test set, this prediction error propagates into an large under-estimation in the damage arising from the flight. Compared to Flight 12 (Figure 4.4), which had a much larger error in terms of stress prediction, the stress magnitudes at which the errors are occurring here are larger, thus propagating to a greater error in the fatigue prediction.

4.3.1 Discussion and conclusions

This chapter has demonstrated the potential of fatigue prediction using a data-driven framework on the Tucano dataset. This is achieved by performing strain prediction on the most damaging sections of all 84 flights available (from a fatigue perspective) and propagating the mean prediction through the fatigue assessment. This is then compared to the results from the measured data and opens an avenue for better understanding model performance given the ultimate use case. As one of the

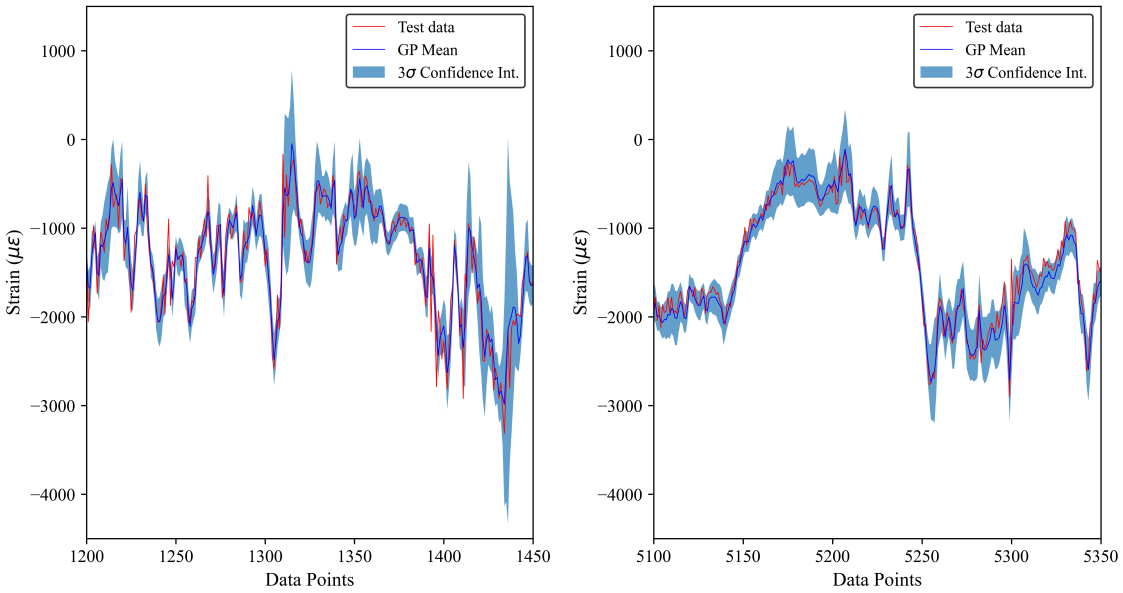


Figure 4.9: Two regions of Flight 77, each showing only 250 data points (of 10,000)

central themes of this work is the utilisation of physics-informed machine learning methods, the benefits of performing *input augmentation* prior to use in the model was demonstrated. The ability to predict stress time-histories is demonstrated, with 98% of the 84 flights in the test set having an NMSE of less than 16% (Figure 4.3). By propagating the strain prediction through the fatigue assessment, the subsequent fatigue damage prediction was shown to be good, with, generally, an acceptable level of error being achieved in the prediction.

One of the advantages of the Gaussian process regression is its probabilistic nature and the provision of a full predictive distribution rather than single point estimate. The remainder of this thesis seeks to utilise the full predictive distribution to both provide a distribution of potential fatigue damage accumulation for use in a probabilistic risk assessment, for example, and to improve the robustness of any such prediction. While the fatigue damage for the majority of flights was able to be predicted with good accuracy, there was one outlier to this trend. For Flight 77, the reason for the poor performance was evident from the strain-time prediction, with the GP posterior mean underestimating the measured value at numerous points in the prediction. Yet, it could be seen that at most of these points, the error was captured within the confidence bounds of the prediction. The hypothesis to be examined in the next chapter is that robustness here will be improved through consideration of the entire

predictive distribution rather than just the mean.

In the next chapter, the uncertainty of the prediction will be propagated through the fatigue assessment. In doing so, one can query if it is possible to mitigate the errors in strain prediction if the posterior variance is sufficiently accurate. Furthermore, even for the flights which are predicted with a very high degree of accuracy, it is proposed that propagating the model error will give a better *true* reflection of the model: if one has two predictions that are identical from the perspective of the GP mean, but one is much more confident about this prediction, it is argued that this should be reflected in the final damage prediction.

This discussion is especially relevant for fatigue assessment. While a static overload assessment could potentially utilise the probabilistic nature of the prediction by simply checking if a given threshold is surpassed at any single point in time, fatigue assessment requires the time history known for a period of time to be able to identify the stress cycles.

This discussion would become increasingly pertinent as the aircraft is in service for longer, as the likelihood of the plane seeing manoeuvres that were not captured by the training data increases (as does the likelihood of seeing increasingly rare environmental conditions e.g. extreme wind gusts or extreme temperatures). Again, while the deterministic results presented in this chapter demonstrated good ability to predict the strain and subsequent damage seen in the test set of 84 flights, by remaining within a deterministic framework, there would be no inherent avenue for assessing the impact of these new conditions.

Chapter 5

Tucano aircraft case study: developing a probabilistic perspective

In this chapter, using the same case study from the previous chapter, the process will be adapted to allow for a probabilistic treatment of the problem by propagating the model uncertainty through the fatigue assessment. In doing so, a distribution of predicted damage, rather than a deterministic assessment, will be acquired. In the previous chapter, it could be seen that even flights with relatively low error on the strain prediction, could result in significant error in the subsequent fatigue assessment: by propagating the model uncertainty, it is hoped that in such cases, a better reflection of the true damage is available. The goal of doing this is to develop a more robust prediction and better interpretation of the model probabilistic prediction.

An overview of the process proposed in this chapter, and how it differs to the deterministic approach taken in the previous chapter is presented in Figure 5.1.

5.1 Propagating model uncertainty

First, a brief overview of the sampling theory, that was discussed in more detail in Chapter 3, will be covered. The GP posterior mean, $\mathbb{E}[\mathbf{f}_*]$, and covariance, $\mathbb{V}[\mathbf{f}_*]$, were defined as follows [112]:

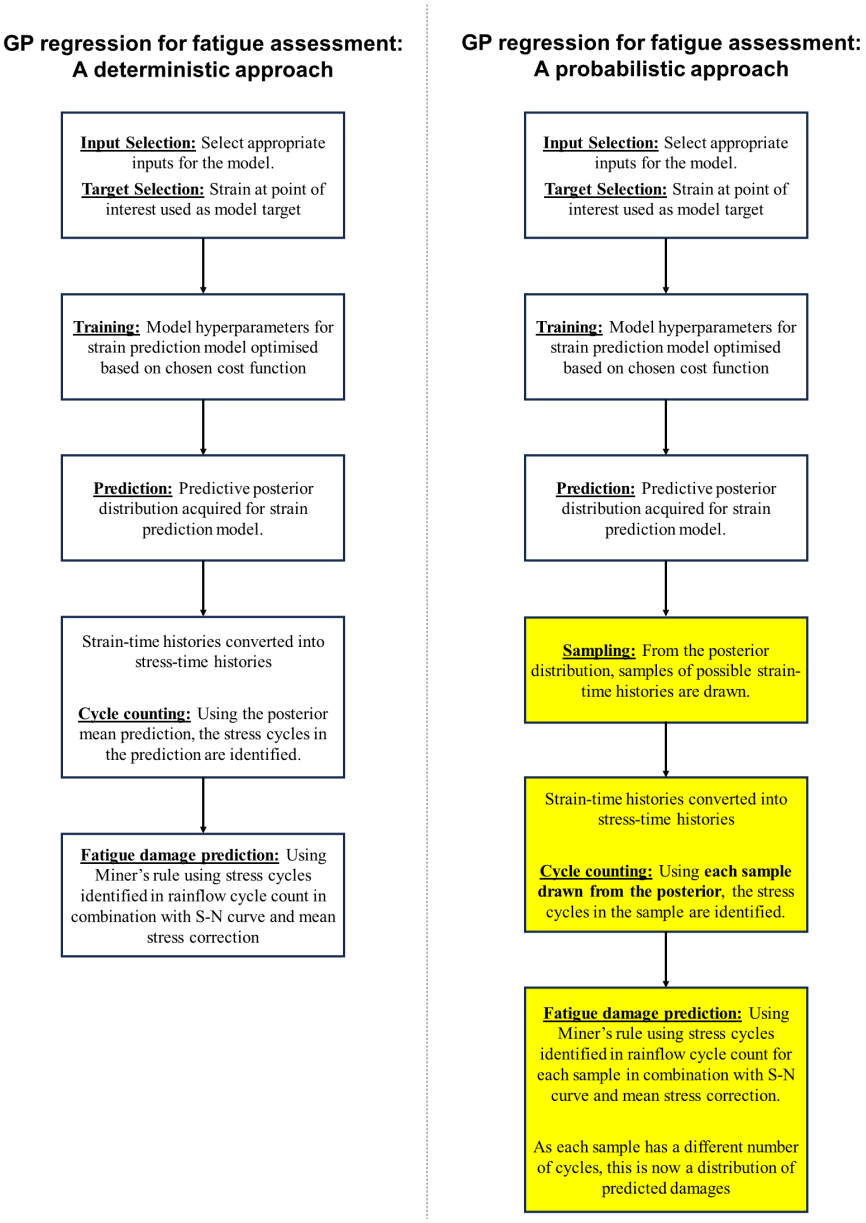


Figure 5.1: A probabilistic perspective on using GP regression for fatigue assessment

$$\mathbb{E}\left[\mathbf{f}_*\right] = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1}\mathbf{y} \tag{5.1}$$

$$\mathbb{V}\left[\mathbf{f}_*\right] = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1}K(X, X_*) \tag{5.2}$$

It is the diagonal values of the posterior covariance matrix that provides the user with the confidence intervals seen throughout this work. However, one can utilise

the full posterior covariance matrix, $\mathbb{V}[\mathbf{f}_*]$, by computing:

$$\mathbf{f}_{sample} = \mathbb{E}[\mathbf{f}_*] + L\mathbf{u} \tag{5.3}$$

where $\mathbf{u}$ follows an independent Gaussian distribution, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, I)$ and L is the lower triangular matrix of the Cholesky decomposition of the posterior covariance (again, see Chapter 3 for further detail). For Flight 12, identified in the previous chapter (Figure 4.4), this is demonstrated in Figure 5.5 for data points 5800-5900.

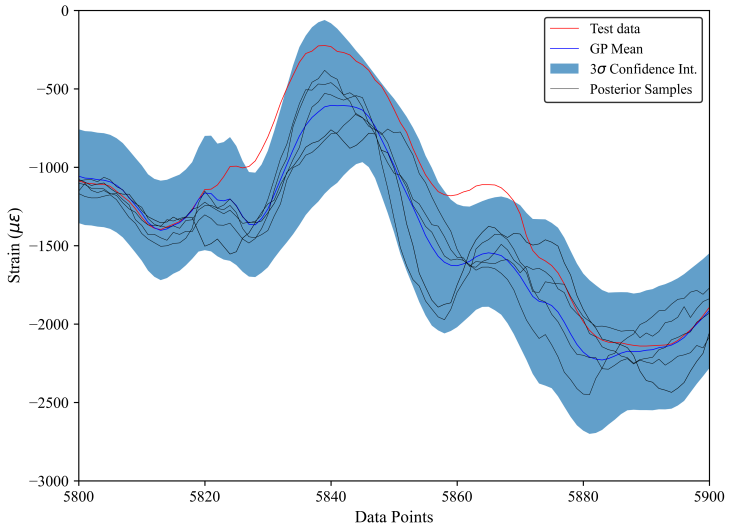


Figure 5.2: Flight 12, identified in the previous chapter, now also with samples drawn from the model posterior.

Now, using the same model as in the previous chapter (including the optimised hyperparameter values), the model uncertainty propagated through the S-N fatigue assessment for all 84 flights. For each flight prediction, 10000 draws are taken from the GP posterior and fed through the fatigue assessment to create a distribution of predicted fatigue damage accrual for each flight. A flowchart, adapted from Figure 5.1, specific to this model is shown in Figure 5.3.

In doing this, the uncertainty of the model can be understood. There are two types of uncertainty; aleatoric and epistemic. The former is related to the uncertainty of the result, which in this case could be considered to be the error of the strain gauges that are used to train the model. This type of uncertainty is considered in [81], where an error of $\pm 5\%$ on the strain gauges is simulated and a subsequent upper and lower limit on the measured fatigue damage is shown. Epistemic uncertainty is related

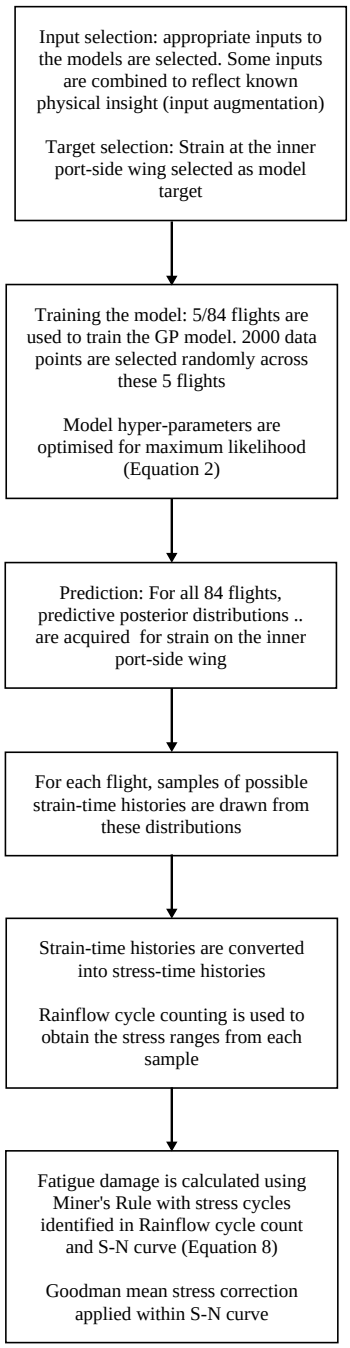


Figure 5.3: Flowchart of proposed methodology

to the uncertainty of the model and is often the result of the availability of data. While epistemic uncertainty can be *reduced* with the use of increasing data, this is not always possible within a data-driven approach: Firstly, the quantity of data is not always the limit to having representative training data, but the availability of training data that is representative of the testing conditions. Secondly, there are

considerable computational demands that need to be considered. This work is looking at uncertainty of both types: the noise hyperparameter represents the uncertainty in the data, while the magnitude of the posterior variance is also a function of the distance between the training data (model inputs) and test data, see equation (3.4).

As per the flowchart, the other parameters in the process are kept the same. The hyperparameters attained during the model optimisation for the previous chapter are kept throughout the results shown in this chapter. The same S-N curve (Equation 4.3 is used throughout the Tucano work.

5.2 Results

As per the previous chapter, inner strain measurement is chosen as the target throughout the work as it experiences the greatest stress and thus, fatigue damage accumulation. The model is maintained as per the last chapter, including the optimised hyperparameters. However, as indicated in 5.3, 10,000 samples are now drawn from the posterior distribution.

After following the methodology set out in Figure 5.3, one now has a distribution of predicted damage for the inner wing for each flight, as opposed to a deterministic value as predicted in the previous chapter. The predicted fatigue accumulation distributions for all 84 flights are visualised in Figure 5.4. Each vertical contour shown represents the damage distribution for a single flight and the solid line marks where the predicted damage is equal to the damage accrual from the measured strain data, as per the previous chapter. The last chapter demonstrated that flights with relatively low error on the strain prediction could still result in error in terms of fatigue damage prediction. The first result that can be noted here is that, for each flight, the predicted damage distributions encompass the measured damage, immediately indicating an increased robustness of the method. Circular markers are added to show the fatigue damage prediction when the GP mean is propagated through the fatigue assessment, as per the previous chapter.

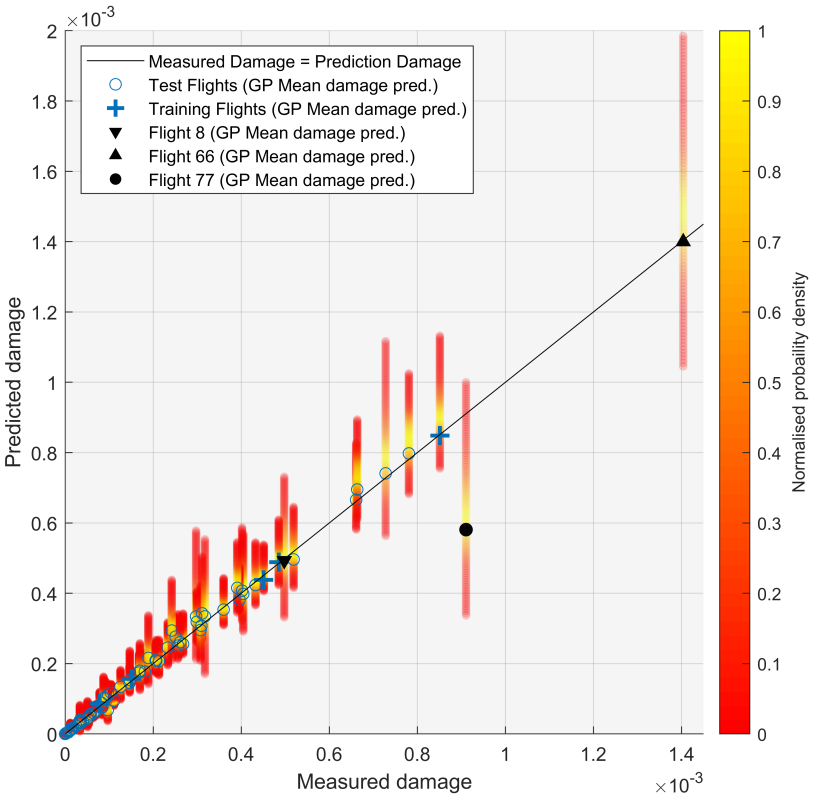


Figure 5.4: Fatigue damage distributions from the GP posterior mean, individual posterior samples and the measured data for all 84 flights. The flights studied in further detail (flights 8, 66 and 77) are indicated.

Looking more closely at the results now, the overall trend shown is that the most likely value from the sampled damages is more conservative than the mean GP prediction: an expected result of sampling from the GP and the subsequent additional damage caused by larger cycles, discussed more below. Each of the contours shown in Figure 5.4 may also be viewed as a histogram. Figure 5.5 shows this for Flight 8, first shown in the previous chapter. In Figure 5.4, one can see that the peak of the damage distribution (i.e. the mode average) of the sampled damages is consistently more conservative than the propagated GP mean prediction across the range of flights. For Flight 8, whose histogram is shown in Figure 5.5, the mean and median of the sampled damages are also shown. For this flight, each of these are more conservative than the measured damage, in part due to the damage prediction being so accurate for this flight using the GP posterior mean.

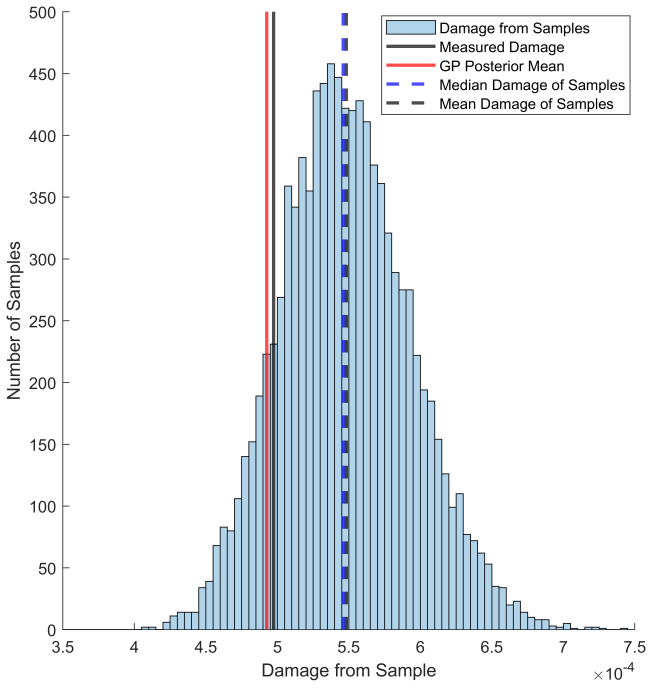


Figure 5.5: Histogram of damages for Flight 8.

From Figure 5.4, we can also see a general increase in the variance of the damage distribution with an increase in flight damage. Again, the non-linear relationship between stress range and cycle damage is partly a cause of this. However, there is a further variable that impacts this: the stress cycle mean value.

As discussed in the previous chapter, we know that the cycle mean stress can also have an impact on its subsequent damage: a tensile mean stress will result in a faster failure and vice versa for compressive mean stresses [123]. As S-N curves are generally developed under zero-mean applied stresses [124], a correction has to be applied for stress ranges with a non-zero mean stress. Methods such as the Goodman [125] correction can, thus, be used to compensate for the altered damage from the non-zero mean. For the fatigue assessment on the Tucano aircraft, the fatigue assessment is performed using the S-N curve specified by 4.3, which has a mean stress term included. To demonstrate the potential for mean stresses to significantly impact damage, Figure 5.6 shows how the variance of damage grows with that of cycle amplitude and mean. The blue line shows the relationship between cycle range and subsequent damage. The uncertainty associated with cycle amplitude is shown in purple (here the cycle range is sampled from a Gaussian distribution with a standard

deviation equal to 2% of the stress amplitude). The effect of having uncertainty on the mean stress of ±100 MPa is shown in pink, where one can see further growth in damage variance. Sampling from the posterior distribution can lead to large changes in the mean stress of the largest (and most damaging) cycles. The impact that this can have over thousands of stress cycles is large and difficult to fully understand or predict due to the 'black-box' nature of the rainflow cycle counting process.

Table 5.1: Comparison of the largest stress cycles for 2 of the samples drawn from Flight 8

Reference	Stress Range [MPa]	Mean Stress [MPa]	Damage
A	389.3	-87.6	1.350e-04
B	389.6	-117.0	1.120e-04

In Table 5.1, we can see the details of the largest stress cycles of two draws from Flight 1 posterior. Despite the stress ranges for the two cycles being almost equal, we can see that cycle B is around 19% more damaging than cycle A as a result of the difference in mean stress. When a large amount of the damage for each flight arises from a small number of large stress cycles (see: Figures 5.8 and 5.10(b)), the impact of mean stress effect of the damage on the sampled flight could be significant.

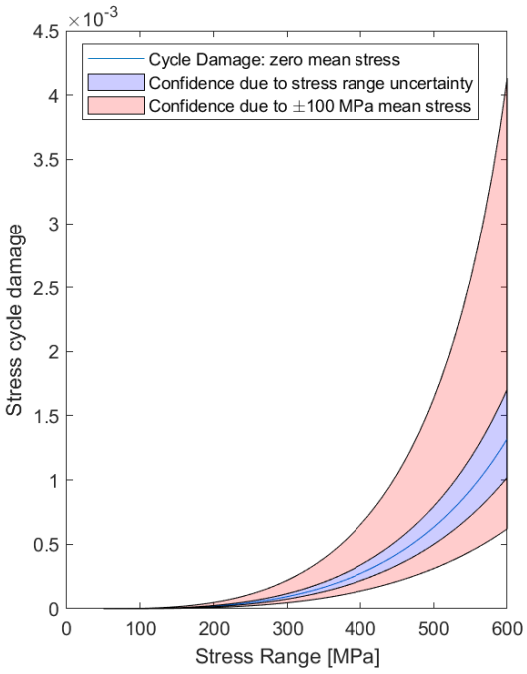


Figure 5.6: Impact of cycle mean stress and uncertainty of stress cycle range on damage

5.2.1 Flight 66

Flight 66 is the most damaging of the 84 flights shown in Figure 5.4. The posterior distribution for this flight can be seen in Figure 5.7, where one can see that the prediction is good across the flight, both in the regions of the local maxima and also the more benign period towards the end of the flight data. We can see that there is a small mean offset when comparing the mean prediction to the measured data, but for the majority of the flight, the measured value is captured within the confidence bounds of the prediction and the important characteristics of the stress response are captured by the prediction. The prediction NMSE value is only 0.9% and the fatigue damage prediction for the GP mean is also very accurate, with an error of less than 0.5%. The distribution of damage demonstrated by the contours in Figure 5.4 shows that, despite this good accuracy (for both strain prediction and using the GP posterior mean to predict the fatigue damage accumulation), the variance of sampled damage can still become large. The result of this is that the most damaging sample is 1.4 times more damaging than the measured data, despite the high accuracy of the prediction. The sampled mean fatigue damage is also more conservative than

the GP mean and measured predictions, although, this is only true by 6%.

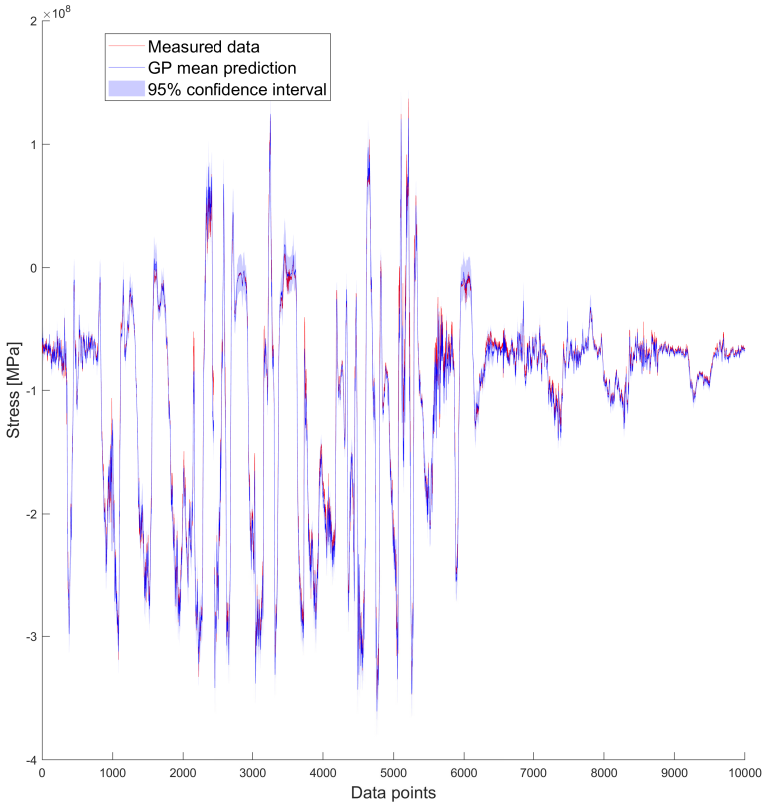


Figure 5.7: Posterior GP stress prediction for Flight 66

The load spectra (stress ranges versus cycle counts) from the measured strain and the most and least damaging samples drawn from the GP are shown in Figure 5.8. The plot also shows the progression of damage across the stress distribution for the measured data and one can see that the majority of damage is caused by a relatively small number of high stress cycles.

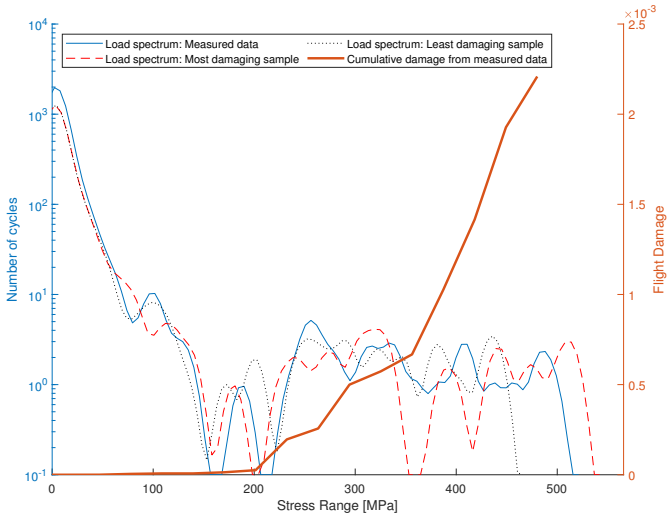
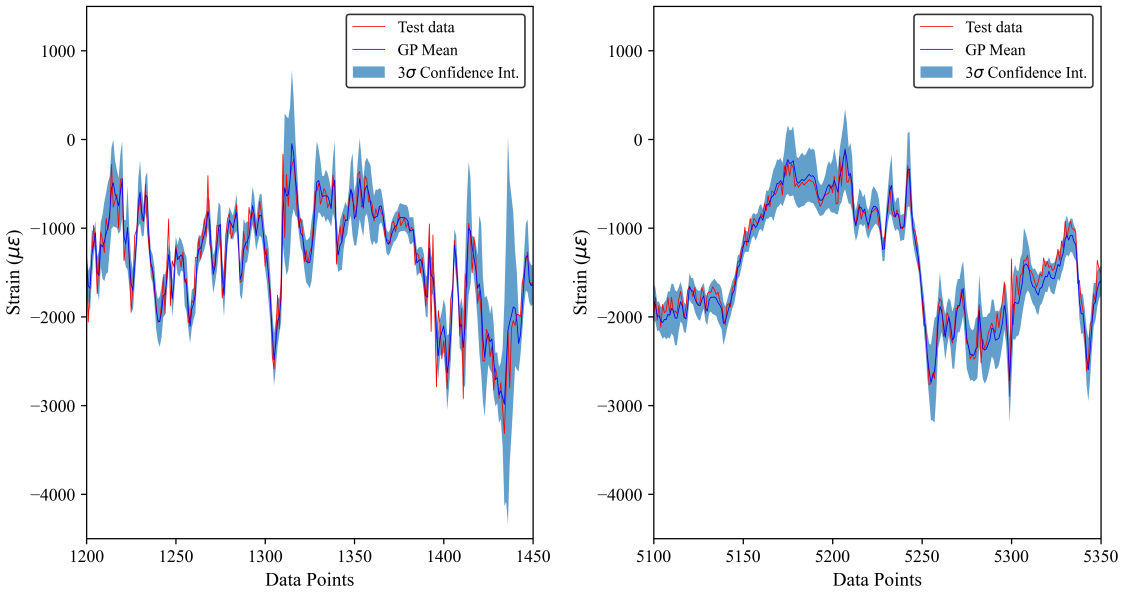


Figure 5.8: Load spectrum for Flight 66

5.2.2 Flight 77

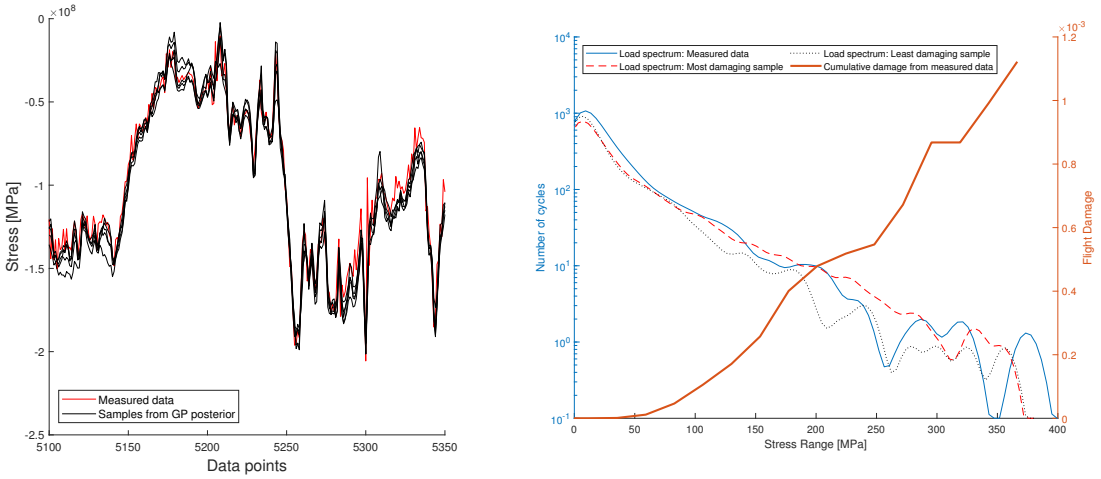
As discussed in the previous chapter, Flight 77 has the largest error of the 84 test flights in terms of fatigue damage prediction accuracy, with the GP mean prediction underestimating the measured damage by 36.2 %. While one could see from Figures 4.8 and 4.9 (reproduced below in Figure 5.9) that the measured data was within the confidence intervals for the majority of the stress-time history, by only using the GP posterior mean, this could not be beneficial. Clearly, by sampling from the posterior distribution, we can attain a better reflection of the true damage state, which is within the subsequent damage distribution (albeit at a low probability). If one was to consider the mean of the sampled damages rather than the GP posterior mean prediction, this would reduce the error from 36.2% from the GP mean to 30%.



(a) (b)

Figure 5.9: Two regions of Flight 77, each showing only 250 data points (of 10,000)

While sampling does indeed give more information about the damage state for Flight 77, the extent to which the underlying phenomena are captured by the samples drawn from the posterior is not always exemplary. For example, in Figure 5.10(a), we can see the same section of data shown Figure 5.9(b). One can see that the fluctuation around data point 5300 is still missed by the samples and GP mean alike. Similarly, in Figure 5.10(b), showing the load spectrum for Flight 77, one can see that the largest stress cycles in the measured data are underestimated by both the most and least damaging samples drawn from the model posterior. Instead, the additional damage that the most damaging cycle is capturing (compared to the GP mean) comes from an increase in medium-sized stress cycles.



(a) (b)

Figure 5.10: (a) Some samples from the posterior of Flight 77: we can see that some detail is still missed by the samples, and (b) The load spectrum for flight 77

5.3 The relationship between stress cycle range and fatigue damage

The results above demonstrate some variation in the damage distribution across the range of flights. If the prediction is accurate, it would be expected for the damage from the GP mean prediction and the measured damage be similar, which is generally seen in the results. As a trend, it could be seen that the most probable damage level of the samples (i.e. where the normalised probability equals 1 in Figure 5.4 and the peak of the histogram in Figure 5.5) is greater than the damage from the GP mean prediction and the measured damage. Furthermore, a common theme of the damage distributions shown in Figure 5.4 is that there are long tail characteristics demonstrated, with the most damaging sample being further from the most likely value, than the least likely value.

To investigate the cause of this, a small investigation into the causes of shifts in the damage distribution is performed below. In S-N fatigue, we know that the relationship between stress cycle range and number of cycles to failure is observed to be linear, but non-linear before log-scales are applied on the x and y axes. Accordingly, we would also expect a non-linear relationship between an increase in range of a given

stress cycle (stress range, σ_R) and the subsequent damage from that stress cycle, as is shown in Figure 5.12.

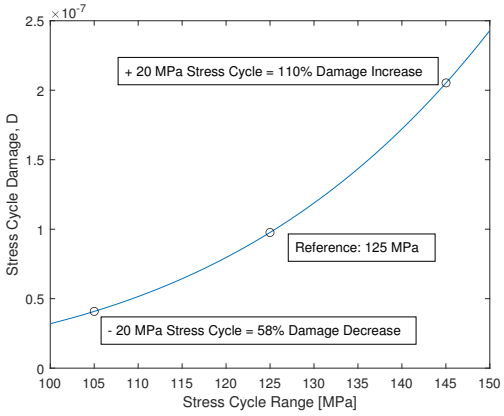


Figure 5.11: The relationship between stress cycle range and damage is nonlinear. Here, we can see that an increase of 20 MPa in stress range is significantly more damaging that an equivalent decrease

To investigate the impact of this in a sampled approach, a stress response is simulated of the form:

$$Y = Acos(\omega t) + Bsin(\omega t) \tag{5.4}$$

where A and B are independent and identically distributed (i.i.d) Gaussian random variables. This is equivalent to

$$Y = Dcos(\omega t - \phi) \tag{5.5}$$

where $D = \sqrt{A^2 + B^2}$ and $\phi \sim U(-\pi, \pi)$. In this case, D is known to follow a Rayleigh distribution with probability density function:

$$y = \frac{x}{b^2} exp\left[\frac{-x^2}{2b^2}\right] \tag{5.6}$$

For the simulated data, A and B are selected as i.i.d. Gaussian random variables with mean equal to zero and a variance of 25^2 ($A = B \sim N(\mu = 0, \sigma^2 = 25^2)$). 100,000 samples are drawn, providing the distribution of D shown in Figure 5.12.

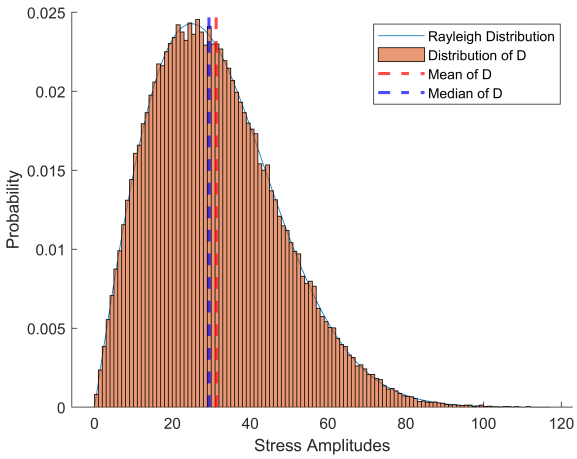


Figure 5.12: Rayleigh distribution of stress amplitudes D (Equation 5.5) for sampling investigation. 100,000 samples are drawn from the distribution

Indeed, this distribution follows a Rayleigh distribution (Equation 5.6) with $b = 25$. The samples taken from this distribution are now propagated through the fatigue assessment.

In the studies on the Tucano aircraft above, the fatigue assessment is performed with the S-N curve described in Equation 4.3. For this study, however, fatigue analysis is performed using two S-N curves: firstly a typical (log-linear) S-N curve with S_{1000}, the stress that would cause the component to fail after 1000 stress cycles equal to 250 MPa and equivalent value for 1×10^8 cycles set at 0.1 MPa. Note that, to simplify the analysis by removing an additional variable, there is no endurance limit. The second S-N curve is one in which the relationship between an increase (or decrease) in stress range is linear with the subsequent damage caused by that stress cycle. i.e. the S-D (where D = damage = 1/N) relationship is linear.

The two S-N, or S-D, curves can be seen below (Figure 5.13). The mean of D, the distribution of stress amplitudes, is 31 MPa, so the new linear S-N curve is scaled so that the damage caused by a stress cycle at 31 MPa is equal to the damage from the regular S-N curve.

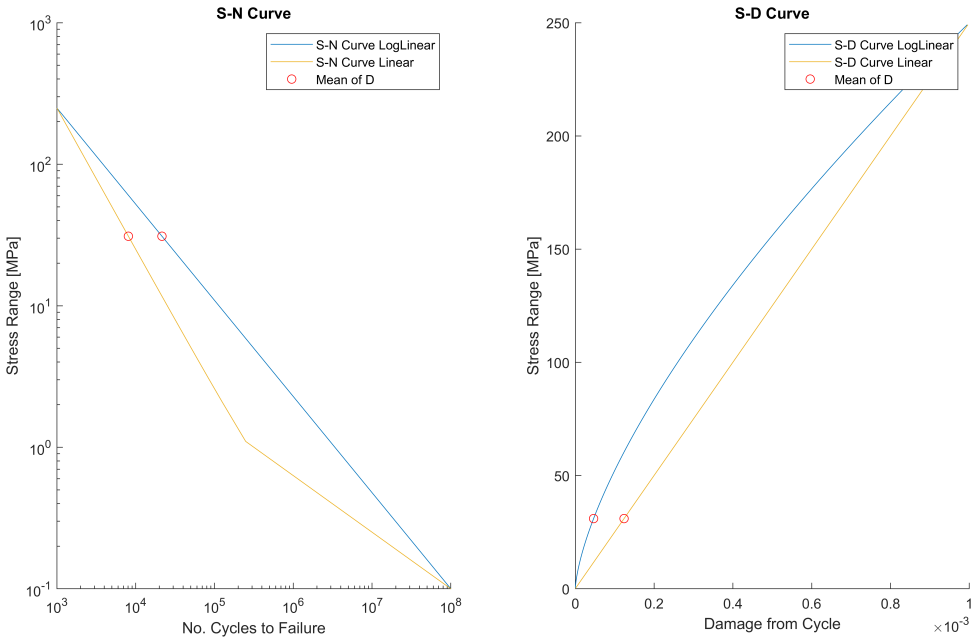


Figure 5.13: The two damage curves used for the study. Note that the axes on the S-N curve are logarithmic, while the axes on the S-D curve are linear.

For each sample drawn from D, fatigue assessment is performed using each of the S-N curves. So that the distributions can be compared easily, the resulting damage distributions are scaled so that the mean damage is equal to 1. The two damage distributions can be seen below.

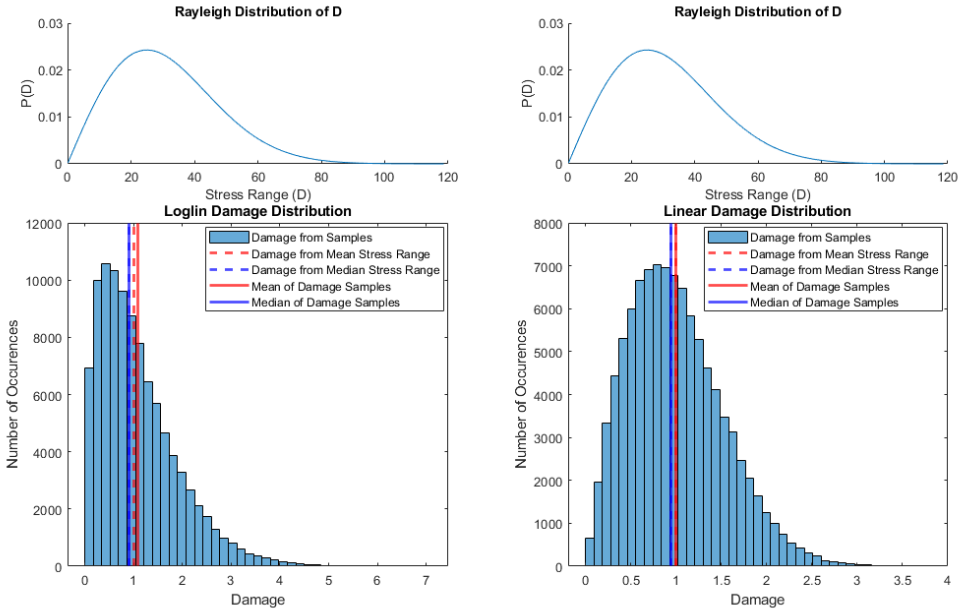


Figure 5.14: Damage distributions from each of the damage curves. The Rayleigh distribution of D is plotted above for comparison. The damages are scaled so that the mean value of distribution, D, is equal to unity.

Clearly, the impact of the non-linear damage curve is significant. The results for the linear damage curve on the right display a damage distribution very close to that of the Rayleigh distributed D. The mean and median values of the damage distribution match (almost precisely) the damage arising from the median and mean values of samples drawn from D.

For the log-linear case, however, it is immediately notable that the most damaging samples are significantly greater than the mean (6.5 times greater, in contrast to 3.8 times greater for the linear damage study). Clearly, while the distribution remains very similar between the samples (D) and their damage for the linear case, this is not the case for the non-linear study. We can see, however, that the mean and median of the damages do agree again closely with the damage arising from the mean and median of distribution, D. Clearly, then, while the log-linear nature of S-N curves does effect the damage distribution, it is not the sole cause of the shifts seen in the Tucano test flights, where the mean of the damage samples was frequently larger than the GP mean. It does however, appear to be part of the reason for the long-tail phenomena shown in the Tucano damage distributions.

5.3.1 Effect of mean stress

In the discussion of the Tucano results above, the effect of the mean of a given stress cycle on damage was shown in Figure 5.6, demonstrating that the uncertainty from the mean stress of a given cycle can easily surpass that due to uncertainty in the stress range. To further investigate the effect this has on the damage distribution, the same stress response is simulated, offset by mean, c:

$$Y = Acos(\omega t) + Bsin(\omega t) + c \tag{5.7}$$

Using the same log-linear S-N curve as above and applying the Goodman mean stress correction

$$\sigma_{a,corrected} = \frac{\sigma_a}{1 - (\frac{\sigma_m}{\sigma_{UTS}})} \tag{5.8}$$

where σ_{UTS} and σ_m are material ultimate tensile strength (UTS) and cycle mean stress, respectively. The UTS is estimated as $1.1 \times S_{1000}$ (defined above). By varying the mean stress, c, between -50 MPa (compressive) and +50 MPa (tensile), we can test to see the effect that mean stress has on the damage distribution (rather than its effect on a single stress cycle, as was shown in Figure 5.6). The results are shown in Figure 5.15, where it can be seen that the mean stress does also clearly have a large impact on the damage distribution. While the median and mean values of the damage distribution shift across the three cases (not shown, for clarity), it is also clear that the most probable damage remains close to constant, peaking around the damage level of 1×10^{-3} in each case.

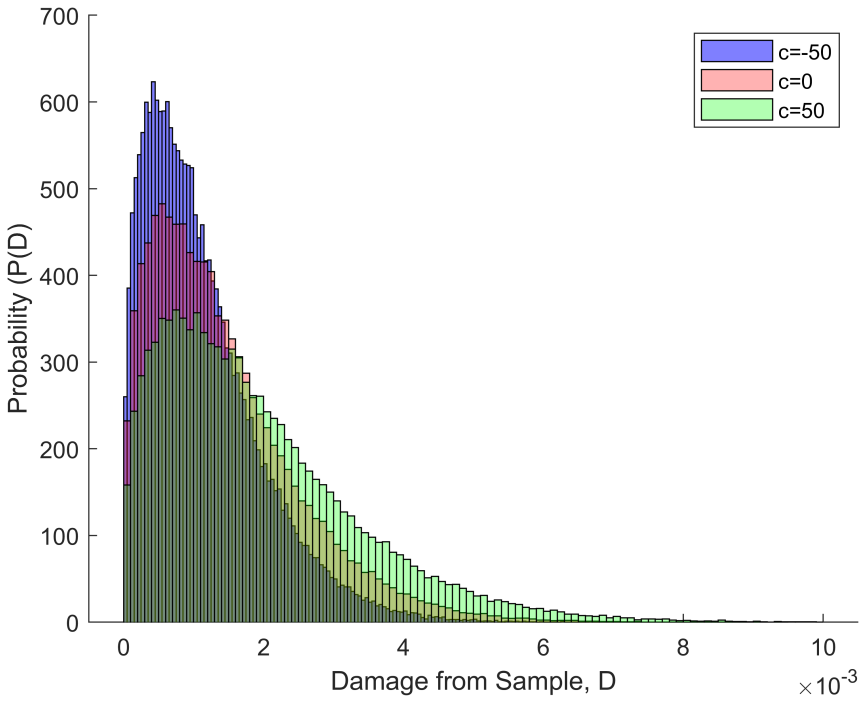


Figure 5.15: Impact of mean stress c on the damage distribution. Tensile mean stresses are known to result in faster failure, and this is born out in the damage distribution, where it can be seen that there are more damaging samples visible in the tensile mean stress histogram. However, the peaks of the three histograms are centred around a similar damage level.

While this study has demonstrated that the non-linear nature of the S-N relationship can (a) significantly impact the damage distribution compared to the distribution stress ranges and (b) further exaggerate the impact of non-zero mean stresses, it does not fully explain the changes in distributions that have been visible throughout this chapter. For example, in Figure 5.5, the GP mean and measured damage agree closely, but the peak of the histogram is more damaging than this value, while throughout this study, it has been consistently less damaging.

5.3.2 Effect of model hyperparameters and distance from training data

The final variables that will be investigated here are those of the model hyperparameters (and, more specifically, the lengthscale) and also the distance between training and testing points. Again, a sinusoidal stress response again is simulated: $Y = 100sin(\omega t)$

(this is equivalent to the stress response sampled from previously (Equation 5.5), with D a constant equal to 100 MPa, rather than a Rayleigh-distributed random variable). 10 fully-reversed stress cycles are again present in the signal. Again, the log-linear S-N curve shown in Figure 5.13 is used for the fatigue assessment here.

Using the log-linear S-N curve used above, two situations are tested: In the first, the sensitivity of the damage distribution to lengthscale is investigated. In the second, the effect of varying the training interval is tested. For the first case, a GPR model is trained on every 15 points in the stress response. Three lengthscales are tested, with the signal and noise hyperparameter kept constant. The results are presented in Figure 5.16 below, with lengthscales, from left to right ($[\sigma_f, \ell, \sigma_n]$): [1,0.5,0.1],[1,1,0.1], [1,2,0.1]. Clearly, as the model lengthscale decreases, there is a greater fluctuations in the posterior samples (note: the GP mean also has more fluctuations in the shorter lengthscale case, but to a significantly lesser extent than the samples drawn from the posterior). The impact on the damage distribution is severe: For the short lengthscale scenario, despite a reasonable GP mean prediction (owing to a simple case study), even the least damaging samples are more damaging than the simulated data. As the lengthscale increases, we see a dramatic shift of the damage distribution to the left (i.e. closer to the GP mean and measured damages). In this case, as the data is narrowband, the increasingly long lengthscale brings the samples closer to the simulated response. However, as will be shown in the coming chapters, if the stress is wideband, this effect will result in a severe underestimation of the true damage level.

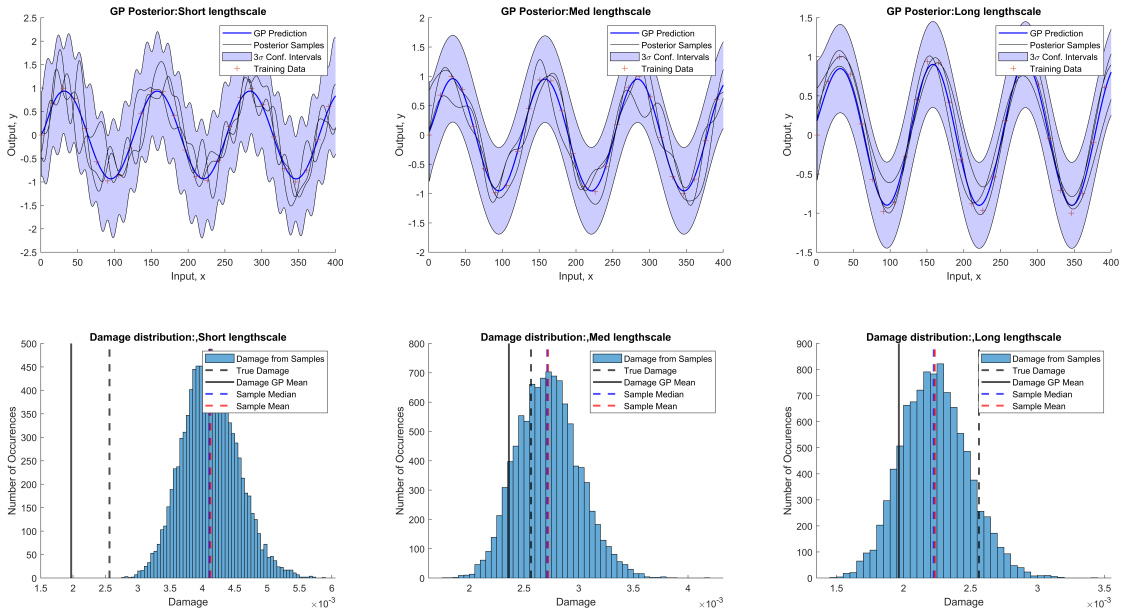


Figure 5.16: Impact of hyperparameter on lengthscale. From left to right, the hyperparameters are $[\sigma_f, \ell, \sigma_n]$:[1,0.5,0.1],[1,1,0.1], [1,2,0.1]

A similar trend is visible in the study on training interval, the results for which are shown in Figure 5.17. This time, the hyperparameters are kept constant (and the same as the short lengthscale example above, Figure 5.16). In this case, we can see that varying the training interval has a similar effect that adjusting the lengthscale had in the previous case. Both the sampled damages and the mean prediction are reasonable for the highest sampling rate (with training indices every 5 data points), but as the training points become increasingly distant, this provides the opportunity for increasing fluctuations in the prediction. Again, this effect applies to the GP posterior mean, but to a lesser extent than the posterior samples drawn. In both cases (varying the hyperparameters, or varying training distance) the results here highlight the importance of having training data that is representative of the test cases.

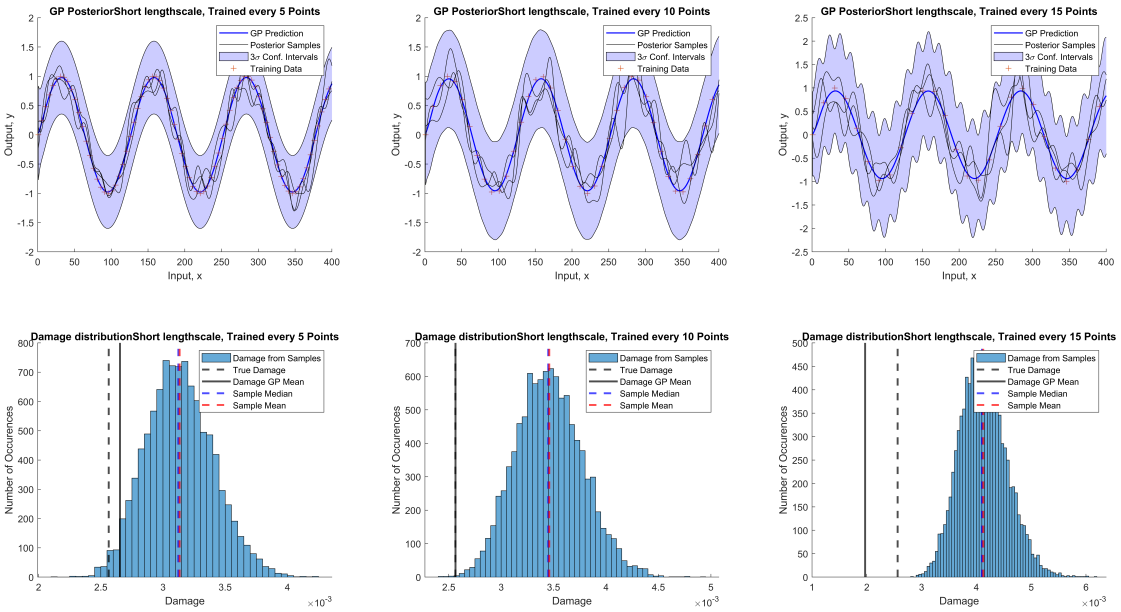


Figure 5.17: Impact of training interval on samples drawn from posterior. In this case, the model hyperparameters are kept constant across the three models ($[\sigma_f = 1, \ell = 0.5, \sigma_n = 0.1]$). Instead, the training interval of the GP is varied, with sampling points every 5 (left hand image), 10 (middle) and 15 (right) points used.

5.4 Discussion

The proposed methodology provides a probabilistic prediction of strain and fatigue accrual at unmeasured locations on a structure under an unknown loading. Having shown in the previous chapter the ability to predict strain using GP regression across the range of flights, it has been demonstrated that, by propagating uncertainty in the model, errors (when the model is reasonably well-performing) can be mitigated. The resulting distributions of fatigue can shed light on the range of potential damage accrued, with the measured damage state captured within the distribution across all of the flights. One natural question is, therefore, how to actually *use* the resulting distribution. Ultimately, a structure's usage will dictate this, and a risk-based maintenance and inspection strategy can be formed accordingly [52]. The original design methodology of the structure, i.e. Safe-Life or Fail-Safe, will influence this, alongside other factors including the level of uncertainty of the prediction.

The central argument of this chapter is that predicting fatigue accumulation indirectly using GP regression as a virtual strain sensor should be considered as a distribution

rather than a deterministic value. By using only the posterior mean to predict fatigue damage, one of the benefits of using GP regression - its inherent probabilistic nature - is lost. Indeed, it could be argued that a single, deterministic, value for fatigue damage is not truly representative of the posterior distribution of the model. We can see from Figures 5.4(a) and 5.4(b) that the probability of detection of the measured damage state for some flights is low. The results from flight 77 (Figure 5.9) provide an excellent example of the usefulness of the approach taken here.

While the applied usage of the damage distribution would depend on the individual structure and the potential cost of failure, if one was employing a deterministic strain prediction for this particular flight (Flight 77), or if one were simply relying on the GP mean, the predicted fatigue would be very misleading. The predicted fatigue distribution from the uncertainty propagation approach, however, incorporates the true damage accumulated (albeit at a low probability). Any following risk assessment employing the distribution would be much more robust to the error in the (mean) prediction of the strain. Regardless of modelling choices, kernel selection, and hyperparameter optimisation, in a data-driven environment, accurately predicting the damage using the mean prediction would be unlikely for all scenarios. Therefore, utilising the proposed methodology of sampling to develop damage distributions provides a more robust framework that can be utilised differently based on the nature of the structure.

One can extract key features of interest from the predicted fatigue distributions. For example, the method provides a principled means of extracting the least and most damaging strain time histories possible for a given flight that may be used to inform an (e.g. interval-based) risk assessment. An alternative means of predicting e.g. the most damaging strain-time history possible for a given flight could be to sample a function that passes through the maxima/minima of the posterior GP mean plus/minus the confidence interval. An example of such a function $\pm 2\sigma$ is shown in Figure 5.18. Such a naive approach neglects the joint Gaussian assumption the model is based on and produces a fatigue prediction that is a 350% over-prediction of the measured value.

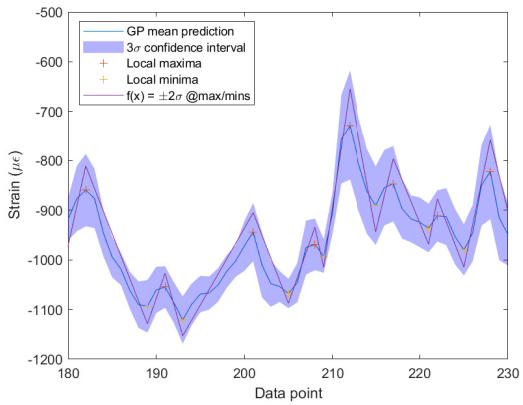


Figure 5.18: Naively sampling from the confidence intervals at local maxima and minima,

Whilst the author advocates for exploiting the probabilistic nature of the GP model, or any probabilistic model, care must be taken in how uncertainty assessments are interpreted. The uncertainty in the GP prediction is based on a measure of the distance between input training and testing points. The inputs to the GP are specified by the user and the distance measure is dictated by the covariance function. This means that the predictive distributions of fatigue shown here are dependent on the selection of both of these things. So long as the strain prediction quality is acceptable, the selection of the covariance function should not be of large consequence.

However, care must be taken in the selection of the model inputs. For example, if an irrelevant input were included that exhibited different behaviour between training and testing sets, this would increase the uncertainty of the fatigue damage assessment in an unrepresentative manner. Here, the use of Automatic Relevance Determination (ARD) covariance function [129] is advocated, with a separate lengthscale for each input to mitigate the effect of a potentially misspecified input variable. Finally, it is important to note that, as with all data-driven methods, prediction quality and reliability of the confidence bounds will decrease in extrapolation. If the relationships between inputs and targets change outside of the training data set, the inference here will not be valid. The importance of a representative training set cannot be understated. Clearly, this is particularly important when there is a significant dynamic component to the stress response. The sampling frequency of the input data (and the training target data) must be sufficiently high to be able to accurately capture the stress signal, otherwise the shifts in the damage distributions, seen here, will be likely. Similarly, the use of a sufficiently high sample rate will make the

adoption of an appropriate lengthscale more likely. This will be covered in more detail in the following chapter.

5.5 Conclusions

In this chapter, uncertainty in the strain prediction model using the Tucano dataset is quantified through the Gaussian Process framework and propagated through to the final fatigue assessment. While other studies have considered probabilistic data-driven models for fatigue, this is generally limited to the DEL approach (which simplifies the stress-time history into a single repeated load). This is the first work to consider the uncertainty of the full stress-time history and propagate the uncertainty of the prediction throughout the fatigue assessment. By propagating the uncertainty of the model prediction, we can achieve an improved reliance upon the predicted damage and a better understanding of the full model prediction.

By doing this, it can be seen the distribution is not always normally distributed and centred around the mean prediction. By investigating what causes the distribution to shift away from the posterior mean prediction, it is shown that the the damage distribution is significantly impacted by the non-linear nature of the relationship between stress and damage (i.e. the non-linearity of S-N curves) and also, the representativeness of the training data to the test data.

The focus of the following chapters will be on how model accuracy can be improved for both strain prediction and probabilistic fatigue damage prediction by integrating knowledge of physics into the Gaussian process model. Motivated by the fact that dynamic behaviour can have significant contribution to fatigue accumulation, a covariance function derived from a single-degree-of-freedom oscillator will be used for strain prediction. Again, the impact of propagating model uncertainty will be investigated.

Chapter 6

Introducing Physics

In the previous chapters, the use of a data-driven model for predicting strain on a Tucano aircraft wing was demonstrated. Under quasi-static loading, the model performed well. At the available sampling rate, however, predicting the strain behaviour could not be achieved satisfactorily. In terms of assessing the fatigue damage accrued during these flights, developing a probabilistic prediction by propagating the uncertainty of the model was able to mitigate the model prediction error to a degree: across the range of flights, the measured damage was captured by the resulting damage distribution.

In this chapter, we will focus on the problem of predicting dynamic behaviour. Specifically, this will look at the challenge of upsampling time-series data to accurately predict fatigue damage accumulation. The motivation for this comes from the high sampling rates generally considered necessary for predicting fatigue damage. In [130], it was found that for an offshore wind turbine, an effective sampling rate of 10 times the last significant frequency was required for predicting fatigue damage accurately. In Chapters 4 and 5, it was the high-frequency data that was particularly difficult to capture. In [95], a 'mathematical network' (see Chapter 2 for more information) is used for predicting the fatigue damage accumulation on a helicopter: despite good results across the majority of data available, accurate fatigue damage prediction could not be achieved during turbulent behaviour due to insufficient sample rates of the training data for vibration-induced fatigue. In this chapter, we will tackle this problem by using a covariance function derived from an single-degree-of-freedom (SDOF) equation of motion under zero mean white noise loading, described further

in the next section.

6.1 Kernel design

As discussed previously, the flexibility of a GP arises from the fact that we can choose an appropriate covariance function for the model, specifying the functional form that the model posterior will take. Figure 6.1 (first presented in Chapter 3), shows some draws from common kernels.

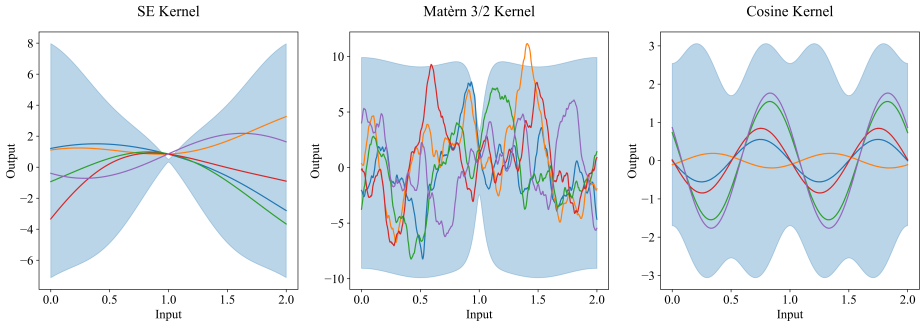


Figure 6.1: The covariance function (and its optimised hyperparameters) define the functional form of the GP posterior. The squared exponential kernel is the most commonly used, but there are a wide range of kernels that can be selected based on prior knowledge of the data. Furthermore, kernels can be combined to provide further flexibility [114]

While typically, data-driven models will use an 'off the shelf' kernel, we can also design kernels to suit our needs. It is possible to embed our physical knowledge of a system into the kernel itself, the periodic kernel is an example of this: as periodic responses are known to exist in many areas, such as robotics [105], this fact can be harnessed to improve the response prediction significantly [105]. The reader is referred to [114] for a review of kernel basics and common forms.

Following [106], this principle is used to develop a kernel for single-degree-of-freedom (SDOF) oscillating systems under zero-mean Gaussian white noise forcing. Here, we will utilise the fact that we can calculate the autocovariance of a process as:

$$k(t_1, t_2) = \mathbb{E}\left[(y(t_1)) - \mu(t_1))\,(y(t_2)) - \mu(t_2))\right] \tag{6.1}$$

where $y(t_1, t_2)$ and $\mu(t_1, t_2)$ are realisations of the process and its mean at two arbitrary time points, respectively. The autocovariance is equivalent to the autocorrelation when the process has a zero mean.

For an externally forced, linear single-degree-of-freedom, dynamic process ($F(t)$) with mass (m), stiffness (k) and damping (c):

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) = F(t) \tag{6.2}$$

By calculating the convolution of the excitation and the impulse response function ($h(t)$), we can predict the response of the system:

$$Y(t) = \int_{-\infty}^{\infty} F(t-r)h(r) \tag{6.3}$$

By assuming external forcing as stochastic, assumed as Gaussian white noise, we can calculate the autocovariance of the process to be:

$$\phi_{Y(t)} = \frac{\sigma^2}{4m^2\zeta\omega_n^3} e^{-\zeta\omega_n|\tau|} \left(\cos(\omega_d\tau) + \frac{\zeta\omega_n}{\omega_d} \sin(\omega_d|\tau|) \right) \tag{6.4}$$

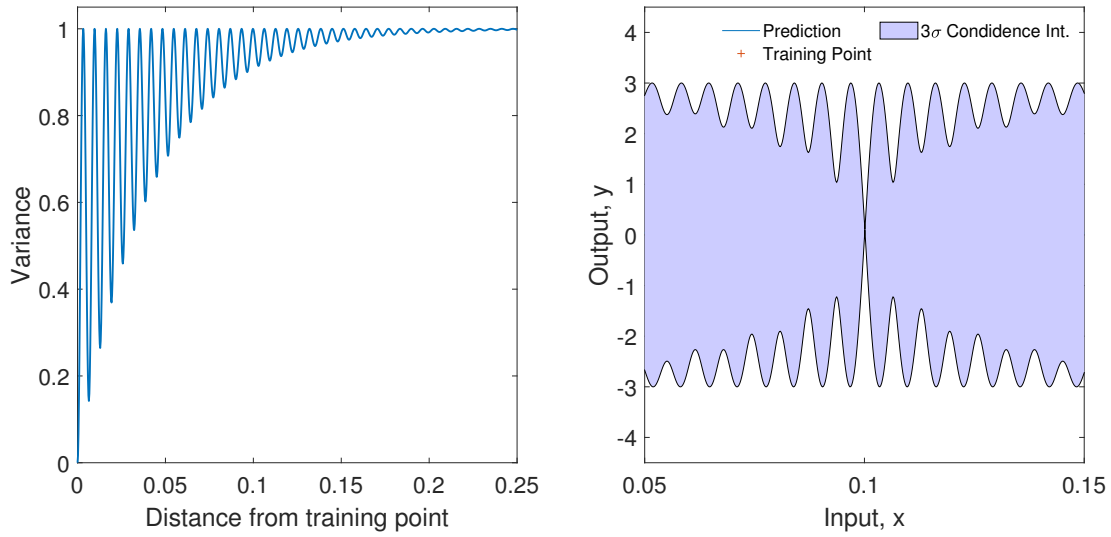
The full derivation of the kernel (Equation 6.4) can be found in [106] and the reader is also referred to [131] and [132] for further information. The use of the SDOF covariance function here is a novel way of bringing standard engineering practices to the typically data-driven machine learning environment. While the hyperparameters of data-driven kernels are values that can be optimised to achieve the best *fit* of the data, the hyperparameters for this physically-derived covariance function are the actual physical parameters of the system: mass (m), damping ratio (ζ), natural and damped frequencies (ω_n and ω_d, respectively). τ is the difference in time in seconds between the two points of interest. The magnitude is scaled by an amplitude term which is a function of the variance of the forcing input as well as these physical parameters of the system.

6.1.1 Interpretation of Model Variance

Interpreting the confidence intervals from physics derived kernels

As the damage distributions presented throughout this work are developed by propagating the model posterior variance, it is worth considering how the choice of kernel affects this. In Chapter 3, it was shown that for a typical, data-driven kernel (an SE in this case, Figure 3.12), the model hyperparameters will determine the rate at which the GP posterior tends towards its maximum variance, which is dictated by the signal variance.

Figure 6.2(a) shows an equivalent plot for the SDOF kernel where we can see that the variance does not simply grow to its maximum with increasing distance from the training location but fluctuates as a result of the oscillatory terms in the covariance function. Figure 6.2(b) shows an example posterior mean and confidence using this kernel conditioned on a single point at x = 0.1. While we will typically use the term confidence intervals for this 3σ posterior limit, it is now useful to remind ourselves that the physical representation of this is related to the family of functions that fit the data that we have conditioned our model by. In the case of the SDOF kernel, there is now a *physical* interpretation to these intervals, they now represent the actual variance of the process under a Gaussian random load.



(a) (b)

Figure 6.2: (a) GP posterior variance with distance from training point with SDOF covariance function and (b): Resulting confidence intervals.

Thus, now when samples are drawn from the posterior confidence using the SDOF covariance function, they will maintain this physical significance.

6.2 Simulated case study

The output of the simulation is displacement, used as a proxy for strain in this chapter. Thus, the simulated displacement data is multiplied by a scalar, such that the maximum stress range in the simulated data is 150 MPa.

A histogram of the resulting stress ranges can be seen in Figure 6.3. Three models are set up for comparison (6A-6C, described below), with this simulated stress-time history used as the target in each case.

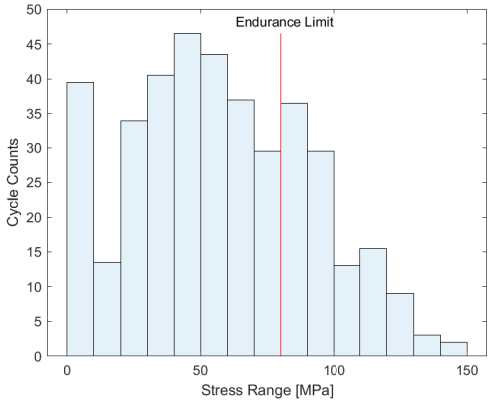


Figure 6.3: Stress ranges of the simulated data set. We can see that the majority of stress ranges are below the endurance limit of the S-N curve

6.2.1 S-N Curve

After using a bespoke S-N curve suitable for the Tucano data, the work in this chapter (and the following) utilises a generic S-N curve, with a slope, m, equal to 5 and an endurance limit of 79 MPa, Figure 6.4. The number of cycles in the simulation data above the endurance limit can be seen in Figure 6.3.

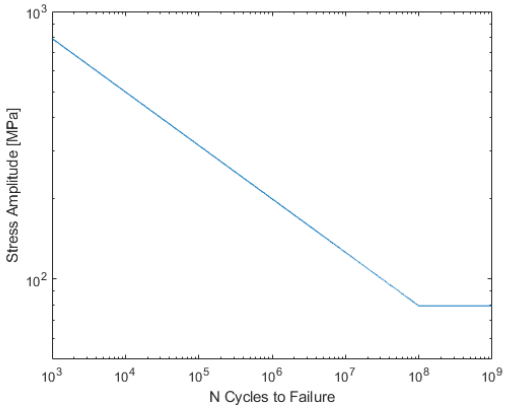


Figure 6.4: S-N curve used for chapters 6 and 7.

6.2.2 GP inputs and targets

Three models are compared for this case study, described in Table 6.1. For the SDOF kernel, time is used as an input (s) (Equation 6.4). Two black-box models will be

compared, one using acceleration as an input to the GP (as this is a typical input, in practice in a virtual sensing environment) and one using time (to be directly comparable to the SDOF case).

Model	**Description**	**Inputs**
6A	SE (black-box) kernel	acceleration
6B	SE (black-box) kernel	time (s) (ms^{-2})
6C	SDOF kernel	time (s)

Table 6.1: Three models for simulated case study

The natural frequency of the system is 45 Hz, making the Nyquist frequency [133] (the minimum sampling frequency at which reconstruction of a signal is possible, equal to two times the frequency of the signal), 90 Hz, while the data is generated at 1024 Hz. For the initial study, the training inputs to the GP will be provided at 128 Hz, i.e. 1 in 8 data points, sampled evenly across the data.

6.2.3 Model assessment

In the previous sections, NMSE has been used to assess the model performance (Equation 4.4). Clearly, this only assesses the performance of the GP mean prediction, while this work shall be continuing to use the full posterior prediction for the fatigue assessment. Furthermore, it was clear that the NMSE did not correlate well to accurate fatigue damage prediction, even from a deterministic perspective. As a result, in this chapter, a second metric of mean standardised log loss (MSLL) shall also be used to assess the model. Rather than simply using the posterior mean, this assesses the predictive *distribution*:

$$\text{MSLL} = \frac{1}{n_*} \sum \frac{1}{2} \log(2\pi\sigma_*^2) + \frac{(\mathbf{y}_* - \bar{\mathbf{y}})^{\mathbf{2}}}{2\sigma_*^2} \tag{6.5}$$

where σ_* is the posterior variance. Lower values here indicate a better predictive distribution, with a reasonable prediction achieved with an MSLL of approximately zero. Increasingly negative values indicate an increasingly good prediction [112].

6.2.4 Black-box model

First, we will look at model (6A), using acceleration as the model input. In Figure 6.5, we can see that there is a reasonable correlation between the input and target data (the Pearson correlation coefficient for the two variables is 0.81), albeit with a reasonably large variance in the data.

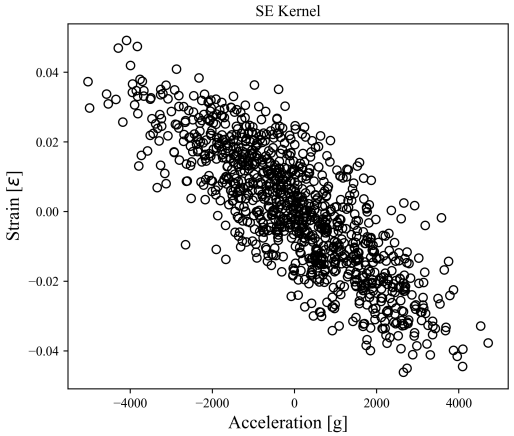
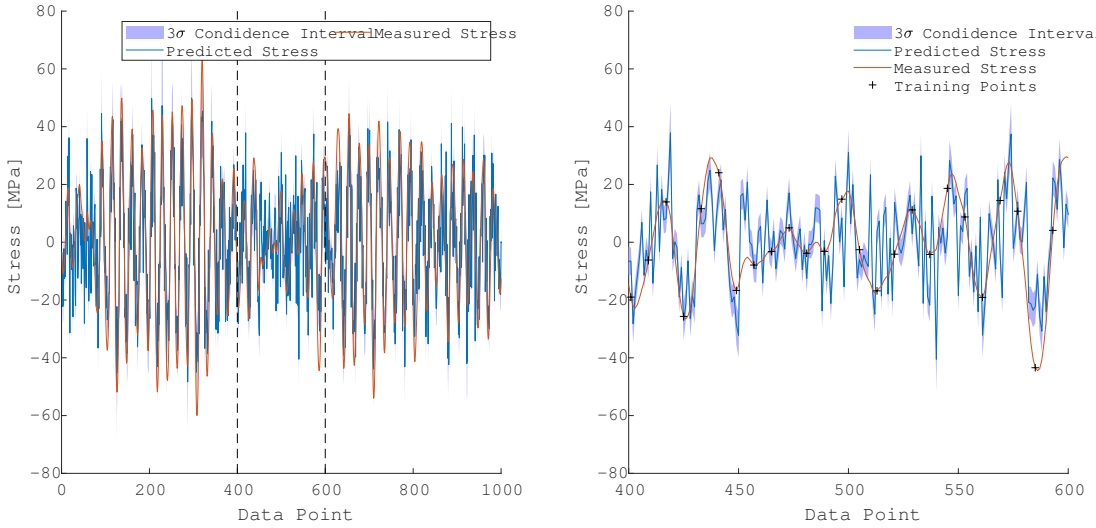


Figure 6.5: Acceleration vs strain for the simulated data

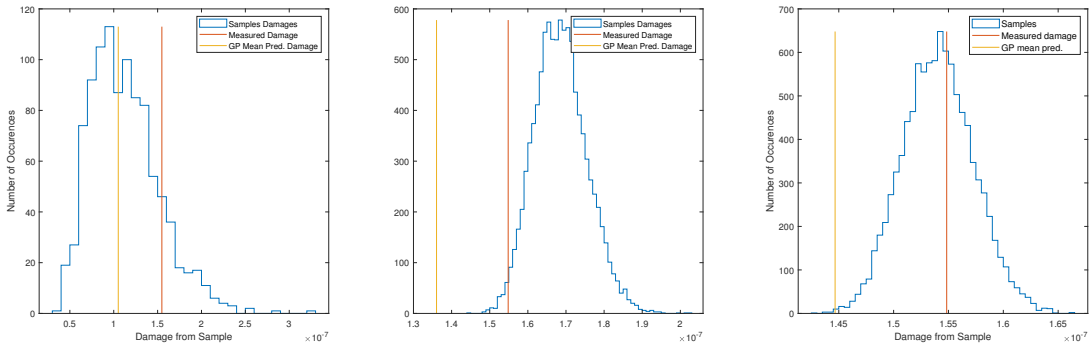
The GP posterior can be seen in Figures 6.6 (a) and (b). The NMSE and MSLL for this prediction are 30% and 25.2, respectively, indicating that the prediction is not accurate. This can be confirmed visually from the GP posterior shown in Figures 6.6 (a) and (b). Despite the good correlation between the input and the target data, it is clear that the GP is overfitting compared to the test data. Furthermore, the posterior shows excessive confidence in this wrong prediction (evidently, in a probabilistic context this is the worst eventuality).



(a) (b)

Figure 6.6: (a) GP Posterior for black-box model using acceleration as an input. (b) Enhanced view of (a).

The resulting damage distribution for Model 6A can be seen in Figure 6.7 (a). Unsurprisingly, given the poor prediction shown in Figure 6.6, the fatigue damage prediction is poor. We can see that both the GP mean and the median and mean of the posterior sampled damage all underestimate the true damage state. Sampling from the posterior helps here, as the actual damage state is captured within the damage distribution. However, the variance of the sampled damages is very large. In this case, this is useful, as the large damage variance is indeed representative of the fact that the model is not able to predict the stress response accurately.

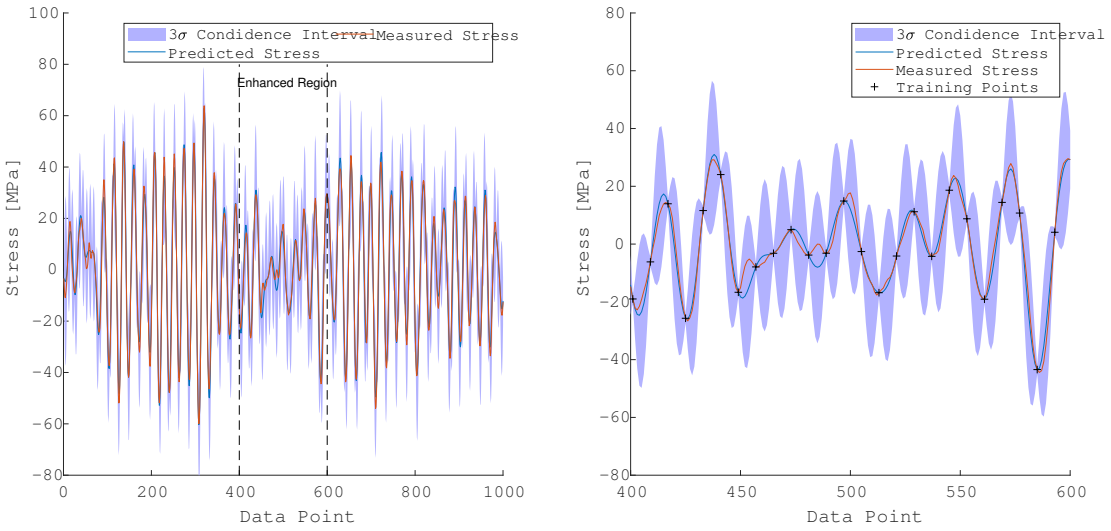


(a) (b) (c)

Figure 6.7: Damage distributions for the three models: (a) Black-box model - Acceleration input (b) Black-box model - Time input (c) Grey-box model using SDOF kernel

6.2.5 Black-box - Time input

In model (6B), we are using an SE kernel with time used as the model input: this is directly comparable to the grey-box model (the same training indices are used throughout the three models).



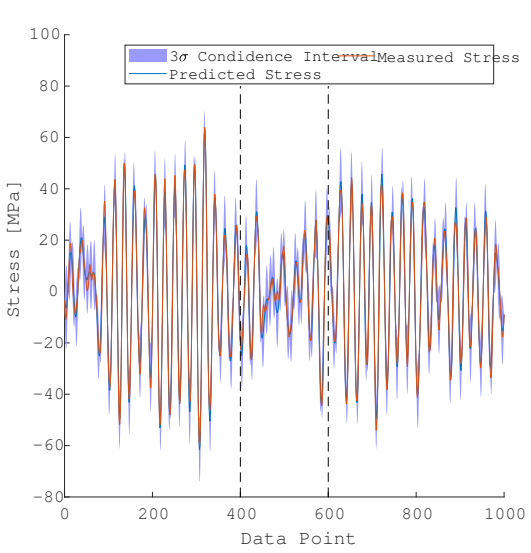
(a) (b)

Figure 6.8: (a) GP Posterior for the black-box model using time as an input (b) Enhanced view of (a)

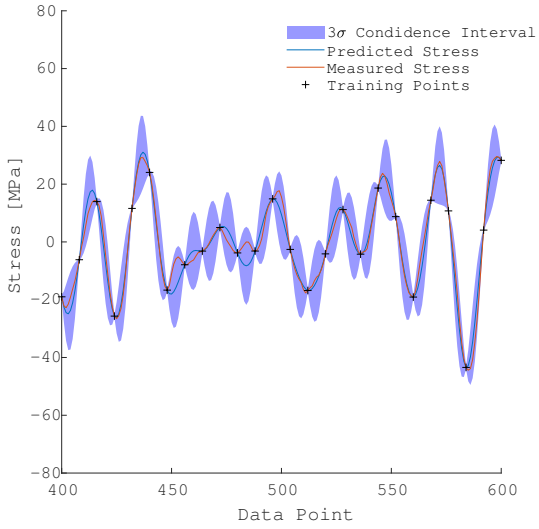
Clearly (Figure 6.8), we can see that the prediction is significantly improved compared to Model (6A). The mean prediction follows the test data and the underlying phenomena is captured well at this (relatively high) sampling rate. The MSE and MSLL are 1.7% and -2.11, respectively, reflecting this improvement compared to Model (6A). The improvement in prediction is reflected in the damage distribution which is shown in Figure 6.7(b) where there is a significant reduction in damage variance, with both the GP mean and the samples median agreeing with the 'true' damage value much more closely. The posterior mean still underestimates the true damage by 12.1%, however, this is reduced from 32.1% for Model (6A).

6.2.6 Grey-box - SDOF kernel

In Model (6C), the SDOF kernel is used as the GP covariance function. As per Model (6B), time is used as the model input with the training indices the same as in the prior models. As it is reasonable to assume that the physical properties of a system can be estimated to a reasonable level, the search parameters of the (physically representative) hyperparameters are limited to ± 20% of the simulation values. The resulting GP posterior is shown in Figures 6.9(a) and (b).



(a) First 1000 data points



(b) Enhanced view (as per black-box model)

Figure 6.9: GP posterior distributions using the SDOF kernel on the simulated data set.

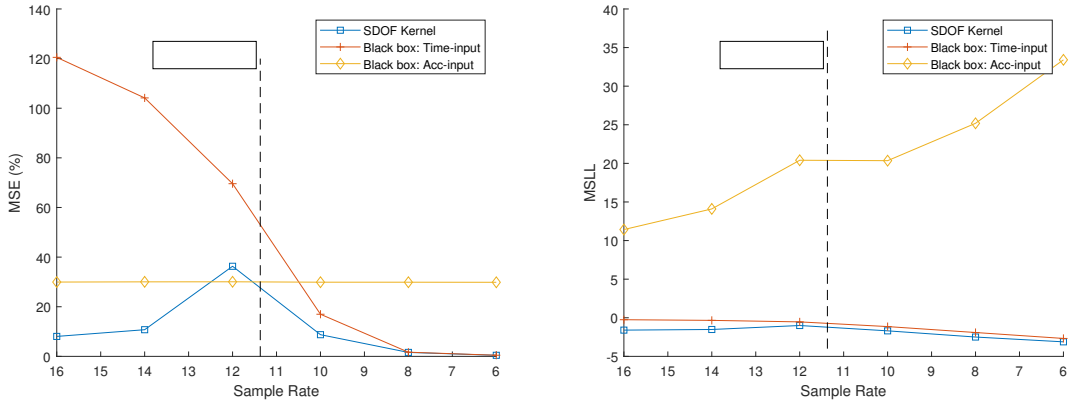
Here, we can again see a very large improvement compared to Model 6A and a significant improvement on Model 6B: the NMSE is reduced to 0.4% and the MSLL is -3.1, confirming this improvement in prediction. Away from the training data, we can see an improved model confidence (which is reflected by the reduced MSLL metric). While the results in terms of MSLL and MSE show an improvement with the black-box models, the largest improvement is shown in the damage distribution shown in Figure 6.7(c). As per the first two models, the posterior mean underestimates the actual damage state (in Figure 6.9(b), we can see that some smaller stress cycles are indeed missed by the posterior mean). However, the magnitude of the error is significantly reduced in Model 6C, underestimating the true damage by 6.6% (while the equivalent values for Models 6A and 6B are 32.1% and 12.1%, respectively).

The median of the GP samples for this model predicts the true damage state accurately as a result of the draws from the posterior taking the form of the physical process, as discussed earlier. Crucially, the variance of the damage distribution is significantly reduced with the most damaging sample only being 10% greater than the least damaging sample, while in Model B, this value was 43%. The overall improvement in performance when using the SDOF kernel is significant.

6.2.7 Reducing sample rate

The results above demonstrate the performance of the three models at an effective sample rate of 128 Hz (while the Nyquist rate is 90 Hz). While the results of the SDOF kernel have already shown an impressive capability to upsample the data to a sufficient sample rate to predict fatigue damage (while the sample rate used for the above study is above the Nyquist rate, it is still significantly below the required sample rates cited in the above studies), we will now test the performance at even lower sample rates. Each of the models is tested with sample rates between 1/16 and 1/6 data points (representing an effective sampling rate of 64 Hz and 171 Hz, respectively). Figures 6.10(a) and (b) show the impact that sample rate has on the MSE and MSLL values for the three models. As per the fatigue damage distributions, the grey-box model clearly outperforms the black-box models when the sample rate is reduced, with MSE values of 10% or below across the range of sample rates tested, except for an outlier just below the Nyquist frequency.

The corresponding Fatigue prediction is shown in Figures 6.11a - 7.15b which show

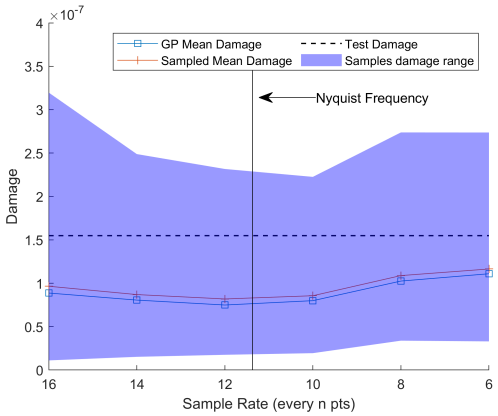


(a) MSE (b) MSLL

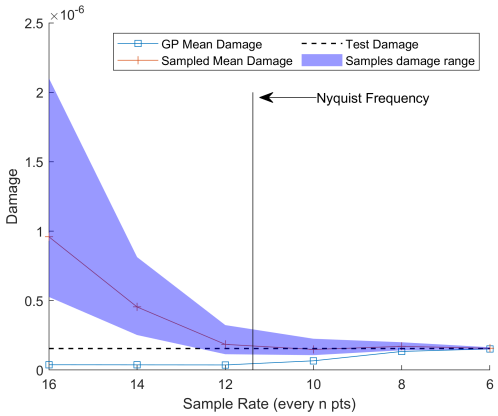
Figure 6.10: Progression of Mean Square Error (MSE) and Mean Standardised Log Loss (MSLL) with sample rate.

the damage progression for each of the models across the sample rates. The range of damages arising from the samples is also indicated here. At 128 Hz, the performance of the SDOF kernel was a significant improvement on the black-box models and when reducing the sample rate, the impact of the physics-informed kernel becomes increasingly significant. The performance of the acceleration-input black-box model (Model 6A, Figure 6.11a) is steady over the range of sample rates, however, even at 128 Hz, this model was not able to predict the strain accurately and the subsequent damage prediction was also unacceptable. The grey-box, prediction, on the other hand, performs well across the range of sampling frequencies from a fatigue damage perspective. The utility of developing a probabilistic prediction of damage is already evident from Figure 6.7(c), where we saw the median sample of the distribution matching the measured damage more accurately than the posterior mean. This is reinforced by the results shown in Figure 6.11c, where the samples drawn from the posterior indeed provide a better understanding of the damage state across the range of sampling rates.

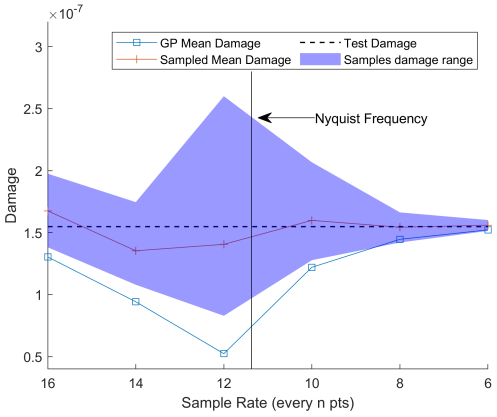
The most notable decrease in performance, from both a MSE and damage perspective, is when using the black-box Model 6B at lower sampling rates. While at 128 Hz, the sample rate was sufficient for the model to be able to reconstitute the signal of this (relatively simple) case study, this is not true as you tend towards, and below, the Nyquist frequency. As the sample rate drops, not only does the GP mean struggle and the model confidence decrease, but we can also see that there ceases to be a utility of the damage distribution, with the measured damage no longer captured by



(a) Black-box



(b) Black-box time



(c) SDOF Kernel

Figure 6.11: Impact of reducing the sampling rate on Fatigue damage prediction distributions - it can be seen that just below the Nyquist rate, the SDOF kernel struggles. Elsewhere, a clear improvement can be seen using the physics inspired kernel.

the distribution. The posterior prediction for Model 6B is shown in Figure 6.12: the NMSE at this sample rate is 120%, the worst of all of the models.

The reason for this improved prediction can be seen in Figure 6.13, we can see the grey-box (Model 6C) posterior prediction at the lowest sampling rate of 64 Hz: even at this low sampling rate, the model is following the functional form of the oscillator, albeit without accurately predicting the amplitude of the stress oscillations (and having a low confidence to reflect this).

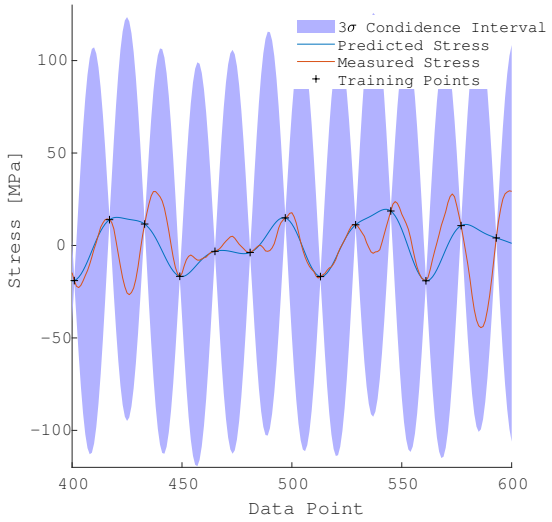


Figure 6.12: Black-box model (time input) with sampling rate of 64 Hz.

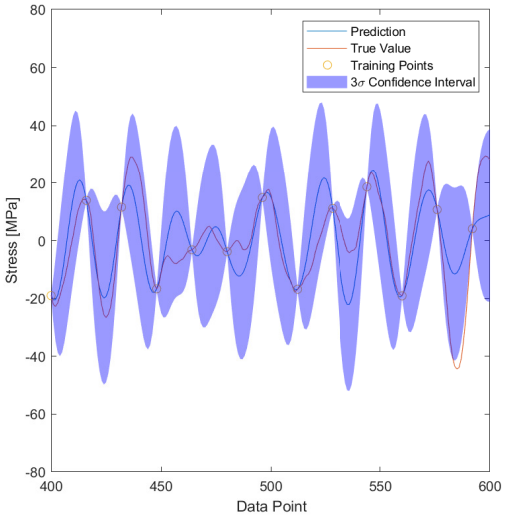


Figure 6.13: SDOF model with sampling rate of 1/16 data points, well below the Nyquist rate of 1/12 points. Despite the low sample rate, the model still follows the functional form of an SDOF oscillator and therefore, a reasonable prediction is achieved.

While the time-input black-box model (Model 6B) performed satisfactorily at 128 Hz, reducing the sampling rate evidently has a large impact on the ability of the model for damage prediction. As per the grey-box model, the MSLL value is below zero across the range of sample rates, however, the posterior prediction, shown in Figure 6.12, shows that the prediction is, as one would expect, smoothing through

the data rather than following the oscillations of the simulated data. It can be noted that despite the poor prediction, the MSLL values for this model indicate a good prediction as the true value is captured within the confidence intervals of the posterior. This comes at a cost of the variance growing very severely away from the training points. As this results in samples drawn from the posterior having much larger amplitudes than the simulated data, the result of this is a very conservative damage prediction, with the most damaging samples from this model being over 10 times greater than the true value at 64 Hz. Conversely, for the grey-box model, the most damaging sample at the same sample rate (64 Hz) is 1.4 times the measured value.

6.3 GARTEUR aircraft case study

The performance of the SDOF kernel is now also tested using measured data from laboratory testing on the Garteur aircraft. The data is introduced in Chapter 3 following [119, 120, 134]. For the testing conducted at the Laboratory for Verification and Validation in Sheffield, in addition to the accelerometers are installed across the aircraft, as per the benchmark study, 15 strain gauges are also installed across the starboard wing. The structure is excited using Gaussian white noise between 4 and 80 Hz using a shaker installed at the port-side wing tip.

As per the Tucano work in the previous chapters, the target of the GP will be the strain at the root of the wing, where the stress is greatest, and, again, two black-box models will be used in comparison to the grey-box model. Again, black-box models a using squared-exponential kernel (with inputs time and acceleration, respectively) are compared to the SDOF kernel. As the benchmark structure was not originally intended to accrue fatigue, the stresses present are not sufficiently large for a typical S-N curve. Therefore, the data is scaled by a factor of 20 to make the data usable for fatigue damage prediction using the same S-N curve used for the simulation above.

The total length of the test data is over 2×10^6 data points. In order to draw samples from the GP, inverting the full posterior covariance matrix is required and, therefore, the number of data points that we will assess is limited to 4×10^4 in this work (indicated in Figure 6.14). Figure 6.14 also shows a region used for estimating the natural frequencies present in the data (indicated 'Parameter estimation' in Figure

6.14, this is separate from the testing region). Again, for the grey-box models, the search limits for optimisation of the kernel during training are set at ±20% of the values identified.

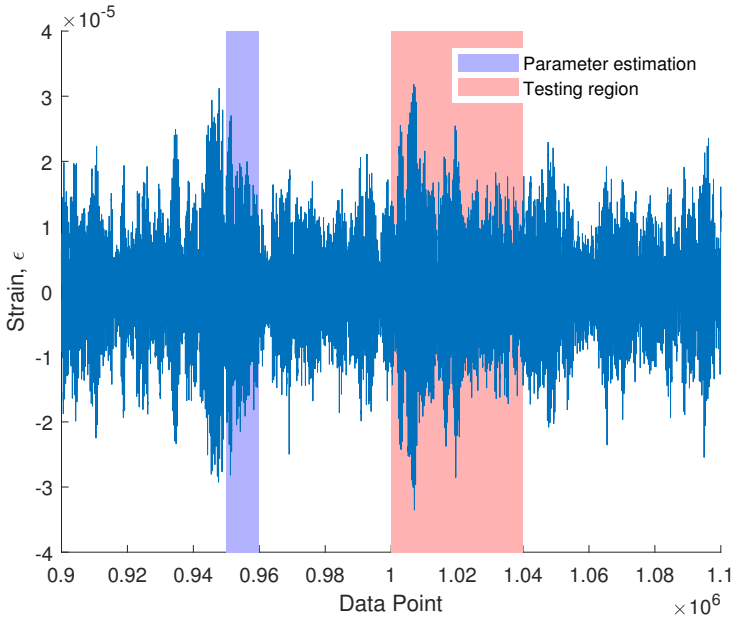


Figure 6.14: Measured strain data for the GARTEUR aircraft under Gaussian white noise forcing. The region of data used for testing and separate section of data used for estimating the natural frequencies of the system are indicated.

6.3.1 Data Processing and model inputs

The data is collected at 2048 Hz. A low-pass filter with a cutoff frequency of 256Hz is used to remove measurement noise in the data. The data is then downsampled to 1024Hz to enable a greater length of data to be assessed while maintaining a sample rate more than 10 times the last significant mode.

From the section of data used estimation of the natural frequencies, we can see that there are natural frequencies in the structure around 6Hz, 16Hz, 35Hz, 49Hz, and 64Hz. As discussed in Chapter 3, kernels can be combined to provide further flexibility and for the grey-box model, five SDOF kernels are used in a sum, with hyperparameter search limits set at ±20% of the natural frequencies identified. The effective sampling rate provided to train the models is 128 Hz, 1/16 data points, which is approximately the Nyquist frequency for the last major natural frequency.

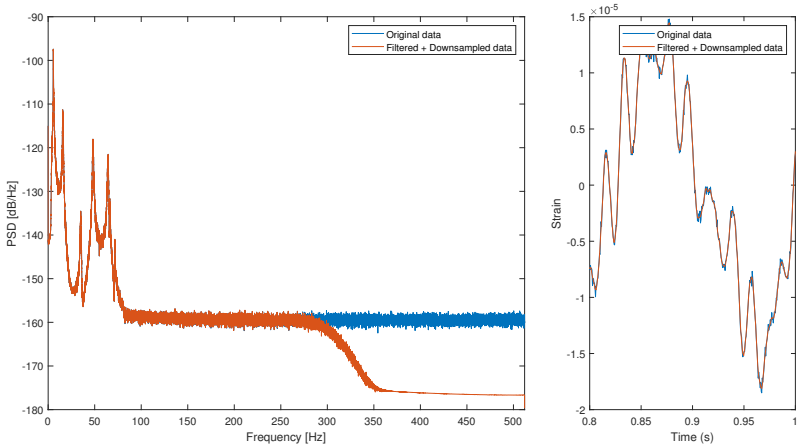


Figure 6.15: A lowpass filter is used to filter high frequency measurement noise with a cutoff frequency of 256Hz. The data is then downsampled from 2048 Hz to 1024 Hz to enable a greater time period of the stress response to be predicted.

The same sampling points are used across the range of models.

Model	Description	Inputs
6D	SE (black-box) kernel	acceleration (ms^{-2})
6E	SE (black-box) kernel	time (s)
6F	SDOF kernel	time (s)

Table 6.2: The three models assessed are the same as the simulated case study

6.3.2 Results

Model	MSE	MSLL	Fatigue Damage GP Mean (% Accuracy)	Fatigue Damage Samps Median (% Accuracy)
6D	57.1	49	-82	-82
6E	3.4	-1.6	-48	-44
6F	0.6	-2.7	-11	-8

Table 6.3: Overview of the results of the three models

6.3.3 Black-box models

The black-box models are again using a squared-exponential kernel. Model 6D, using acceleration as the model input is presented first. The first 3000 data points of the posterior prediction can be seen in Figure 6.16(a), with some samples shown in Figure 6.16.

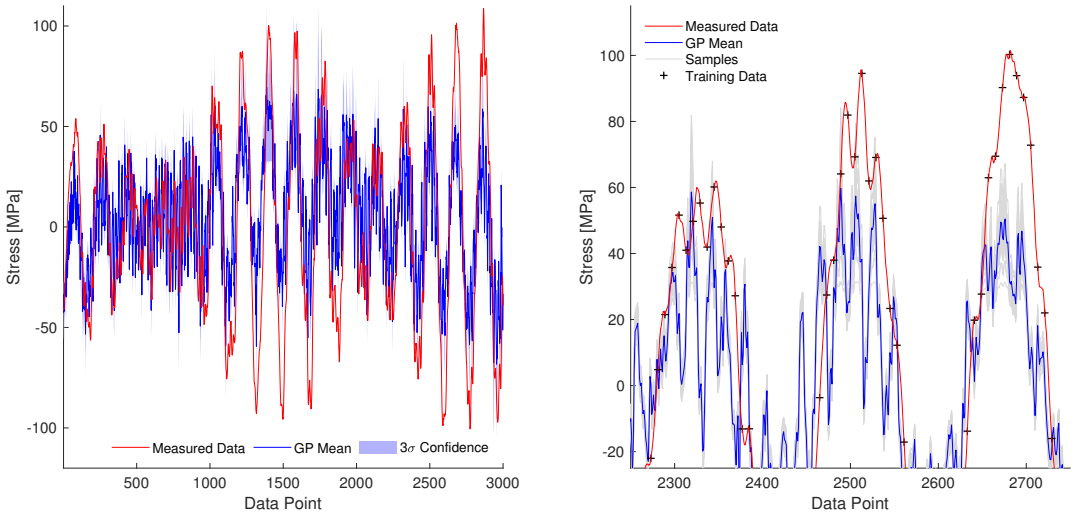
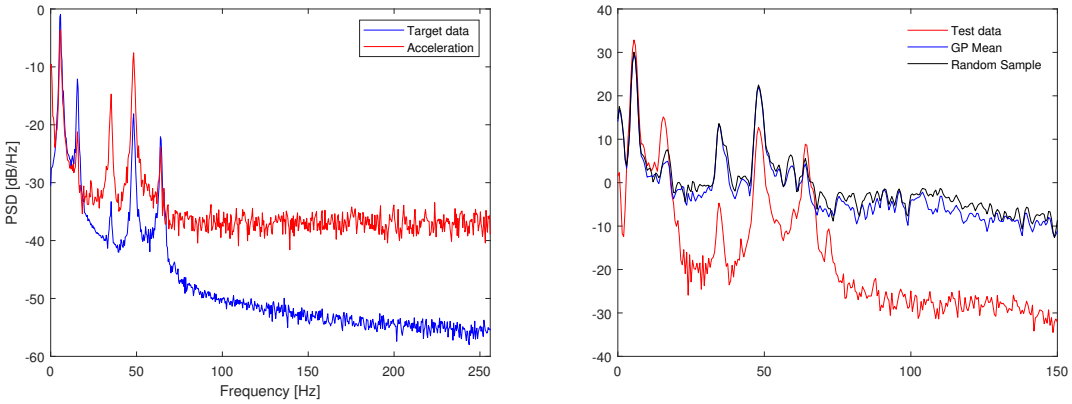


Figure 6.16: GP posterior prediction using squared-exponential kernel with acceleration as a model input: (a) First 3000 data points (b) An enhanced view of (a) with samples shown

Clearly, the model is overfitting with high-frequency fluctuations in the prediction which are not present in the measured data. Furthermore, the strain prediction is underestimating the peaks of the measured data. The MSE and MSLL values are both reflective of the fact that neither the GP mean, nor the posterior confidence, are reflective of the test data. By comparing the power spectral density (PSD) of the input and target data (Figure 6.17(a)), we can see that while the first two modes match, there is more behaviour around the 35 Hz and 45 Hz peaks in the acceleration data, in addition to a less severe drop-off in the power content above 64Hz. Figure 6.17(b), showing the frequency content of the GP posterior compared to the target data, shows that this undesirable high-frequency content is also present in the GP prediction. Furthermore, we can see that there is a lot less drop-off between the natural frequencies of the model compared to the measured data.



(a) (b)

Figure 6.17: (a) Frequency content of the target (strain) data in comparison to the input (acceleration) data. (b) Frequency content of the posterior predictions compared to the target data.

In addition to this undesirable high-frequency content, we can see that the confidence intervals of the model are not of any real utility, with the measured data lying outside of these bounds (as is, at some points, the training data). In Figure 6.18, showing an input-space representation of the model, we can see that, despite a large noise hyperparameter being adopted to capture the weak correlation of the training data, the posterior variance is still not able to capture the test data variance.

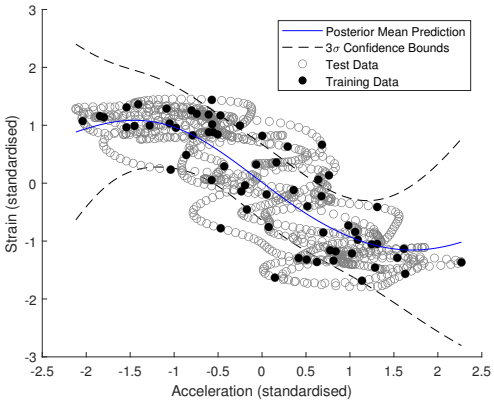


Figure 6.18: Input-space view of the acceleration-input black-box model (limited to first 1000 data points)

Unsurprisingly, the prediction of the fatigue damage accrued is poor (Figure 6.23a). From a GP mean perspective, the measured damage is underestimated by 82%, and sampling from the posterior does not significantly improve this (for the most damaging sample, this is reduced to 24%, but this sample is an outlier and the

majority of samples underestimate the measured damage by over 50%). The median of the sampled damages matches the prediction of the GP mean. Furthermore, the uncertainty of the model does not compensate effectively for this poor prediction. While there is a clear, albeit weak, correlation shown in Figure 6.18, it would arguably be more useful for the GP to revert to a zero-mean posterior to highlight the uncertainty of the model.

For the black-box model using time as an input (Model 6E), we can see a slight improvement in the model: the adopted model variance can capture the training points within the posterior confidence intervals (Figure 6.19). However, this comes at the cost of effectively being able to predict the higher frequency content in the model. This is visible in Figure 6.19(b), in which we can see that both the mean and the individual samples are smoothing through the dynamic turning points of the stress response. Figure 6.20 confirms this, where we can see that the spectral content for both the samples and GP mean drops off rapidly after the second natural frequency at 15 Hz.

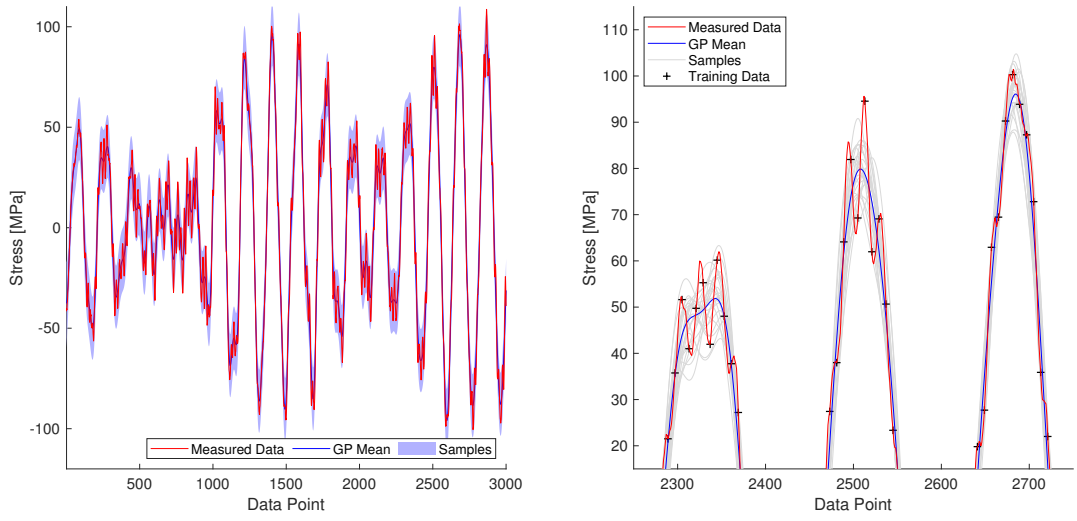


Figure 6.19: GP posterior prediction using squared-exponential kernel with time as a model input: (a) First 3000 data points (b) An enhanced view of (a) with samples shown

The result of this is a model that is performing better in terms of MSLL and MSE, but a fatigue damage prediction that is not significantly improved compared to Model 6D (the results for the three models are summarised in Table 6.3). The fatigue damage prediction from the GP mean and sampled median damages are

both improved compared to the acceleration-input model, yet they are both still significantly underestimating the measured damage state. Furthermore, the variance of the damage distribution does not satisfactorily grow to indicate the uncertainty of the model.

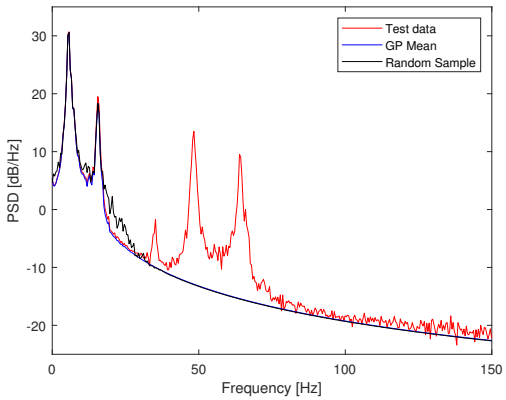


Figure 6.20: Frequency content for the black-box GP using time as an input to the model in which the first two natural frequencies are captured. However, oversmoothing through the data means that the higher modes are not captured.

6.3.4 Grey-box model

The results of the simulated case study demonstrated the performance of the SDOF kernel for a single natural frequency, while the data here displays five natural frequencies over the 4-80Hz excitation range. For this, we can demonstrate the ability to use multiple single-degree-of-freedom kernels in addition to modelling such multiple-degree-of-freedom (MDOF) problems.

Visually, from the posterior distribution shown in Figure 6.21, we can see that the prediction is significantly improved compared to the black-box models for both the GP mean and the posterior samples. This is confirmed in the prediction PSD (Figure 6.22) in which we can see that the 5 modes are captured well by both the mean prediction and the samples, although we can see that there is more spectral content in the samples around the third mode at 35Hz. The MSE and MSLL for this prediction are 0.6% and -2.7, respectively. The improvement is reflected in the fatigue damage prediction: while both the GP mean and samples median both still underestimate the damage from the test data, the extent of this error is significantly reduced to 12% and 8%, respectively. The damage distribution still under-predicts the measured

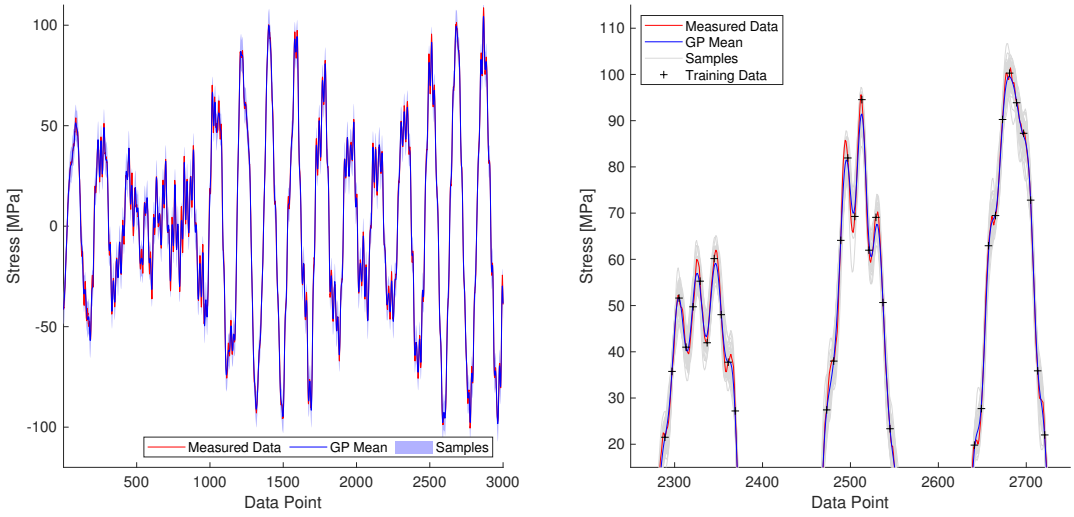


Figure 6.21: GP posterior prediction using SDOF kernel: (a) First 3000 data points (b) An enhanced view of (a) with samples shown

damage, with only the most damaging samples correctly predicting the measured damage state. However, the damage distribution is still significantly more accurate than the black-box models.

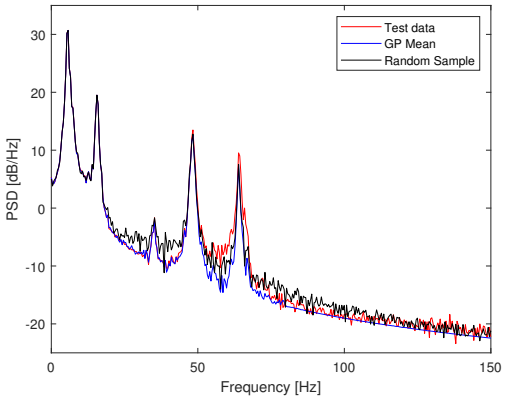
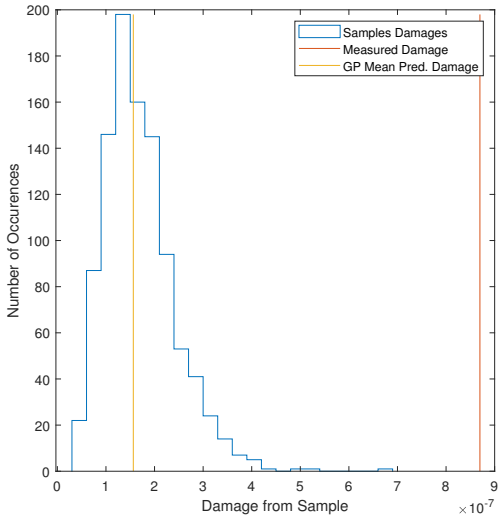
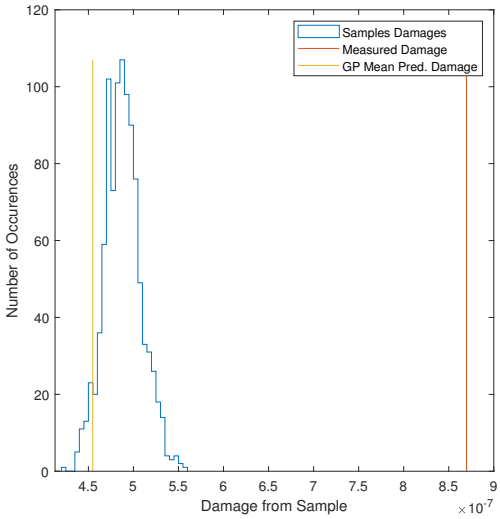


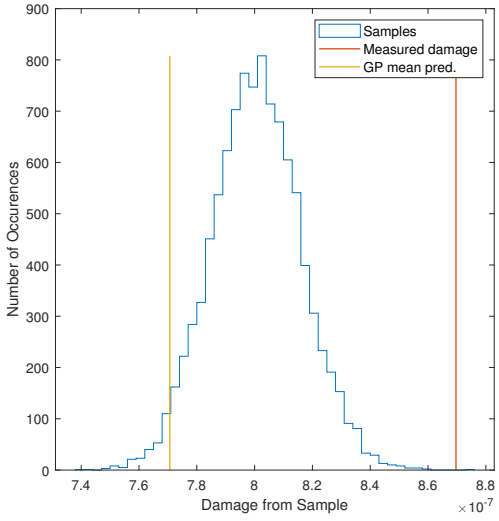
Figure 6.22: PSD for the test data, GP posterior mean and a random sample taken from the posterior. We can see that the model captures the spectral content for the test data well.



(a) Black-box: Acceleration input



(b) Black-box: Time input



(c) Grey-box: SDOF Kernel

Figure 6.23: Predictions of fatigue damage for the Garteur aircraft.

6.4 Conclusions

In this chapter, the concept of introducing knowledge of the physics of the process through the GP covariance function was introduced. Motivated by the fact that dynamic behaviour can have significant contribution to fatigue accumulation, a covariance function derived from a single-degree-of-freedom oscillator is used to predict the loads on a simulated case study. First, using a simulated case study, it was shown that using the SDOF covariance function, fatigue damage is able to be reconstituted to a good accuracy at sample rates significantly below those generally considered minimum. Using a black-box model, it was also possible to predict fatigue to a reasonable accuracy at sample rates above the Nyquist frequency of the simulation. Following this, a study was performed using the same simulation to test performance of the models at lower sampling rates. Here, the performance of the grey-box model outperformed the black-box models by a significant amount. While the black-box models became rapidly unuseable for damage prediction, the grey-box model was able to provide a reasonable prediction, even at the lowest sample rate tested.

Following this, the GARTEUR dataset is used to demonstrate the performance on a measured dataset. As there were multiple modes present in the data, it was demonstrated that, by using multiple kernels in addition, multi-modal data could also be captured. As per the simulated dataset, the grey-box approach outperformed the black-box models, both with respect to strain prediction and fatigue damage prediction. Power spectral density plots were used to demonstrate that the frequency content of the measured data was able to be captured, both by the posterior mean and the samples, using the grey-box model.

The field of physics-informed machine learning is a rapidly emerging and developing subject of research. While not the first work to utilise physics-informed approaches for fatigue [100, 101], it is the first to consider the problem of loads monitoring for S-N fatigue and, furthermore, the first to do this using a kernel-design approach. By utilising the probabilistic methodology developed in the previous chapter, this has further demonstrated both the utility of the probabilistic framework and also highlighted the improvement that can be attained using the physics-informed machine learning approach proposed here. In this chapter, this is limited to purely dynamic loading. As it is the superposition of both quasi-static and dynamic loads that is

often of importance, such problems will be investigated in the next chapter. The same physics-derived covariance function will be used in combination with data-driven methods, to assess the predictive ability for such problems.

Chapter 7

Using physics in more complicated cases: Combined loading

In the previous chapter, the use of the SDOF kernel had a clear, positive impact on the ability to predict fatigue damage under zero-mean white noise loading. However, it is often the superposition of multiple phenomena that we wish to predict. An example of this would be the superposition of quasi-static and dynamic loading on a wing due to turbulence while undergoing an aircraft manoeuvre. This is of particular interest here, as these phenomena in combination can have a large impact on fatigue damage due to the increase in severity of the maximum and minimum loads of a stress response [5]. In this chapter, the performance of the SDOF covariance function used in combination with a data-driven kernel to account for a quasi-static unknown load is demonstrated. As per the last chapter, this will be compared to a purely data-driven approach. Furthermore, the impact of using a mean function will be investigated: as discussed in Chapter 3, this is a relatively easy way of imparting knowledge of a process into a Gaussian process regression model. In this case, this is used to demonstrate an alternative option for capturing the quasi-static behaviour, if one is able to characterise such knowledge easily. Adopting such an approach will have a knock-on effect on the damage distribution, as using a deterministic mean function in this way would remove this load prediction from the probabilistic framework.

7.1 Combining Covariance Functions

For the MDOF problem in the previous chapter, a combination of SDOF kernels is used to model the strain history reliably. For this problem of combined loading, it is again difficult for a single kernel to perform adequately, especially if the phenomena are dissimilar in terms of the frequency content or causal factors. In such situations it is possible to use a combination of different kernels. Where a practitioner has some partial knowledge of the process of interest, this can then be incorporated into one of the kernels, while allowing the other kernel(s) to predict the unknown behaviour(s).

As discussed in Chapter 3, covariance functions can be combined through any linear operator, though most commonly, through multiplication or addition. In this work, the physical and data-driven kernels will be combined in addition: Figure 7.1 visualises this addition operation for a toy function using the SDOF and an SE kernel.

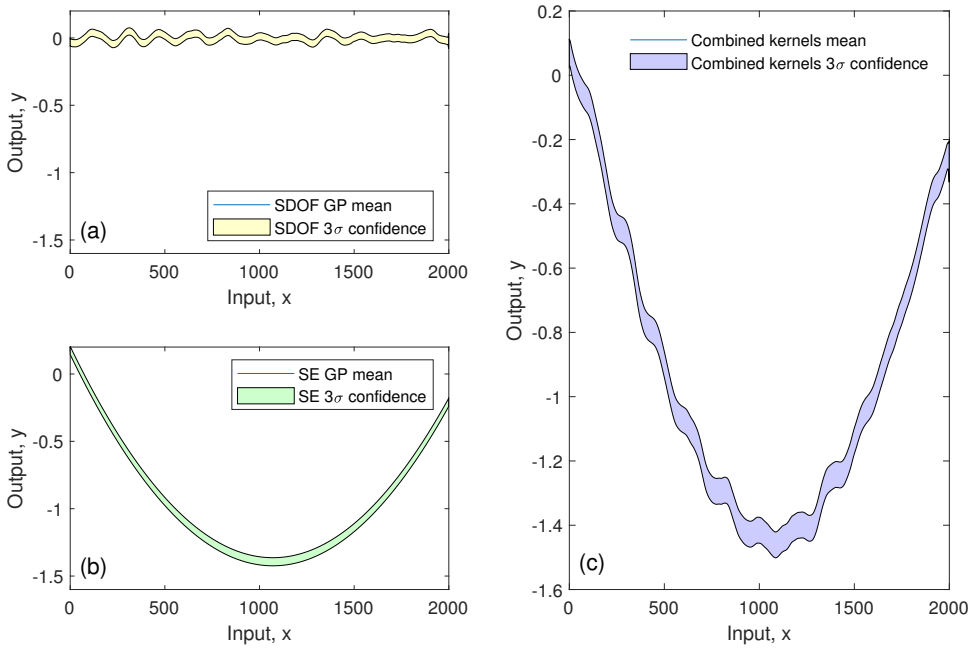


Figure 7.1: (a): Posterior mean and confidence using SDOF kernel, (b): Posterior mean and confidence using SE kernel and (c): Posterior mean and confidence using combined kernels.

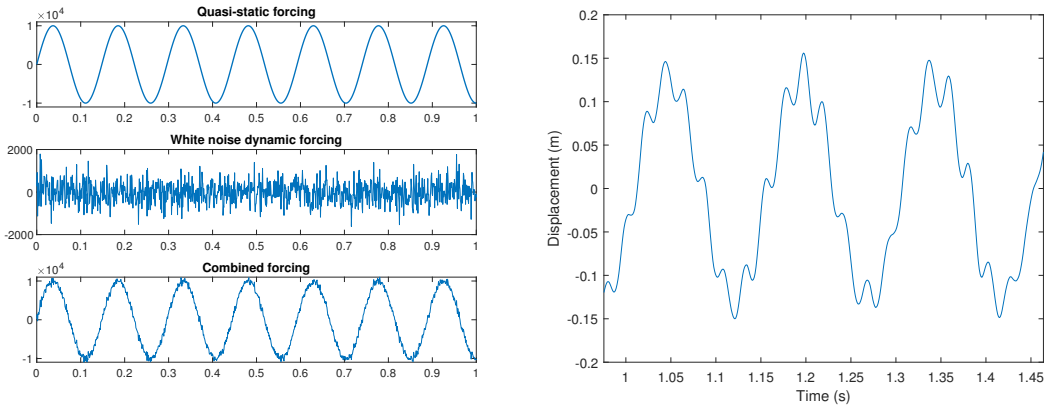
7.2 Simulated case study

As per the previous chapter, to assess the performance of the grey-box approach, a simulated case study shall initially be used. The same linear oscillator is simulated as in the previous chapter (Equation 6.2). However, in this case, rather than using a Gaussian white noise function to force the oscillator, the simulated system is forced using the combination of a Gaussian random load (the dynamic component of forcing) and a sinusoidal load (the static component of the load, with a frequency of 6.8 Hz, significantly below the natural frequency of the system of 45 Hz).

Mass m	**Natural Freq.:** ω_n	**Damping Ratio:** ζ	**Static Force Freq.:** f_{static}	**Sampling Freq.:** f_s
1 Kg	45.1 Hz	0.0301	6.8 Hz	1024 Hz

Table 7.1: Parameters of the simulated system

The forcing signal can be seen in Figure 7.2(a) and the resulting displacement can be seen in Figure 7.2(b).



(a) (b)

Figure 7.2: (a) Quasi-static, dynamic and combined forcing signals. The combined forcing is used as the input to the simulation (b) Simulated displacement data (first 1000 data points of 10000)

7.2.1 GP inputs and targets

For the black box models, force (N) is used as the input to the model while the SDOF kernel again uses time as a model input (Equation 6.4). Beyond using the

kernels in addition, in this chapter, we will also demonstrate how a mean function can be used, derived from the linear relationship seen in the data, to incorporate more intuitive knowledge of physics into the problem. As per Chapter 6, the output of the simulation is displacement, which is taken as a proxy for stress, as per [135]. This is scaled such that the maximum stress range is 600 MPa and then this is used as the target of the GP. Fatigue assessment for this chapter utilises the same S-N curve introduced in the last chapter, Figure 6.4.

The Nyquist frequency of the dynamic system is 90 Hz, while the data is generated at 1024 Hz. Initially, the training inputs to the GP will be provided at 171 Hz, i.e. 1 in 6 data points, sampled evenly across the data.

Model	**Description**	**Inputs**
7A	SE (black-box) kernel	Static input force (N)
7B	SE (black-box) kernel	Filtered Force (N)
7C	SDOF kernel	time (s)
7D	SDOF + SE kernel	time (s) + Static force (N)
7E	SDOF kernel (+ linear mean function = $f(F)$)	time (s)

Table 7.2: Three models for simulated case study

7.2.2 Black box model

Static forcing input

We will first look at a black-box model using only the static component of the forcing signal as an input to the model, Model 7A. As the forcing is static, one would realistically only expect this to be capable of predicting the static behaviour of the data. Indeed, this prediction is borne out in the GP posterior, Figure 7.3. The trend of the static behaviour is captured by the model, but the smaller, dynamic, fluctuations in stress are not predicted by either the GP posteior mean or the samples drawn from the posterior (some samples from the posterior are shown, in which we can see that the large model variance adopted does not result in the samples being more representative of the data, simply providing uncertainty on the amplitude of the quasi-static stress). Even though the prediction (the GP mean and its samples) are over-smoothing through the data, the model error in terms of NMSE is low at 1.1%.

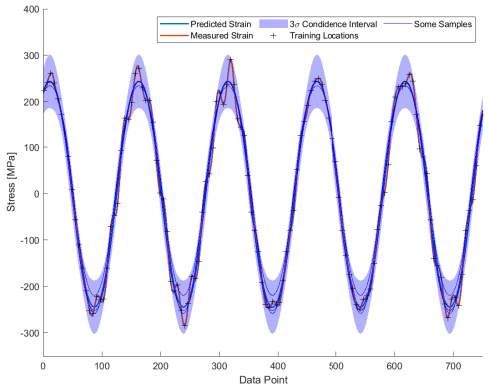


Figure 7.3: GP posterior for the black box model (static forcing input)

Unsurprisingly, Figure 7.4 shows the result of this is a poor prediction of fatigue damage. The more conservative samples from the damage distribution do indeed capture the true value yet the posterior mean underestimates the true damage state by 56%. However, we can see that the median sample damage is also underestimating the accrued fatigue damage by a similar amount. Furthermore, the variance of the sampled damages is also large, with a large range of possible values for damage (the most damaging sample is almost 12 times more damaging than the least damaging). The MSLL, assessing the predictive distribution, is -2.1 - owing to the fact that the simulation data is captured within the confidence intervals of the model. As attested by the fatigue damage distribution, the fact that this distribution is capturing the measured data is not of high significance if the trend of the underlying physical process is not captured.

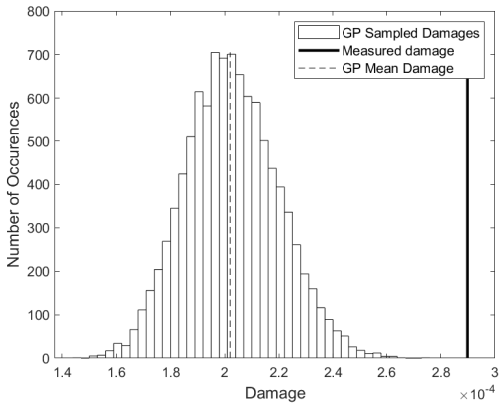


Figure 7.4: Damage distribution for the black box model (static-forcing input)

Combined forcing input

For Model 7B, the full forcing data is now used as the input to the model, itself a combination of static and dynamic forcing. However, as the dynamic forcing has spectral content significantly above the frequency of the natural frequency of the system, a low-pass filter with a cut-off frequency of 50 Hz is used. The original and filtered data can be seen in Figure 7.5

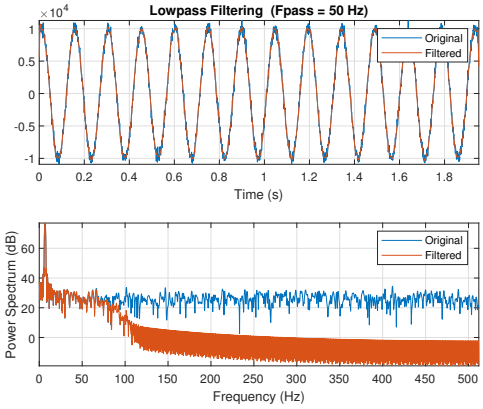
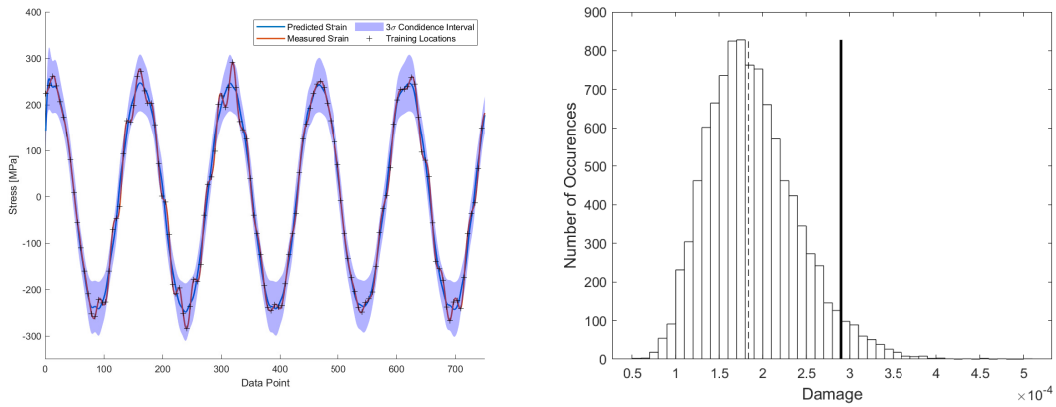


Figure 7.5: A lowpass filter is used to remove spectral content from the forcing signal before use as an input into the model.

From Figure 7.6, a slight improvement in the model performance can be seen compared to model 7A, with less over-smoothing of the data when compared to using the static-only forcing. However, the model is still clearly not capturing the underlying physical process. The MSE and MSLL for this prediction are 1.15% and -2.15, respectively, again indicating a good result. However, there is very little change reflected in the damage distribution shown in Figure 7.6(b), where a similar damage variance and mean prediction error can be seen as seen in the static-only forcing model. The GP mean still underestimates the damage state by 64%. The true damage state is captured by the damage distribution again, but the variance is such that the most damaging sample is 11 times the damage of the least damaging cycle (meanwhile, the least damaging sample is almost 8 times less than the damage of the test data). The level to which the results shown here underestimate the damage of the simulated data is severe, but, in general, the results are coherent with the study into expected distributions shown in Chapter 5.

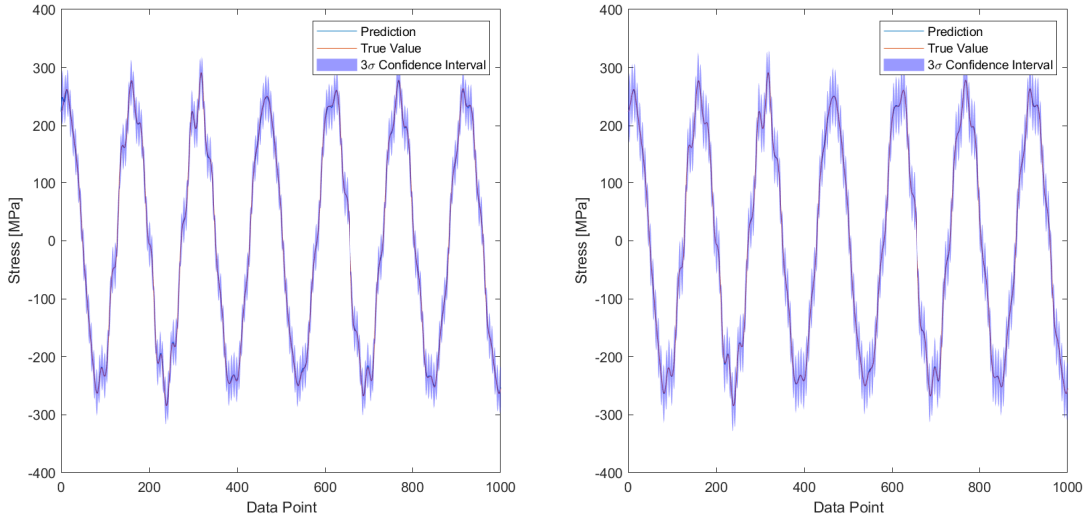


(a) (b)

Figure 7.6: Black-box GP with filtered combined forcing input (a) GP posterior (b) Damage distribution.

7.2.3 Grey box - SDOF kernel used in isolation

The SDOF kernel is now used as the covariance function for the GP (Model 7C). As discussed, time is now used as the input for this kernel, with the training indices the same as those used for the black-box model. For the same reasons as in the previous chapter, the search parameters of the (physical) hyperparameters are limited to ± 20% of the actual values. The resulting GP posterior is shown in Figure 7.7(a).



(a) SDOF Kernel (7C) (b) SDOF + SE Kernel (7D)

Figure 7.7: GP posterior distributions using the SDOF kernel on the simulated dataset.

As was the case in the last chapter, we can see a large improvement compared to the black-box models with the grey-box approach. The NMSE for this prediction is 0.1% and the MSLL is -3.6, but most importantly, the GP is capturing the underlying phenomena very accurately: the dynamic behaviour of the stress response is captured here in addition to the static behaviour.

The GP posterior mean can predict the sampled damages very accurately, with an error of only 0.5% compared to the test data. We can see that the measured damage is in the less damaging end of the damage spectrum. The variance of the damage from the samples is significantly reduced: while for both black box models, the most damaging sample was over ten times the damage of the least damaging, this multiple is reduced to only 1.65 for the grey-box prediction.

We can see a clear conservative shift in the damage distribution compared to the GP posterior mean prediction: this was discussed in the previous chapters but can be illustrated further in Figure 7.8, where an enhanced view of the posterior is shown for the following case study (model 7D, however, the same conclusions apply for Model 7C). In the study into the causes of shifts in the damage distribution carried out in Chapter 5, it could be seen that higher frequency content (or a shorter lengthscale) caused a damaging shift in the distribution. Although the prediction is accurate here, uncertainty on the amplitude of the dynamic behaviour is causing this damaging shift. Here, the effective sampling rate is around two times the Nyquist frequency (171 Hz vs 91 Hz): as the effective sampling rate increases, the conservative shift of this should reduce, as was found for the case study in Chapter 5. While the sampled damages here are indeed overly conservative, the improvement in comparison to the black-box models is significant. In the concluding chapter of this thesis, ways of mitigating this shall be considered.

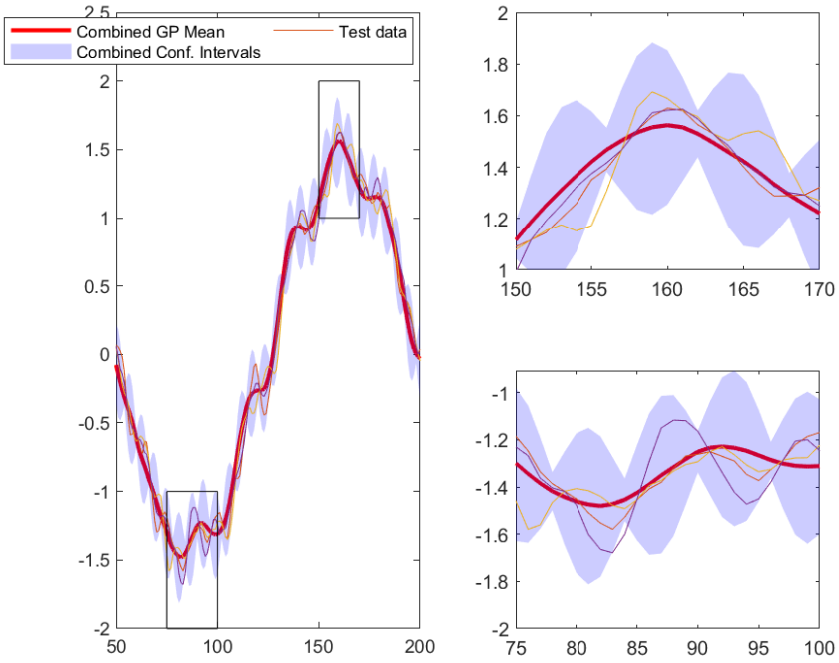
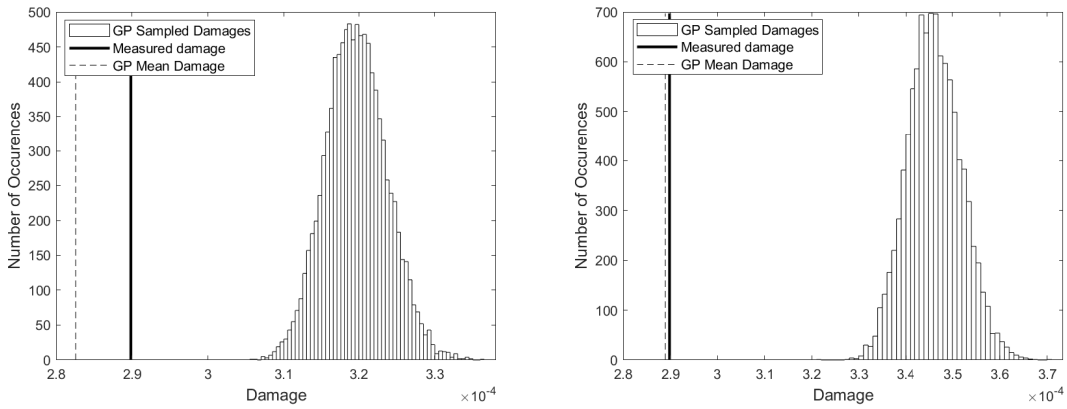


Figure 7.8: Enhanced view of combined SE + SDOF kernel (see Figure 7.7(b). It can be seen that each of the samples shown exceed the simulation data at one point of the local maxima. Over the full prediction, this means that many of the larger stress cycles will increase as a result of the sampling process. The same phenomena can be seen in the SDOF only model



(a) SDOF Kernel (7C) (b) SDOF + SE Kernel (7D)

Figure 7.9: Damage distributions using the SDOF kernel on the simulated dataset. While the distribution in both cases is conservative (see earlier discussion), the accuracy of mean of the sampled damages and variance is much improved for both cases compared to the black-box models

7.2.4 Combination of kernels: SDOF + SE

In the black-box models above, the SE kernel was capable of predicting the static component of the data accurately. While attempting to predict the combined loading was not successful, we can use the knowledge that the model is capable of predicting the static loading by using a black-box kernel in combination with the SDOF. The static forcing signal is used for this.

The optimised GP posterior can be seen in Figure 7.7(b). The MSLL for the prediction is -4.8, indicating a small improvement over the already very well-performing example of the SDOF kernel above (the MSE is similar, and low enough in both cases to have little significance).

The contribution of each kernel can be calculated according to [114]:

$$f_1(x^*)|f_1(x), f_2(x) \sim \mathcal{N}(\mu_1^* + K_1^*(K_1 + K_2)^{-1}\left[f_1(x) + f_2(x)\right) - \mu_1\mu_2], \\ K_1^{**} - K_1^{*T}(K_1 + K_2)^{-1}K_1^*) \tag{7.1}$$

By computing the contributions of each kernel to the GP (shown in Figure 7.10), we can see that the Kernel is performing exactly as one would hope: the SE kernel is predicting the static component (with a significantly improved level of confidence compared to this kernel used alone, see Figure 7.3) and the SDOF kernel is predicting the dynamic component. The subsequent prediction is not significantly changed compared to the SDOF kernel used in isolation and, indeed, the damage distribution shown in Figure 7.9(b) shows a slightly less accurate predictive distribution of the damage state than the SDOF kernel used alone. However, here the SDOF kernel is not needing to guide the mean of the posterior prediction (which it is capable of doing by nature of having target data within the test region). Future work will look to extrapolate prediction to locations on a structure where measured data is not available, thus increasing the importance of being able to use separate kernels to complement each other.

As per the SDOF kernel used alone, we can see that, despite the high accuracy of the model, the fatigue damage prediction overestimates the damage of the test data. The reasons for this are the same as described above for the SDOF-only case, visualised

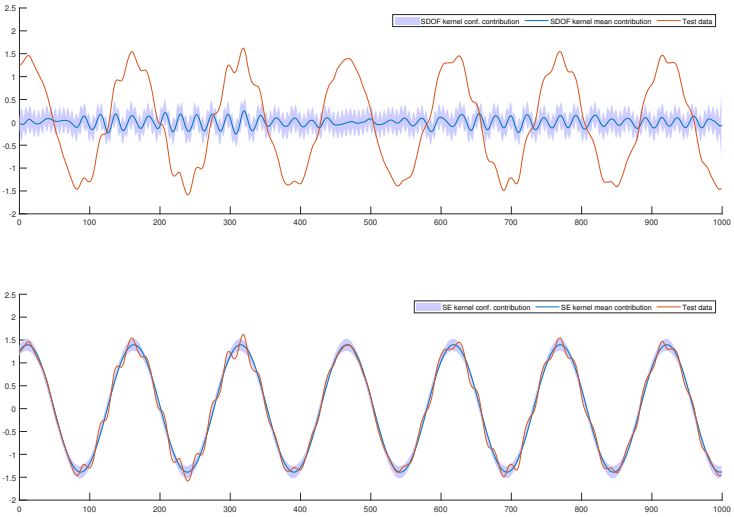


Figure 7.10: The contributions of the individual kernels to the result

in Figure 7.8.

7.2.5 Use of a mean function

Lastly, we will consider the use of a *mean function* (Model 7E) as an alternative, or additional, way of including our knowledge of the physics of the problem into the model. As we have a known force and can calculate the stress as $\sigma = \frac{F}{A}$, even if we do not have much knowledge of the geometry of the structure (as may often be the case for virtual load sensing problems) we can estimate the static component of the data as:

$$\sigma_{static} = cF_{static} \tag{7.2}$$

where c is a constant that can be derived using linear regression, using the training data available. After estimating the static forcing using this, we are left with a *residual*, shown in Figure 7.11, that resembles the dynamic component of the data. If used in isolation, the mean function would underestimate the fatigue damage of the data by 37%. However, by training a GP to predict the residual, we can attempt to reduce this or gain an understanding of the magnitude of the error.

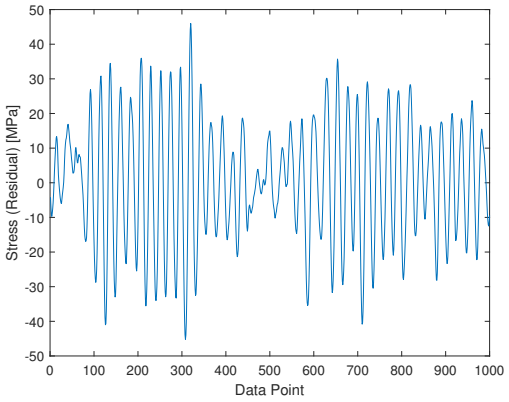


Figure 7.11: Residual of error using linear mean function. Despite the mean function being empirically estimated, the remaining data to be predicted resembles the dynamic component of the data

The SDOF kernel is used to model this residual with the same hyper-parameter search limits used as above. The GP posterior can be seen in Figure 7.12 where we can see that, while some fluctuations are not captured by the mean prediction, dynamic behaviour is overall predicted very well. The NMSE and MSLL values for this residual prediction are 0.4% and -3.2, respectively.

To predict the full stress response, we can add this to the GP posterior: the combined prediction can be seen in Figure 7.13(a). The stress prediction matches the data very closely and we can see from the damage distribution in Figure 7.13(b) that the fatigue damage distribution matches the true value closely. It can also be noted here that the damage distribution has a smaller variance in this case: the most damaging sample is less than 2% greater than the least damaging sample. This arises from the fact that the mean function does not have an uncertainty to propagate (compared to using a black-box GP to predict the static component of the data). Naturally, this could be a good, or bad, thing depending on how much certainty we have about the static behaviour. A similar effect could be achieved by severely limiting the noise hyperparameter for the black-box GP and restricting the potential posterior covariance as a result. A major advantage of this approach, compared to using a combination of kernels, is that the draws from the confidence intervals maintain their physical significance (as it is only the SDOF covariance that is being inverted, while this is not the case when using a combination of kernels, see equation 3.4).

For comparison, the performance of an SE kernel for predicting this residual can be seen in Figure 7.14. It can be seen that the GP is not able to capture the underlying

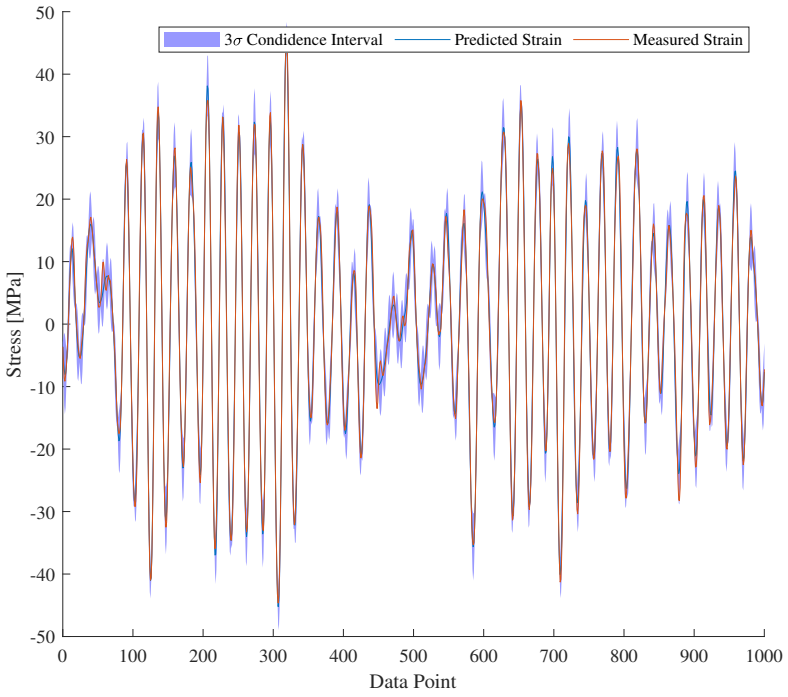
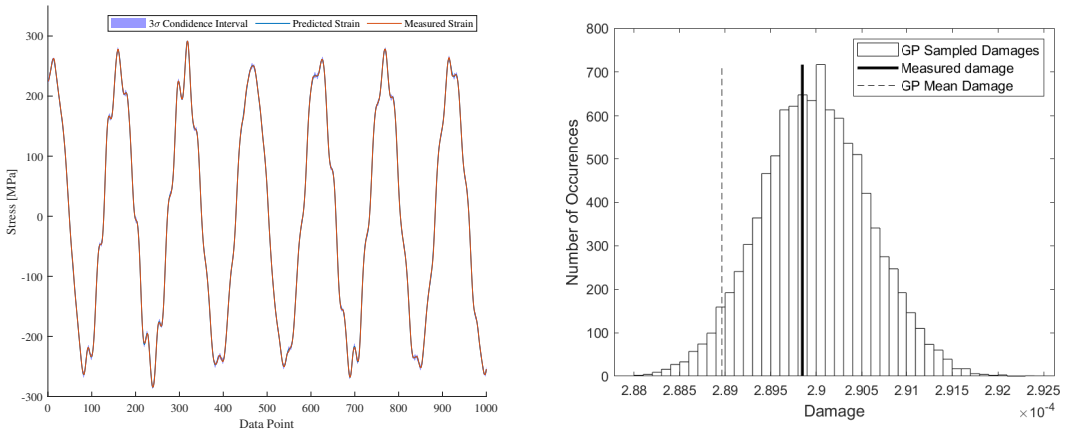


Figure 7.12: GP posterior: SDOF covariance function used to predict the residual shown in Figure 7.11 The combined mean function + GP prediction can be seen in Figure 7.13(a). A black-box prediction can be seen in Appendix Figure 7.14



(a) Posterior

(b) Damage distribution

Figure 7.13: Posterior and damage distributions using SDOF kernel with linear mean function. A significant improvement can be seen here, especially in the damage distribution

nature of the data. Instead, a large posterior variance is adopted to account for the fluctuations in the data. The GP mean shows some fluctuation, but is close to zero mean and not representative of the dynamic process as per the SDOF model. Drawing samples (not shown) from the distribution is not of help here, as they simply take the form of a static offset of the GP posterior mean.

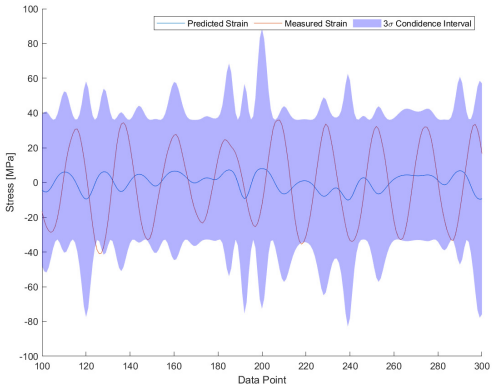


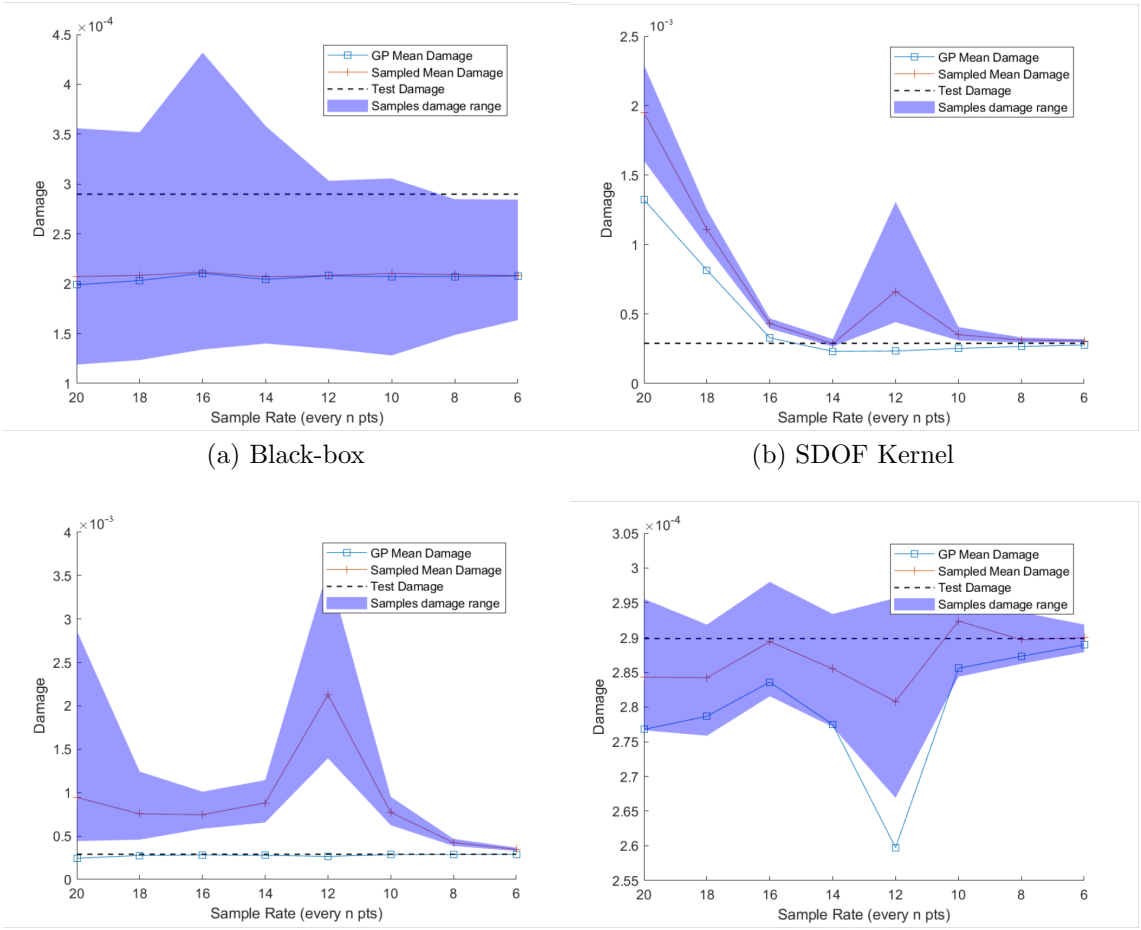
Figure 7.14: Prediction of the residual shown in Figure 7.11 using black-box SE kernel

7.2.6 Reducing sampling rate

As in the previous chapter, we will now test the performance of each approach under lower sampling rates. The effective sampling rate for the above studies was 171 Hz. For this study, sample rates between 51 and 171 Hz shall be used. Figures 7.15a - 7.15d show the damage progression for each of the models.

Figure (a) shows the prediction for the black-box model (with combined (filtered) forcing input, Model 7B). Unsurprisingly (as the prediction quality at the highest sampling rate was not good), the prediction is poor across the range of sample rates. It can be noted that the prediction does not get significantly worse with reducing sample rates, and instead, is relatively steady in terms of the GP mean prediction and the sampled mean prediction. In fact, we can note that, by reducing the sample rate, the model variance has increased such that the model confidence is encapsulating the actual damage state at the lower sampling rates (the slight variance increase here arises from the distance between training and testing point increasing).

In models 7C and 7D, the damage distribution was conservative at the original sampling rate of 171 Hz. We can see that this remains true as the sample rate



(a) Black-box

(b) SDOF Kernel

(c) SDOF + SE Kernel

(d) SDOF Kernel + Mean Function

Figure 7.15: Impact of reducing the sampling rate on Fatigue damage prediction distributions - it can be seen that just below the Nyquist rate, the SDOF kernel struggles. Elsewhere, a clear improvement can be seen using the physics inspired kernel.

decreases. While for Model 7D, which utilised the data-driven kernel in combination with the SDOF, the damage prediction does not get significantly worse (with the exception the sampling frequency just below the Nyquist rate), as was the case with the SDOF kernel for dynamic problems in the previous chapter. By contrast, in Model 7C, the performance significantly decreases from both a fatigue and stress prediction perspective. This is visible in Figure 7.16, which shows the posterior prediction for this case. While clearly the training data is sufficient to estimate the static behaviour, the optimisation is attempting to fit a posterior that is coherent with the search limits that were provided of $\omega_n \pm 20\%$ As a result, the model is not able to estimate the amplitude of this dynamic process, resulting in severe overestimation of

the model damage and poor MSE and MSLL values.

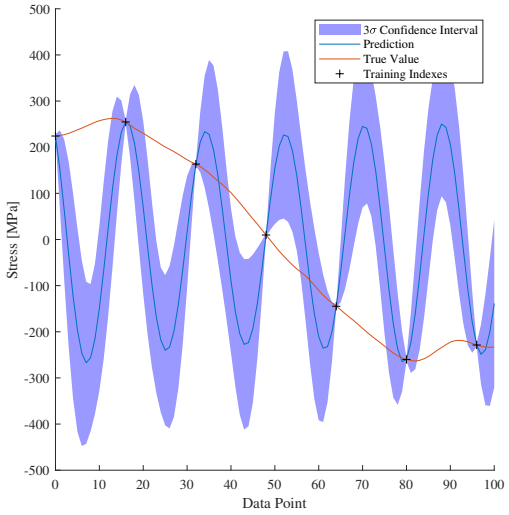
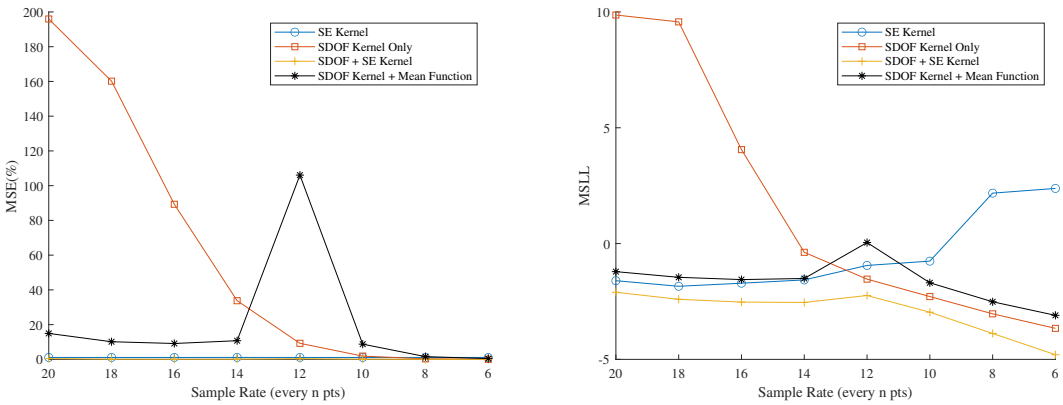


Figure 7.16: SDOF model with sampling rate of 1/16 data points, well below the Nyquist rate of 1/12 points. While clearly the sufficient training data is available to estimate the static behaviour, the optimisation is attempting to fit a posterior that is coherent with the search limits that were provided of $\omega_n \pm 20\%$ As a result, the model is not able to estimate the amplitude of this dynamic process, resulting in severe overestimation of the model damage and poor MSE and MSLL values

The excellent performance of the SDOF kernel in combination with the mean function across sample rates is notable. Given the performance of the SDOF kernel under varying sample rates in the previous chapter, this could be somewhat expected given that the residual prediction indeed resembles the dynamic component of the stress very closely. However, in comparison to the results of the other models here, the performance is excellent across the range of sample rates. Again, this is only possible in this case (to be used as a point of discussion), as estimating an appropriate mean function is trivial.

Figure 7.17 shows the impact that sample rate has on the MSE and MSLL values. From both this figure and the damage distributions, it can be noted that while the SDOF kernel performed well at the original sampling rate, its performance is the worst at lower sampling rates, both for fatigue prediction and also in terms of NMSE/MSLL. Again, this is explained by the prediction shown in Figure 7.16 (and the text in the associated caption).

Again, the performance of the SDOF kernel is worst at a sampling rate just below the



(a) MSE (b) MSLL

Figure 7.17: Progression of Mean Square Error (MSE) and Mean Standardised Log Logg (MSLL) with sample rate. The values for the *SDOF Kernel + Mean Function* (Model 7E) refer only to the residual prediction, amplifying the error.

Nyquist frequency, as per the previous chapter. While this is still the case, the impact of this is significantly reduced when the mean function is used in combination with the SDOF kernel. As the residual model can achieve the same degree of accuracy, regardless of sampling rate, we see much less variation in the performance of the model with varying sample rates. It can be argued, then, that it could be beneficial to use a mean function for such problems when a given component of the data can be estimated with accuracy, thereby removing this known entity from the probabilistic framework and allowing the GP to focus on the prediction of the more complex phenomena, to be discussed further in the conclusions.

7.3 Case study using Tucano dataset

We will now use the Tucano data to further demonstrate the performance of the SDOF kernel for combined loading problems when used in combination with a data-driven kernel. As per Table 3.1, strain data on the wings is available at 128 Hz for this data, but the majority of the data is only available at 16 Hz. The work in Chapters 4 and 5 of this thesis focussed on predicting the quasi-static behaviour of the data. While the discrepancy between sampling rates was acknowledged, the high-frequency component of the strain-time history was not considered. For the majority of flights, this is acceptable, however, some flights show additional dynamic

behaviour at around 11 Hz and 33 Hz. This higher frequency behaviour is the subject of this case study.

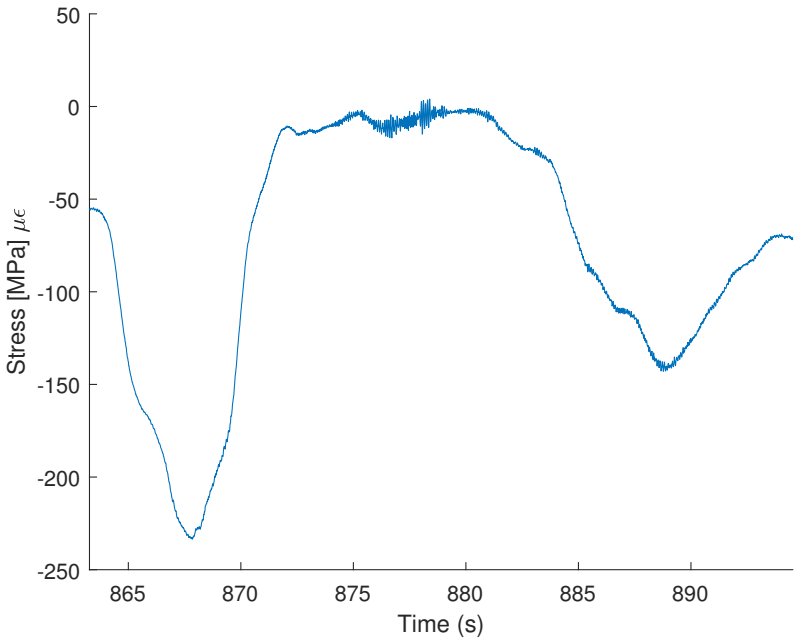


Figure 7.18: Tucano data under consideration for case study

Rather than predicting the full range of complete flights, for this study, only a 4000 data point section of Flight 4 shall be assessed. The strain-time history under consideration can be seen in Figure 7.18 where we can intuit that the data is mostly static, with some dynamic behaviour between 875 and 885 seconds. The PSD for the 4000 data points can be seen in Figure 7.19.

7.3.1 Model Inputs

Following the results of the previous section, two models shall be compared for prediction (summarised in Table 7.3): firstly, a black box model using acceleration at the aircraft center-of-gravity. This data is only available at 32 Hz and therefore has to be upsampled to 128 Hz for the test case: this is achieved using spline fitting. Note that this is not using the full range of inputs used in Chapters 4 and 5.

The second model will use the SDOF kernel in addition with an SE kernel. A mean function approach is not considered for this data as a linear fit cannot satisfactorily estimate the static component of the data with the acceleration data used. The

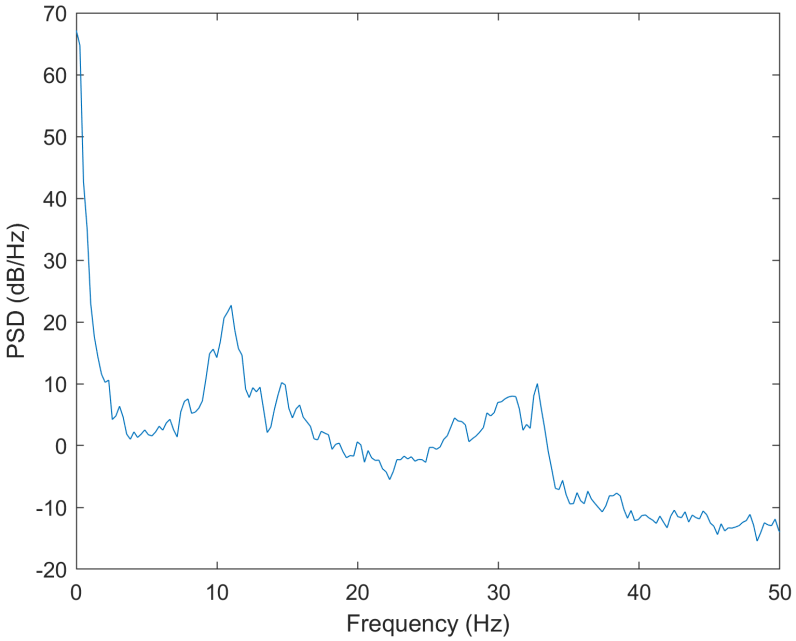


Figure 7.19: Power spectral density plot for the 4000 data points from the Tucano dataset used for this case study. While in Chapters 4 and 5, the data was downsampled to 16 Hz, in this chapter, the focus shall be on predicting the higher frequency component of the data.

inputs used in this case are time for the SDOF kernel and the same acceleration measurement as above (with upsampling) for the SDOF kernel. The same training indices are used for the two models.

Model	**Description**	**Inputs**
7F	SE (black-box) kernel	Acceleration (g), upsampled
7G	SDOF + SE Kernels	Acceleration (g), time (s)

Table 7.3: Two models under comparison for the Tucano dynamic prediction

7.3.2 SE Model

The GP posterior for the black box model (7F) can be seen in Figure 7.20. As expected (following the results of chapters 4 and 5), the static prediction is reasonable, with some errors around the dynamic section of the data (note that the black-box model is not using the full range of inputs used for the work in Chapters 4 and 5). The NMSE of the prediction is only 0.2% and the MSLL is -1.7, indicating a

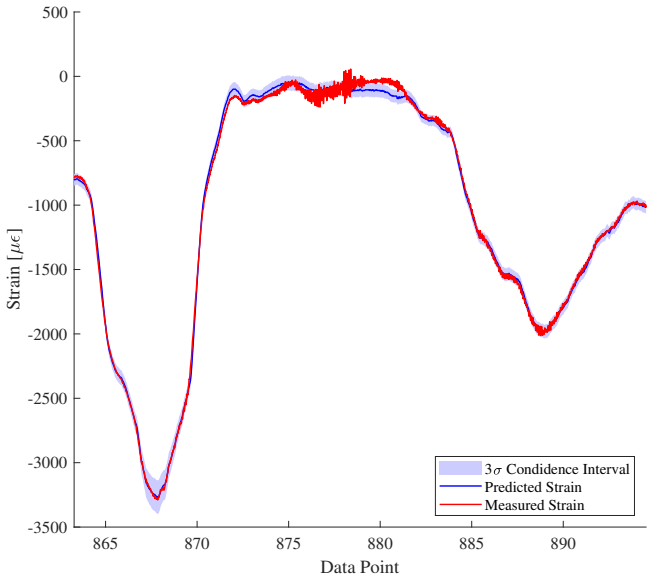


Figure 7.20: Black box strain prediction for Tucano dataset. The prediction is reasonable, however, is not able to predict the high-frequency behaviour around 870-885 seconds.

good prediction. Clearly, however, the posterior is smoothing through the dynamic behavior. The result of this on estimated fatigue damage accumulation (fatigue assessment is performed with the same S-N curve used for the Tucano work in chapters 4 and 5) can be seen in Figure 7.21: even for this short section of data, the impact of error prediction is severe. As per the simulated case study, we can see that only the most conservative samples are indicative of the actual measured damage, with the sampled mean and posterior mean both underestimating the measured damage by 19.9% and 19.7%, respectively. Again, the variance of the data distribution is very large, with the most damaging cycles being 72% more damaging than the least damaging sample. Overall, despite a reasonable strain prediction, the fatigue damage prediction exaggerates the model error and uncertainty. Again, the distribution shown in Figure 7.21 agrees with the study carried out in Chapter 5: by underestimating the high frequency component of the data, even with good strain prediction (in terms of NMSE and MSLL), there is a non-conservative shift in the damage distribution.

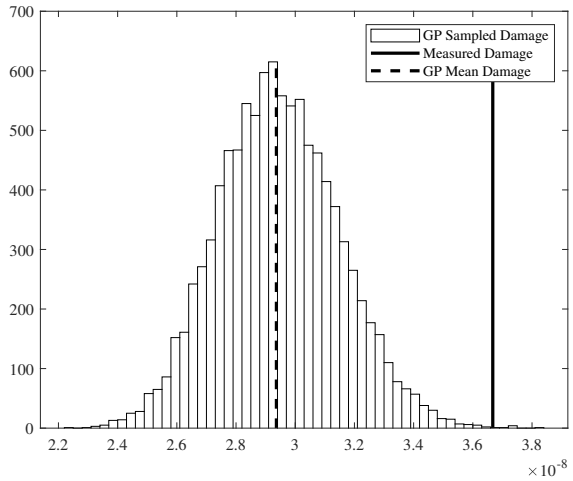


Figure 7.21: Damage distribution from Tucano case study using SE kernel

7.3.3 SE and SDOF in addition

Following this, the SDOF kernel is used in combination with an SE kernel (model 7G). Again, the search limits for the hyperparameters are limited to ±20% of the prior estimated values (estimated by plotting the PSD of an alternative flight to the one under consideration).

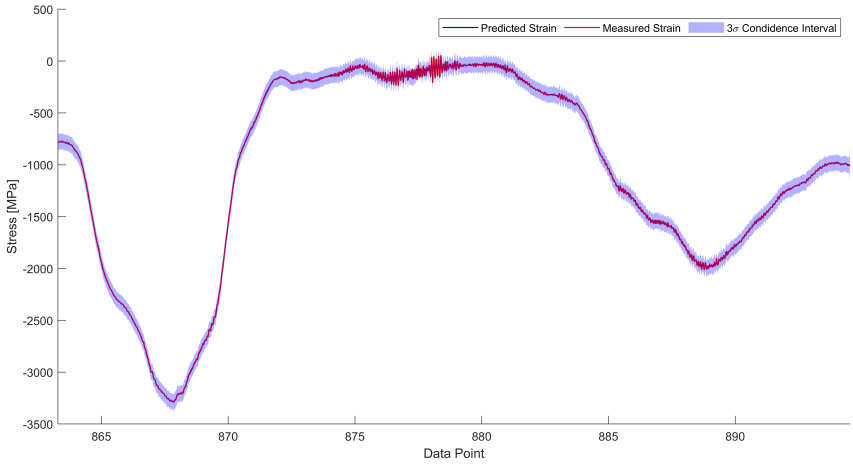


Figure 7.22: Using SDOF + SE Kernels in combination for the Tucano dataset.

The posterior can be seen in Figure 7.22. We can see that the prediction performs much better than the black box model. The model NMSE and MSLL are 0.1% and

-4.7, respectively.

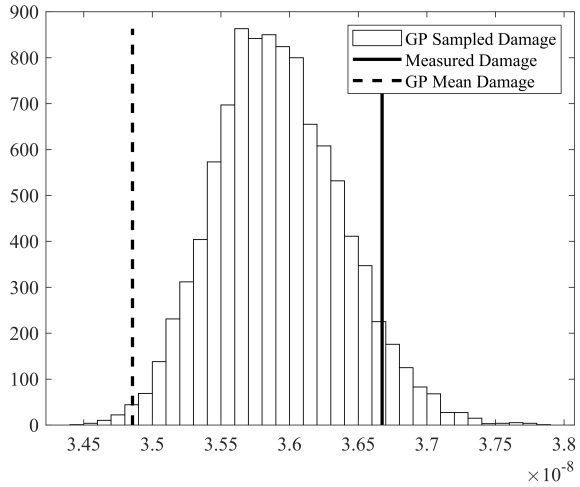


Figure 7.23: Damage distribution from Tucano case study using SE + SDOF Kernels in Combination

Again, we can plot the contribution of the individual kernels, shown in Figure 7.24. Again, the model is (approximately) behaving as one would hope: The SE kernel is performing the prediction of the static component of the behaviour, as we know it is capable of performing from the black box study (Model 7F, above). We can see that the SDOF kernel is compensating slightly for the poor performance of the SE kernel. For example, around data point 3000, we can see that the SE model is overestimating the strain value slightly, and the SDOF kernel is correspondingly correcting for this. More attention would have to be paid in such areas for future work when extrapolating to unknown locations on a structure. Overall, however, the combined strain prediction is very good.

This follows through to the damage prediction, shown in Figure 7.23, where we can see a significant improvement in model confidence. While the GP mean and many of the samples are underestimating the measured damage state, the prediction is much closer to the true value. It can be noted that we are not seeing the conservative shift shown in the case study: as only one SDOF kernel is used to capture the dynamic behaviour around 11 Hz, there is some additional high-frequency content that is not captured. Again referring to the discussion in Chapter 5, it is hypothesised that while the 11 Hz component of the data is most important (from a strain prediction and fatigue damage perspective), this higher frequency content is the cause of the additional damage in the measured data compared to majority of samples. However, the overall fatigue prediction is good: the damage variance is low, and the most

damaging sample is only 10% greater than the least damaging sample, representing a significant improvement on the damage variance seen in the original Tucano case study, chapters 4 and 5.

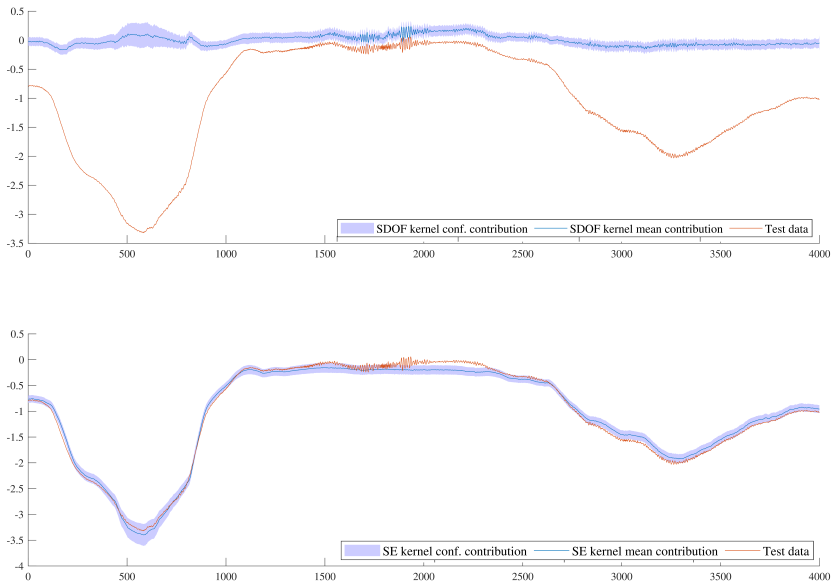


Figure 7.24: Kernel contributions for the SDOF (top) and SE (bottom) kernels.

7.4 Discussion and Conclusions

Having introduced a novel way of integrating physics into data-driven models for strain-prediction in the previous chapter, this chapter demonstrates this for increasingly complicated loading cases. This is the first work to demonstrate the flexibility that can be achieved using the single-degree-of-freedom kernel in combination with data-driven kernels to predict combined loading scenarios and, furthermore, utilise this approach for damage assessment.

The results shown here show very promising progress in the practical applications of GP regression for fatigue damage prediction. The earlier chapters showed that propagating model uncertainty is a crucial step in enabling a greater level of trust to be attained in data-driven models, but by incorporating our physical knowledge into the work, some of the major problems demonstrated in earlier chapters have

been mitigated. In particular, the large variance of damage distributions has been mitigated. In the previous chapter, this was shown for dynamically forced problems. In this chapter, it has been shown that, by combining data-driven and the physics-derived kernel, we can extend this to problems of combined loading. Using this methodology, we see a greater capture of the nature of the underlying phenomena both in terms of the posterior mean and also the samples drawn from the posterior covariance.

Again, the black-box models have shown that 'good' results in terms of NMSE and MSLL do not result in reliable fatigue assessment, even when propagating the uncertainty of the model. Instead, here, the main benefit to arise from this was drawing attention to the uncertainty in the model via the large damage distribution variance. Again, it is highlighted that propagating the uncertainty of the model is not a 'silver bullet' for a poor model to capture the damage state *somewhere* within the predictive distribution.

The introduction of physical knowledge demonstrated clear improvements. The SDOF kernel when used in isolation demonstrates sufficient flexibility to model the combined loading at the original sampling rate. However, as the sampling rate is reduced, the kernel is not able to accurately predict the amplitude of the dynamic behaviour. A combination of kernels can provide a better physical representation also, with each kernel aiming to model defined phenomena. Furthermore, future work which will aim to extrapolate strain prediction to locations on a structure where measured data is not available will benefit from the use of (an) additional kernel(s).

The use of a mean function for the simulated data improved the results further. For this data, calculating such a mean function is trivial, and thus, such an approach can be easily achieved. However, it is recognised that this is not always the case, as per the Tucano case study. While the authors advocate for a probabilistic approach, this case highlights the importance of limiting this to elements of the data that cannot be estimated easily. By removing the component of the data that could be estimated trivially from the probabilistic framework, the variance of the damage distribution could be reduced significantly. In this specific case, this has the additional benefit of maintaining the physical interpretation of the SDOF covariance function, while this is not true of using the kernel in combination with another.

Throughout this work, and highlighted again during the simulated case study, it could be seen that the NMSE and MSLL are not good predictors of damage prediction

accuracy. In the case of NMSE, this is to be somewhat expected, as the metric is simply an assessment of the mean prediction compared to the target data. The MSLL, however, is an assessment of the predictive distribution. While the physics-inspired models that performed better at fatigue damage prediction were also improvements in terms of MSLL, the scores attained by the black-box models (at higher sample rates) also indicated a 'good' prediction according to this metric. As discussed, capturing the measured data within the confidence intervals is not sufficient: the underlying phenomena must be captured to perform well as a predictor of fatigue damage accumulation.

These findings from the simulated case study were confirmed by the Tucano case study. The use of the kernel here achieves the dual goals of improving accuracy and reducing the variance of the damage distribution. Again, it is shown that NMSE and MSLL do not give a good predictor of damage prediction.

While the SDOF kernel performs very well in the work presented here, it is limited by the requirement to have measured data at the training locations, i.e. the model is performing very well in interpolation but cannot, in its current form, be used in extrapolation. In certain situations, and the Tucano case study here is an example of predictive ability previously unavailable, this is not an inherently limiting factor as it has enabled a level of insight not otherwise attainable. However, this requirement will frequently limit the usability of the kernel.

Chapter 8

Conclusions

There is a significant body of literature regarding the fatigue failure of structures, yet fatigue remains the most common cause of the failure of structures. In this work, the following questions were posed:

- How can we use a data-driven framework to predict the fatigue of structures?
- Can using a probabilistic framework help build trust in models when working within a data-driven framework
- Can we use knowledge of physics within such a framework to inform our prognosis?

In this work, a novel probabilistic approach is presented for tackling these questions. Firstly, by fully utilising the probabilistic nature of the Gaussian process regression strain prediction model, a probabilistic view of fatigue damage is achieved. By embracing uncertainty in the modelling approach, it is the authors opinion that there is a pathway to both increase structural safety and also enable a greater lifespan to be attained from the structure. The new probabilistic approach has indeed shown increased robustness to errors in model prediction. People have now starting considering how one can improve data-driven models by inserting physics into their structure. This is some of the first work that considers how physical knowledge can be used in a machine learning environment to predict fatigue damage. This is achieved by inserting knowledge of the dynamics of the problem. Using same probabilistic framework described above, a clear improvement in the predictive capability is visible,

both from the perspective of strain prediction models and also probabilistic damage prediction. Equally important is the fact that the strain prediction model now has physical meaning, and a greater level of trust can be achieved as a result.

In chapters 1 and 2, the overall problem was discussed and scope of this work defined. The fact that the use of data is becoming increasingly common across the engineering industry was discussed, and applications across the structural health monitoring field were presented. In Chapter 2, the scope of this work was defined while reviewing the relevant literature, highlighting that the focus of this work is on the data-driven prediction of structural loads. This work develops a probabilistic methodology for propagating uncertainty of such predictions. It was highlighted that the focus of the probabilistic framework was here, and not on the other sources of uncertainty within the S-N methodology.

In Chapter 3, a on overview of the theory of Gaussian process regression was presented. Following this, the two datasets that are used throughout this work were presented: The first is the Tucano dataset, in which a military aircraft was instrumented with strain gauges and additional accelerometers for a monitoring period. The second is the GARTEUR dataset, a model aircraft structure reproduced based on previous benchmark projects. For this work in this thesis, one of the wings was instrumented with strain gauges in order to predict fatigue accumulation on the structure. Having introduced the data and methodology, a small case study is used to demonstrate the theory and data further using the Tucano dataset. As one of the focuses of this work is to utilise the probabilistic nature of the model, part of the focus of this case study is on interpretation of confidence intervals using a data-driven kernel.

In Chapter 4, the Tucano dataset is used to demonstrate the capability of a data-driven methodology for predicting fatigue damage, while remaining within a deterministic framework. The benefits of input-augmentation was demonstrated by demonstrating the impact of simple adjustments to input data on Tucano data. This was shown to be beneficial even when only using two inputs to the model. Following this, the fleet of 84 aircraft is studied. The potential for using data-driven modelling is demonstrated but the associated risks also highlighted. By using a deterministic prediction, the damage level is predicted for the majority of flights, but in some cases is predicted incorrectly. It is demonstrated here, for the first time in this work, that there is not necessarily a linear relationship between error metrics used for strain (or stress) prediction models, such as normalised mean square error, and the accuracy of

a subsequent S-N fatigue damage prediction.

In Chapter 5, the same Tucano case study was used to demonstrate that by propagating the uncertainty of the model, we can develop a better understanding of the full model prediction. By understanding the uncertainty of the model, the new probabilistic approach has showed increased robustness to errors in model prediction. In doing so, it is the authors opinion that there is a pathway to both increase structural safety and also enable a greater lifespan to be attained from the structure. However, due to the black box nature of the S-N process (and in particular, the rainflow cycle count), the relationship between probabilistic strain prediction model and the resulting damage distribution is not clear. Thus, a controlled case study is set up, where it could be seen that the damage distribution is significantly impacted by the non-linear nature of the relationship between stress and damage (i.e. the non-linearity of S-N curves) and also, the representativeness of the training data to the test data. In particular, the negative effect of employing an overly short lengthscale or excess distance between training and test data is demonstrated. This is the first work to consider the uncertainty of the full stress-time history and the impact of this on fatigue damage prediction.

In chapter 4 and 5, some physical knowledge was introduced into models via input augmentation. However, in Chapter 6, the focus is on the increased use of physical knowledge in the model with respect to fatigue assessment. Motivated by the fact that dynamic behaviour can have significant contribution to fatigue accumulation, a covariance function derived from a single-degree-of-freedom oscillator is used to predict the loads on a simulated case study. The GARTEUR dataset is then used to demonstrate the strain prediction performance on a physical structure, with model uncertainty propagated through the fatigue assessment as per the previous chapters. By using a physics-informed covariance function, the draws from the posterior distribution now had a physical interpretation. For both studies (simulated and the GARTEUR aircraft), the performance of the model with respect to fatigue damage prediction is significantly improved. This is true both when using the model posterior mean and also, the samples drawn from the posterior distribution. However, this was shown to be especially true when reducing the sample rates, where the physics-informed model was able to maintain good predictive capabilities at sample rates below the Nyquist frequency.

In Chapter 7, the same physics-derived covariance function was used in combination

with a data-driven kernel to demonstrate its performance for combined static and dynamic loading. A simulated case study was again used which demonstrated a significant performance improvement compared to a purely-data-driven approach and, again, this was especially true as the sample rate is reduced. Using the same simulated case study, the potential of using a mean function was also presented. In this case, as the static of component of the data could be easily estimated with good accuracy, the benefit of removing this component of the data from the probabilistic framework was discussed. Following this, the Tucano dataset was used, this time to predict strain at the full sample rate available of 128 Hz. By using the single-degree-of-freedom covariance function in combination with a data-driven (in this case, a squared exponential) kernel, it was possible to capture high frequency strain measurements with a high-degree of accuracy. The resulting fatigue prediction showed a significant increase in accuracy.

8.1 A practical view: limitations and future work

The methodology proposed in this thesis has enabled an insight to be gained on the fatigue damage state of structures which may otherwise be difficult. In chapters 4 and 5, looking at the Tucano aircraft, utilising other methods such as finite element modelling would have been very difficult, as understanding the forces acting on the aircraft would have been challenging (in addition to the large computational expense that would have been required for analysing time-series data for all 84 test flights).

By introducing physics into the modelling approach, significant improvement has been achieved in Chapters 6 and 7. This is true both from a strain prediction, and fatigue damage, perspective. In this work, accurate fatigue damage prediction has been achieved at sample rates significantly below sample rates typically considered the minimum required for effective fatigue damage prediction, while maintaining good levels of accuracy. Even at sample rates below the Nyquist frequency, fatigue damage prediction has remained acceptable in many cases. Furthermore, in doing so, this work has provided a valuable real-world example of the potential benefits of physics-informed machine learning.

While GP regression has been the form of data-driven model used throughout this work, it is the author's opinion that many of the findings and the methods can be used

for other methods of data-driven learning. Naturally, the specific methodology for e.g. drawing samples from strain distributions would change, but the work contained in this thesis has demonstrated the value of propagating time-series prediction uncertainty.

Before the practical implementation of such a methodology, however, a more comprehensive evaluation of how to use such probabilistic predictions will be required. For example, it is worth considering how one could use the damage distributions attained to justify further modelling in certain cases: one way of achieving this is could be to set a threshold of damage variance that, when passed, justifies a more detailed study into the given flight. The results of the Tucano dataset were achieved with only one training set of 2000 data points from across the training flights. Yet, the study in Chapter 5 into the cause of shifts in the damage distribution highlighted the importance of having representative training data. So, if a hypothetical threshold of damage variance were to be exceeded, there could be avenue to 'cherry pick' training data that is more representative of that flight.

One downside of the method developed in this thesis is that, to fully utilise this probabilistic form, drawing samples from the GP posterior is computationally expensive. The cost sampling from the Gaussian process posterior is $\mathcal{O}(n^3)$, where n is the number of testing points. It is for this region that the number of testing points was limited for both the Tucano and GARTEUR datasets. In recent years, research has been conducted to develop more efficient ways of sampling from the posterior. Future work shall include testing methods, such as that developed by Wilson et al. [136] to see if any accuracy is lost in the damage prediction by utilising such efficient sampling techniques.

Another criticism of the method in its current form is that the work in Chapters 6 and 7, thus far, is limited to interpolation in time for prediction. While the performance for acquiring useful damage prediction from (relatively) low frequency data was very good, this limits its usefulness. Future work will, therefore, look to extend this to extrapolation in both the spatial and temporal domains. This would enable use as a virtual sensor, as per chapters 4 and 5, and improve the usability of the approach significantly. Furthermore, while the work in Chapters 6 and 7 did lead to increasingly complex systems being predicted, this could be extended further. The single-degree-of-freedom covariance function treats the system deterministically with stochastic forcing. Extending this to higher degrees of freedom with non-linear

material behaviours would require treating the system stochastically.

Normalised mean square error and mean square log-loss are used throughout this work as a mean for assessing the model error. Yet throughout, we have seen that they are not particularly successful at predicting the fatigue damage accumulation of the structure (or, equally, the underlying process that the structure is withstanding). Thus, another area that warrants further research is the development of cost functions and error metrics to better optimise, and assess, models. By developing cost functions and error metrics that prioritise predictive capability with respect to fatigue damage, it is proposed that this could lead to improved damage prediction and thus, better real-world applicability. Similarly, by utilising methods that optimise the strain prediction models based on posterior samples, as opposed to the prediction mean or predictive distribution, it is proposed that the shifts often seen in the damage distribution throughout this work could be mitigated.

Lastly, it was discussed in the introduction that probabilistic methods are seen as an important tool in reducing the remaining useful life at the point of retirement of structures. In this work, the probabilistic framework considered only uncertainty from the load prediction, while acknowledging that there are many other sources of uncertainty in stress-life fatigue. A natural item of future work will be to integrate the methodology developed in this work with other probabilistic methodologies. This will be a crucial step in working towards the goal stated in the first page of this thesis: to simultaneously increase structural safety and enable a greater lifespan to be attained from structures.

Bibliography

[1] W. Cui. A state-of-the-art review on fatigue life prediction methods for metal structures. *Journal of marine science and technology*, 7:43–56, 2002.

[2] C. R. Gagg and P. R. Lewis. In-service fatigue failure of engineered products and structures–case study review. *Engineering Failure Analysis*, 16(6):1775–1793, 2009.

[3] P. Wirsching. Probabilistic fatigue analysis. In *Probabilistic structural mechanics handbook*, pages 146–165. Springer, 1995.

[4] M. Tapavicza and F. Och. Application of Damage Tolerance Concepts for MBB Helicopters. In *AGARD Conference Proceedings: Helicopter Fatigue Life Assessment*, volume 297, pages 4.1 – 4.9, Neuilly Sur Seine, France, 1980. NATO.

[5] J. Schijve. *Fatigue of Structures and Materials*. Springer Science+Business Media, 2nd edition, 2008. ISBN 9781402068072. doi: 10.1007/978-1-4020-6808-9_12.

[6] M. A. Miner. Cumulative damage in fatigue. *Journal of Applied Mechanics*, 1945.

[7] P. C. Paris. A rational analytic theory of fatigue. *Trends Engin*, 13:9–14, 1961.

[8] P. Paris and F. Erdogan. A critical analysis of crack propagation laws. 1963.

[9] M. Matsuishi and T. Endo. Fatigue of metals subjected to varying stress. *Japan Society of Mechanical Engineers, Fukuoka, Japan*, 68(2):37–40, 1968.

[10] P. Jia, H. Liu, C. Zhu, W. Wu, and G. Lu. Contact fatigue life prediction of a bevel gear under spectrum loading. *Frontiers of Mechanical Engineering*, 15: 123–132, 2020.

[11] M. Rodzewicz and A. Przekop. Experimental investigation of the load spectrum and fatigue tests of the pw-5 world class glider. *Technical soaring*, 24(1):15–20, 2000.

[12] Y. Murakami. Material defects as the basis of fatigue design. *International Journal of Fatigue*, 41:2–10, 2012.

[13] H. Soyama. Comparison between the improvements made to the fatigue strength of stainless steel by cavitation peening, water jet peening, shot peening and laser peening. *Journal of Materials Processing Technology*, 269:65–78, 2019.

[14] J. Pan and S. H. Lin. Fracture Mechanics and Fatigue Crack Propagation. In *Fatigue Testing and Analysis*, chapter 6, pages 237–284. Butterworth Heinemann, 2005. ISBN 9780750677196. doi: 10.1016/B978-075067719-6/50007-5.

[15] D. L. McDowell. Basic issues in the mechanics of high cycle metal fatigue. *International Journal of Fracture*, 80(2):103–145, 1996. ISSN 15732673. doi: 10.1007/BF00012666.

[16] J. Hoole, P. Sartor, and J. Cooper. Safe-Life Fatigue and Sensitivity Analysis : A Pathway Towards Embracing Uncertainty. In *5th Aircraft Structural Design Conference*. Royal Aeronautical Society, 2016.

[17] G. Liu, D. Wang, and Z. Hu. Application of the rain-flow counting method in fatigue. In *2nd International Conference on Electronics, Network and Computer Engineering (ICENCE 2016)*, pages 236–240. Atlantis Press, 2016.

[18] D. Benasciutti and R. Tovo. Rainflow cycle distribution and fatigue damage in gaussian random loadings. *Report of engineering department. University of Ferrara (Italy)*, 2004.

[19] C. R. Farrar and K. Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):303–315, 2007.

[20] A. Rytter. Vibrational based inspection of civil engineering structures. 1993.

[21] A. Haldar. *Health assessment of engineered structures: bridges, buildings and other infrastructures*. World Scientific, 2013.

[22] C. R. Farrar and K. Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley and Sons, nov 2012. ISBN 9781119994336. doi: 10.1002/9781118443118.

[23] R. Ross. Integrated vehicle health management in aerospace structures. In F.-G. Yuan, editor, *Structural Health Monitoring (SHM) in aerospace structures*, chapter 1, pages 3–66. Elsevier Science & Technology, 2016. ISBN 978-0-08-100148-6. doi: https://doi.org/10.1016/C2014-0-00994-X.

[24] H. Rocha, C. Semprimoschnig, and J. P. Nunes. Sensors for process and structural health monitoring of aerospace composites: A review. *Engineering Structures*, 237(July 2020), 2021. ISSN 18737323. doi: 10.1016/j.engstruct.2021.112231.

[25] B. Main, R. Evans, K. Walker, X. Yu, and L. Molent. Lessons from a fatigue prediction challenge for an aircraft wing shear tie post. *International journal of fatigue*, 123:53–65, 2019.

[26] G. Ersdal. Assessment of existing offshore structures for life extension. *Department of Mechanical and Structural Engineering and Material Sciences, University of Stavanger, Norway*, 2005.

[27] T. Rubert, D. McMillan, and P. Niewczas. A decision support tool to assist with lifetime extension of wind turbines. *Renewable energy*, 120:423–433, 2018.

[28] L. Ziegler, U. Smolka, N. Cosack, and M. Muskulus. Brief communication: Structural monitoring for lifetime extension of offshore wind monopiles: can strain measurements at one level tell us everything? *Wind Energy Science*, 2 (2):469–476, 2017.

[29] K. Anup, J. Whale, S. P. Evans, and P. D. Clausen. An investigation of the impact of wind speed and turbulence on small wind turbine operation and fatigue loads. *Renewable Energy*, 146:87–98, 2020.

[30] K. Worden and G. Manson. The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):515–537, 2007.

[31] W. Zhao, A. Stacey, and P. Prakash. Probabilistic models of uncertainties in fatigue and fracture reliability analysis. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 36134, pages 557–578, 2002.

[32] B. Sudret, P. Hornet, J.-M. Stephan, Z. Guede, and M. Lemaire. Probabilistic assessment of fatigue life including statistical uncertainties in the sn curve. 2003.

[33] J. Hoole, P. Sartor, J. Booker, J. Cooper, X. V. Gogouvitis, A. Ghouali, and R. K. Schmidt. Probabilistic Fatigue Methodology for Safe-Life Design and Analysis. Technical report, DSTL, 2021.

[34] L. Xie, J. Liu, N. Wu, and W. Qian. Backwards statistical inference method for p–s–n curve fitting with small-sample experiment data. *International Journal of Fatigue*, 63:62–67, 2014.

[35] F. Bastenaire. New method for the statistical evaluation of constant stress amplitude fatigue-test results. In *Probabilistic aspects of fatigue*. ASTM International, 1972.

[36] C. Hübler, C. G. Gebhardt, and R. Rolfes. Methodologies for fatigue assessment of offshore wind turbines considering scattering environmental conditions and the uncertainty due to finite sampling. *Wind Energy*, 21(11):1092–1105, 2018.

[37] S. C. Saunders. A probabilistic interpretation of miner's rule. ii. *SIAM Journal on Applied Mathematics*, 19(1):251–265, 1970.

[38] K. Miller, H. Mohamed, and E. De Los Rios. Fatigue damage accumulation above and below the fatigue limit. *Mechanical Engineering Publications, The Behaviour of Short Fatigue Cracks,*, pages 491–511, 1986.

[39] W. Fricke. Fatigue analysis of welded joints: state of development. *Marine structures*, 16(3):185–200, 2003.

[40] W. Zhao and M. J. Baker. On the probability density function of rainflow stress range for stationary gaussian processes. *International Journal of Fatigue*, 14(2):121–135, 1992.

[41] Z. Li, T. H. Chan, and J. M. Ko. Fatigue analysis and life prediction of bridges with structural health monitoring data—part i: methodology and strategy. *International Journal of Fatigue*, 23(1):45–53, 2001.

[42] E. Altus. Fatigue, fractals, and a modified miner's rule. 1991.

[43] J. Velarde, A. Mankar, C. Kramhøft, and J. D. Sørensen. Probabilistic calibration of fatigue safety factors for offshore wind turbine concrete structures. *Engineering Structures*, 222:111090, 2020.

[44] G. Cavallini and R. Lazzeri. A probabilistic approach to fatigue risk assessment in aerospace components. *Engineering Fracture Mechanics*, 74(18):2964–2970, 2007.

[45] N. Zentuti, J. Booker, J. Hoole, R. Bradford, and D. Knowles. Probabilistic structural integrity. *FESI Bulletin, FESI Publishing, Preston*, 12:16–23, 2018.

[46] J. Clausen, S. O. Hansson, and F. Nilsson. Generalizing the safety factor approach. *Reliability Engineering & System Safety*, 91(8):964–973, 2006.

[47] R. A. Heller. *The Symposium on Probabilistic Aspects of Fatigue*. ASTM International, 1971.

[48] S. Saunders. *On the probabilistic determination of scatter factors using Miner's rule in fatigue life studies*. ASTM International, 1972.

[49] C. Sundararajan and F. Witt. Stress-strength interference method. In *Probabilistic Structural Mechanics Handbook*, pages 8–26. Springer, 1995.

[50] J. Hoole. *Probabilistic fatigue methodology for aircraft landing gear*. PhD thesis, University of Bristol, 2020.

[51] N. Lindsey, J. Dawson, D. Sheldon, A. DiVenti, and L. Sindjui. Nasa physics of failure (pof) for reliability. In *PSAM-16 Conference (June 2022) Abstract only submittal.*, 2022.

[52] A. J. Hughes, R. J. Barthorpe, N. Dervilis, C. R. Farrar, and K. Worden. A probabilistic risk-based decision framework for structural health monitoring. *Mechanical Systems and Signal Processing*, 150:107339, 2021.

[53] W. Keller and M. Modarres. A historical overview of probabilistic risk assessment development and its use in the nuclear power industry: a tribute to the late professor norman carl rasmussen. *Reliability Engineering & System Safety*, 89(3):271–285, 2005.

[54] W. E. Vesely. Probabilistic risk assessment. *System Health Management: With Aerospace Applications*, pages 253–263, 2011.

[55] M. Stamatelatos, H. Dezfuli, G. Apostolakis, C. Everline, S. Guarro, D. Mathias, A. Mosleh, T. Paulos, D. Riha, C. Smith, *et al.* Probabilistic risk assessment procedures guide for nasa managers and practitioners. Technical report, 2011.

[56] C. R. Farrar, S. W. Doebling, and D. A. Nix. Vibration–based structural damage identification. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 359(1778): 131–149, 2001.

[57] M. Martinez Luengo and A. Kolios. Failure mode identification and end of life scenarios of offshore wind turbines: A review. *Energies*, 8(8):8339–8354, 2015.

[58] N. Bishop and M. Frimley. Finite element based fatigue analysis. In *Americas User Conference, Oct*, pages 5–9, 1998.

[59] A. Beis. A finite element analysis of the nps autonomous underwater vehicle (auv) hull intended to operate in deep waters. Technical report, NAVAL POSTGRADUATE SCHOOL MONTEREY CA DEPT OF MECHANICAL ENGINEERING, 2001.

[60] V. Wakchaure and G. S. Rao. International journal of engineering sciences & research technology cae analysis of automotive lamp.

[61] D. Yu, A. Al-Yafawi, T. T. Nguyen, S. Park, and S. Chung. High-cycle fatigue life prediction for pb-free bga under random vibration loading. *Microelectronics Reliability*, 51(3):649–656, 2011.

[62] D. Taylor, N. Barrett, and G. Lucano. Some new methods for predicting fatigue in welded joints. *International Journal of Fatigue*, 24(5):509–518, 2002.

[63] W. Schütz. Fatigue life prediction of aircraft structures—past, present and future. *Engineering Fracture Mechanics*, 6(4):745–762, 1974.

[64] D. Hajializadeh, E. J. OBrien, and A. J. O'Connor. Virtual structural health monitoring and remaining life prediction of steel bridges. *Canadian Journal of Civil Engineering*, 44(4):264–273, 2017.

[65] T. Guo, D. M. Frangopol, and Y. Chen. Fatigue reliability assessment of steel bridge details integrating weigh-in-motion data and probabilistic finite element analysis. *Computers & Structures*, 112:245–257, 2012.

[66] D. Giagopoulos, A. Arailopoulos, V. Dertimanis, C. Papadimitriou, E. Chatzi, and K. Grompanopoulos. Structural health monitoring and fatigue damage estimation using vibration measurements and finite element model updating. *Structural Health Monitoring*, 18(4):1189–1206, 2019.

[67] C. Papadimitriou, E. Ntotsios, D. Giagopoulos, and S. Natsiavas. Variability of updated finite element models and their predictions consistent with vibration measurements. *Structural Control and Health Monitoring*, 19(5):630–654, 2012.

[68] F. Xu, C. Li, T. Jiang, and D. Zhang. Fatigue life prediction for pbga under random vibration using updated finite element models. *Experimental Techniques*, 40:1421–1435, 2016.

[69] M. Moatamedi and H. A. Khawaja. *Finite element analysis*. CRC Press, 2018.

[70] A. Halfpenny. A frequency domain approach for fatigue life estimation from finite element analysis. *Key Engineering Materials*, 167:401–410, 1999.

[71] Y. Gao and Y. M. Low. An efficient importance sampling method for long-term fatigue assessment of deepwater risers with time domain analysis. *Probabilistic Engineering Mechanics*, 45:102–114, 2016.

[72] X.-P. Niu, R.-Z. Wang, D. Liao, S.-P. Zhu, X.-C. Zhang, and B. Keshtegar. Probabilistic modeling of uncertainties in fatigue reliability analysis of turbine bladed disks. *International Journal of Fatigue*, 142:105912, 2021.

[73] C. Li, S. Mahadevan, Y. Ling, S. Choze, and L. Wang. Dynamic Bayesian network for aircraft wing health monitoring digital twin. *AIAA Journal*, 55(3): 930–941, 2017. ISSN 00011452. doi: 10.2514/1.J055201.

[74] S. Chakraborty, S. Adhikari, and R. Ganguli. The role of surrogate models in the development of digital twins of dynamic systems. *Applied Mathematical Modelling*, 90:662–681, 2021. ISSN 0307904X. doi: 10.1016/j.apm.2020.09.037. URL `https://doi.org/10.1016/j.apm.2020.09.037`.

[75] Y. Gao, Q. Liu, and D. Zhao. Kriging surrogate model based optimisation of welded bogie frame for fatigue improvement. *International Journal of Vehicle Structures & Systems*, 11(4):349–354, 2019.

[76] B. Echard, N. Gayton, M. Lemaire, and N. Relun. A combined importance sampling and kriging reliability method for small failure probabilities with time-demanding numerical models. *Reliability Engineering & System Safety*, 111:232–240, 2013.

[77] R. Teixeira, M. Nogal, A. O'Connor, J. Nichols, and A. Dumas. Stress-cycle fatigue design with kriging applied to offshore wind turbines. *International Journal of Fatigue*, 125:454–467, 2019.

[78] N. S. Gulgec, M. Takáč, and S. N. Pakzad. Structural sensing with deep learning: Strain estimation from acceleration data for fatigue assessment. *Computer-Aided Civil and Infrastructure Engineering*, 35(12):1349–1364, 2020.

[79] D. Hess, W. Faller, W. Smith, and T. Huang. Neural networks as virtual sensors. In *37th Aerospace Sciences Meeting and Exhibit*, page 259, 1999.

[80] S. Reed. A parametric-based empennage fatigue monitoring system using artificial neural networks. In *Proceedings of the 23rd Symposium of the International Committee on Aeronautical Fatigue*, pages 693–704. DGLR-Breicht Hamburg, 2005.

[81] S. Reed. Development of a parametric-based indirect aircraft structural usage monitoring system using artificial neural networks. *The Aeronautical Journal*, 111(1118):209–230, 2007.

[82] B. Mora, J. Basurko, I. Sabahi, U. Leturiondo, and J. Albizuri. Strain virtual sensing for structural health monitoring under variable loads. *Sensors*, 23(10): 4706, 2023.

[83] E. Risaliti. Model based virtual sensors for wheel center loads and full strain field on vehicle suspension components. 2019.

[84] J. Bilbao, E.-M. Lourens, A. Schulze, and L. Ziegler. Virtual sensing in an onshore wind turbine tower using a gaussian process latent force model. *Data-Centric Engineering*, 3:e35, 2022.

[85] J. Zou, E.-M. Lourens, and A. Cicirello. Virtual sensing of subsoil strain response in monopile-based offshore wind turbines via gaussian process latent force models. *Mechanical Systems and Signal Processing*, 200:110488, 2023.

[86] R. Fuentes, E. Cross, A. Halfpenny, K. Worden, and R. J. Barthorpe. Aircraft parametric structural load monitoring using gaussian process regression. In *EWSHM-7th European workshop on structural health monitoring*, 2014.

[87] L. Ziegler, N. Cosack, A. Kolios, and M. Muskulus. Structural monitoring for lifetime extension of offshore wind monopiles: Verification of strain-based load extrapolation algorithm. *Marine Structures*, 66:154–163, 2019.

[88] L. D. Avendaño-Valencia, I. Abdallah, and E. Chatzi. Virtual fatigue diagnostics of wake-affected wind turbine via gaussian process regression. *Renewable Energy*, 170:539–561, 2021.

[89] G. Holmes, P. Sartor, S. Reed, P. Southern, K. Worden, and E. Cross. Prediction of landing gear loads using machine learning techniques. *Structural Health Monitoring*, 15(5):568–582, 2016.

[90] E. Cross, P. Sartor, K. Worden, and P. Southern. Prediction of landing gear loads using machine learning techniques. *Structural Health Monitoring*, pages 1056–1063, 2012.

[91] M. T. Manry, C. H. Hsieh, and H. Chandrasekaran. Near-optimal flight load synthesis using neural nets. In *Neural Networks for Signal Processing - Proceedings of the IEEE Workshop*, pages 535–544, 1999. ISBN 078035673X. doi: 10.1109/nnsp.1999.788173.

[92] J. J. Valdes, C. Cheung, and A. L. Rubio. Extreme learning machines to approximate low dimensional spaces for helicopter load signal and fatigue life estimation. In *2017 International Joint Conference on Neural Networks (IJCNN)*, volume 2017-May. IEEE, 2017. ISBN 9781509061815. doi: 10.1109/IJCNN.2017.7966095.

[93] H. Wang, B. Li, J. Gong, and F.-Z. Xuan. Machine learning-based fatigue life prediction of metal materials: Perspectives of physics-informed and data-driven hybrid methods. *Engineering Fracture Mechanics*, page 109242, 2023.

[94] E. J. Cross, S. Gibson, M. Jones, D. Pitchforth, S. Zhang, and T. Rogers. Physics-informed machine learning for structural health monitoring. In *Structural Health Monitoring Based on Data Science Techniques*, pages 347–367. Springer, 2022.

[95] H. Azzam. A practical approach for the indirect prediction of structural fatigue from measured flight parameters. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 211(1):29–38, 1997. ISSN 09544100. doi: 10.1243/0954410971532479.

[96] D. J. Pitchforth, T. J. Rogers, U. T. Tygesen, and E. J. Cross. Grey-box models for wave loading prediction. *Mechanical Systems and Signal Processing*, 159:107741, 2021.

[97] H.-P. Wan and W.-X. Ren. A residual-based gaussian process model framework for finite element model updating. *Computers & Structures*, 156:149–159, 2015.

[98] G. Ersdal, J. D. Sørensen, and I. Langen. Updating of structural failure probability based on experienced wave loading. In *The Thirteenth International Offshore and Polar Engineering Conference*. OnePetro, 2003.

[99] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[100] W. Hao, L. Tan, X. Yang, D. Shi, M. Wang, G. Miao, and Y. Fan. A physics-informed machine learning approach for notch fatigue evaluation of alloys used in aerospace. *International Journal of Fatigue*, 170:107536, 2023.

[101] E. Salvati, A. Tognan, L. Laurenti, M. Pelegatti, and F. De Bona. A defect-based physics-informed machine learning framework for fatigue finite life prediction in additive manufacturing. *Materials & Design*, 222:111089, 2022.

[102] A. Dourado and F. A. Viana. Physics-informed neural networks for missing physics estimation in cumulative damage models: a case study in corrosion fatigue. *Journal of Computing and Information Science in Engineering*, 20(6): 061007, 2020.

[103] A. Dourado and F. A. Viana. Physics-informed neural networks for corrosion-fatigue prognosis. In *Proceedings of the Annual Conference of the PHM Society*, volume 11, 2019.

[104] T. Zhou, S. Jiang, T. Han, S.-P. Zhu, and Y. Cai. A physically consistent framework for fatigue life prediction using probabilistic physics-informed neural network. *International Journal of Fatigue*, 166:107234, 2023.

[105] N. HajiGhassemi and M. P. Deisenroth. Approximate inference for long-term forecasting with periodic gaussian processes. In *AISTATS*, 2014.

[106] E. J. Cross and T. J. Rogers. Physics-derived covariance functions for machine learning in structural dynamics. *IFAC-PapersOnLine*, 54(7):168–173, 2021.

[107] B. M. Ayyub and R. H. Mccuen. Simulation-based reliability methods. In *Probabilistic structural mechanics handbook*, pages 53–69. Springer, 1995.

[108] L. Nie, W. Wang, L. Deng, and W. He. Ann and lefm-based fatigue reliability analysis and truck weight limits of steel bridges after crack detection. *Sensors*, 22(4), 2022.

[109] H.-P. Wan and W.-X. Ren. Parameter selection in finite-element-model updating by global sensitivity analysis using gaussian process metamodel. *Journal of Structural Engineering*, 141(6):04014164, 2015.

[110] J. Kullaa. Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring. *Mechanical Systems and Signal Processing*, 25(8):2976–2989, 2011.

[111] J. Hensman, R. Mills, S. Pierce, K. Worden, and M. Eaton. Locating acoustic emission sources in complex structures using gaussian processes. *Mechanical Systems and Signal Processing*, 24(1):211–223, 2010.

[112] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[113] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999.

[114] D. Duvenaud. *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge, 2014.

[115] M. Morgan. Aircraft fighter plane embraer emb-314 super tucano brazilian air force 1, 2009. URL `https://flickr.com/photos/24656754@N00/3864956510`. [Online; accessed December 2023].

[116] S. C. Reed. *Indirect Aircraft Structural Monitoring Using Artificial Neural Networks*. PhD thesis, The University of Sheffield, 2006.

[117] M. Degener and M. Hermes. Ground vibration test and finite element analysis of the garteur sm-ag19 testbed. 1996.

[118] E. Balmes and J. R. Wright. Garteur group on ground vibration testing: results from the test of a single structure by 12 laboratories in europe. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 80401, page V01AT03A004. American Society of Mechanical Engineers, 1997.

[119] M. Degener. Ground vibration tests on an aircraft model performed as part of a european round robin exercise. In *International Forum on Aeroelasticity and Structural Dynamics, 17-20 June 1997, Rome, Italy*, volume 3, pages 255–262, 1997.

[120] M. Link and M. Friswell. Working group 1: generation of validated structural dynamic models—results of a benchmark study utilising the garteur sm-ag19 test-bed. *Mechanical Systems and Signal Processing*, 17(1):9–20, 2003.

[121] O. Basquin. The exponential law of endurance tests. In *Proc Am Soc Test Mater*, volume 10, pages 625–630, 1910.

[122] ASTM. E1049-85: Standard practices for cycle counting in fatigue analysis. *ASTM Standard*, 85(Reapproved 2017):1–10, 2017. doi: 10.1520/E1049-85R17.2.

[123] M. Mitchell. Fundamentals of modern fatigue analysis for design. 1996.

[124] B. Cowles. High cycle fatigue in aircraft gas turbines—an industry perspective. *International Journal of Fracture*, 80:147–163, 1996.

[125] J. Goodman. *Mechanics applied to engineering*. Longmans, Green, 1919.

[126] H. Gerber. *Bestimmung der zulässigen spannungen in eisen-constructionen*. Wolf, 1874.

[127] C. R. Soderberg. Factor of safety and working stress. *Transactions of the American Society of Mechanical Engineers*, 52(2):13–21, 1930.

[128] S. Reed. A parametric-based empennage fatigue monitoring system using artificial neural networks. Technical report, QinetiQ, 2005.

[129] D. J. MacKay *et al.* Bayesian nonlinear modeling for the prediction competition. *ASHRAE transactions*, 100(2):1053–1062, 1994.

[130] P. D'Antuono, W. Weijtjens, and C. Devriendt. On the minimum required sampling frequency for reliable fatigue lifetime estimation in structural health monitoring. how much is enough? In *European Workshop on Structural Health Monitoring*, pages 133–142. Springer, 2023.

[131] A. Papoulis. Random variables and stochastic processes. 1965.

[132] T. Caughey. Nonlinear theory of random vibrations. *Advances in applied mechanics*, 11:209–253, 1971.

[133] H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

[134] L. Zhang, R. Brincker, and P. Andersen. Modal indicators for operational modal identification. In *Proceedings of IMAC 19: A Conference on Structural Dynamics: februar 5-8, 2001, Hyatt Orlando, Kissimmee, Florida, 2001*, pages 746–752. Society for Experimental Mechanics, 2001.

[135] C. Lalanne. *Fatigue Damage: Mechanical Vibration and Shock Analysis.* ITSE, 3 edition, 2002. ISBN 978-1-84821-647-1.

[136] J. Wilson, V. Borovitskiy, A. Terenin, P. Mostowsky, and M. Deisenroth. Efficiently sampling functions from gaussian process posteriors. In *International Conference on Machine Learning*, pages 10292–10302. PMLR, 2020.