Monitoring a population of structures from a spatiotemporal perspective: An application to offshore wind farms.

A Thesis submitted to the University of Sheffield

for the degree of Doctor of Philosophy in the Faculty of Engineering

by

Weijiang Lin

Department of Mechanical Engineering

University of Sheffield

September 2023

Acknowledgements

I wish to express my considerable gratitude to many people with whom I have interacted during the work for this thesis. In particular, my supervisor Professor Elizabeth Cross (Lizzy) has been continuously supporting, motivating and supervising my work. Her patience, enthusiasm, and immense knowledge have inspired me to develop a deeper interest in research. Lizzy has also been a role model and emotional support throughout my journey, and I am really grateful that I have been working by her side during this period of intense personal development. I would also like to thank Professor Keith Worden for being another supervisor of mine, who has been sharing with me his endless reservoir of knowledge together with his impeccable sense of humour. His supportive attitude and continuous advice have also encouraged me throughout this project.

I also owe a debt of gratitude to Matt Jones, who has read parts of the work and provided valuable feedback. All the other members of the DRG have also been generous in providing useful comments and mental support throughout the journey. I am grateful to Yangli Liu, Tingna Wang, Feiyi Duan, Ronnie Zhao, and Elaine Ng for sharing motivating conversations with me during my work towards this thesis. Finally, I reserve my deepest appreciation to my family for their patience and understanding during this period.

Abstract

Population-based structural health monitoring (PBSHM) is an emerging field that is motivated by the lack of labelled data necessary for developing diagnostic tools that assess the type and severity of damage in structures. A key theme in PBSHM is to share data across multiple structures within a population whenever possible. Recent advances in the field focus on methods of knowledge transfer through abstract spaces. This thesis concerns PBSHM for a population of structures whose behaviours are spatiotemporally correlated, for example, an offshore wind farm. A main aim of the work is to utilise the spatiotemporal correlation across a population for the development of physics-inspired, population-level methods for structural condition assessment. Within the scope of the thesis, the focus is to explore how the knowledge of spatiotemporal correlations may contribute to the task of wind farm monitoring.

A key contribution of the thesis is the development of a population-level anomaly detector that accounts for the spatiotemporal correlation between structures, known as the mapping method. The mapping method consists of a spatiotemporal model, which predicts the normal undamaged condition of a population, and a detection criterion, based on which error maps are created to visualise the detection of anomalies.

With regard to the modelling techniques, two approaches are developed to capture the spatiotemporality across a wind farm, both of which are based on Gaussian process (GP) regression. The first approach applies GP regression directly to model the turbine performance across the entire wind farm, capturing the spatiotemporal variation through its covariance function. The second approach models the wind farm performance as a spatial autoregressive process, which leads to the development of the Gaussian process spatial autoregressive model with exogenous inputs (GP-SPARX).

The model structure of GP-SPARX is motivated and optimised by considering the physics of wind farms; in this sense, the GP-SPARX is a physics-inspired model.

The developed methods are demonstrated with simulated examples and the data from an operating offshore wind farm. The results show that population-level anomaly detection is viable using a spatiotemporal approach, which is worth investigating into as part of the ongoing research in PBSHM.

Contents

Chapter 1

Introduction

This work records an investigation into the modelling methods for population-based structural health monitoring from a spatiotemporal perspective. The focus is on understanding the spatiotemporal dependence in the monitoring data collected across a population of structures and how new methods are developed to account for such information for monitoring purposes.

1.1 The Task of Structural Health Monitoring

In the simplest terms, structural health monitoring (SHM) is the monitoring of structure conditions. The structure may refer to any structural or mechanical system, such as an aircraft, a bridge, a building, or a wind turbine. The need for health assessments arose initially to extend the service life of existing structures that are approaching the end of their design lifespans. With several critical advances made in the field of SHM over the past three decades, monitoring strategies have also become parts of the considerations made at the design phase and have been implemented to continuously assess the conditions of a growing number of new structures. Although a large amount of effort is put into developing structure-specific monitoring strategies, the fundamental task of SHM is the same across a wide range of methods. This task of SHM has been firstly summarised as the Rytter's Hierarchy [1] and later augmented into a five-step process [2]. The steps include:

1. Detection: detection of whether damage/anomaly exists in a structure, some-

times including the specification of the onset of such damage/anomaly

2. Localisation: determination of where damage/anomaly occurs in the structure, typically referring to the physical location of damage/anomaly
3. Classification: assessment of the type of damage/anomaly
4. Quantification: evaluation of the level of severity of damage/anomaly
5. Prognosis: prediction of the remaining useful life of the structure

The meeting of all five objectives is considered to have provided a "full picture" of the health or condition of a structure. Nonetheless, information obtained from any of these steps is likely to aid the decision making regarding the operation and maintenance of the structure.

1.2 Population-Based Structural Health Monitoring

Population-based structural health monitoring (PBSHM) is an emerging branch in the field of SHM that aims to explore the combined use of data sourced from multiple structures [3–11]. It is primarily motivated by a key challenge in SHM — the scarcity of data, especially damage-state data. Towards the higher levels of SHM, e.g. classification and quantification, labelled data are required to train data-based models to assess damage types and severity. However, the data collected from a single, operating structure are likely to cover, if at all, a limited range of potential damage conditions. Experiments for damage-state data collection can be economically unviable to carry out on operating structures, especially for high-cost structures such as wind turbines. On the other hand, the available damage-state data for SHM purposes are mostly collected from low-cost scale models or retired structures in a tightly-controlled laboratory environment. PBSHM addresses this issue of data scarcity by encouraging knowledge transfer between individual structures of the same type or between different types of structures.

The main idea of PBSHM boils down to knowledge transfer, with the first step being the search for a representative subset of the entire population. This first attempt at PBSHM is found in [12, 13], where the features extracted from an individual turbine

are used to make predictions across other turbines in an offshore wind farm. In this example, an individual structure is assumed sufficient to represent the entire population, given that the turbines are very similar across the wind farm; inferences are then transferred across the population. Building on this simple example, more studies are carried out to formalise the definition of the representative subset. Before introducing the definition, the types of population are to be specified, as they determine what can be a representative subset and how knowledge can be transferred through it.

A population can be homogeneous or heterogeneous based on how similar the members of the population are to each other. A homogeneous population consists of a group of structures that are nominally-identical with respect to geometry, topology, and material [4, 5, 14]. An offshore wind farm can be an example of a homogeneous population given that the turbines are of the same design, configuration, ISO standards etc. Variations are inevitable across a population; two types of variations are allowed in the definition of a homogeneous population. The first type of variation lies in the embodiment of the structures, e.g. minor imperfections within the manufacturing tolerances. Other variations may arise from the environmental and operational conditions, such as different temperature variations, boundary conditions at the seabed, aerodynamic properties of the local wind environment, and differing operational settings. The population *form* is defined to represent the essence (in an abstract sense) of a homogeneous population as well as the variations encountered in reality [5, 14]. To build upon the earlier attempt that uses a single structure as a representative subset of the population, the representative subset is now the population form, which is defined to not necessarily be the structure itself but be a feature that represents the population behaviour; hereby, SHM inferences are transferred across the members of a homogeneous population, with more details provided in [14].

A heterogeneous population refers to any group of structures that are not nominally-identical. The behaviour of a heterogeneous population can be represented as a model that generalises across the entire population [15]. To train such a model, the first step is to determine what knowledge can be transfered between which structures, i.e. an assessment is required to quantify how different/similar the structures are in the heterogeneous population; this is done by, firstly, decomposing the structures into components that demonstrate well-established dynamic behaviour (e.g. beams and plates) and, secondly, looking for the common substructure among the disparate

structures [16]. It allows one to determine the region (at a substructure level), in which knowledge transfer is reasonable, and gives insight into the scope of the SHM problems solvable via this knowledge transfer. The next step is to present the behaviour of the population as a "common feature" obtained by mapping data associated with different structures onto a common latent space, which is achieved using techniques of domain adaptation [15]. A later study approaches the task of knowledge transfer through the lens of modern physics, while accounting for the confounding influence at the same time [17].

As the structures in a population become more disparate, it is more difficult and, thus, less urgent to find a representation of the population behaviour. Hence, the focus of many recent studies has shifted towards the actual methods of knowledge transfer from the source structure(s) to a target structure in various kinds of abstract spaces [15, 17–20].

1.3 PBSHM for Wind Turbine Farms

In light of the current climate situation and power crisis, urgent actions are needed to accelerate the energy transition from fossil fuels to renewables. Wind power is a key type of renewable source and currently constitutes 9% of the world's electricity. Although the offshore wind sector has been growing rapidly over the past few years, the rate of offshore wind installations needs to increase by a factor of four in order to meet the objectives of net zero by 2050 [21]. The rapid expansion in wind power leads to significant challenges as well as ample opportunity. Given the trend of building wind turbines of increasing sizes in offshore locations, future wind turbines tend to be higher-value structures with increasing maintenance difficulties; the development of effective strategies for wind farm operation and maintenance becomes more crucial than ever. This demand for a rapid expension in the wind energy sector motivates the SHM community to focus on monitoring strategies applicable to wind farms.

Once wind turbine farms are built and start operating, wind farm operators will begin to continuously control and monitor turbine performance; based on the measurements taken during operation, wind farm operators will then seek to reduce the cost of turbine operation and maintenance. Wind power cost reduction is generally achieved by minimising the average cost per unit of electricity generation for a turbine over its lifetime — the levelised cost of energy (LCOE) — which can be divided into

three objectives: (a) prolonging turbine lifetime, (b) reducing service cost , and (c) maximising power production [22]. The first objective coincides with the overall aim of SHM, and the five-step task of SHM is designed for this exact reason. The second objective can be achieved through a working SHM system, where cost is reduced by replacing preventive maintenance with predictive maintenance and by the early detection of failure. The third objective requires a control strategy that monitors the turbine performance in a wind farm from a population-based viewpoint. In short, a successful SHM system could play a very significant role in wind power operation and maintenance.

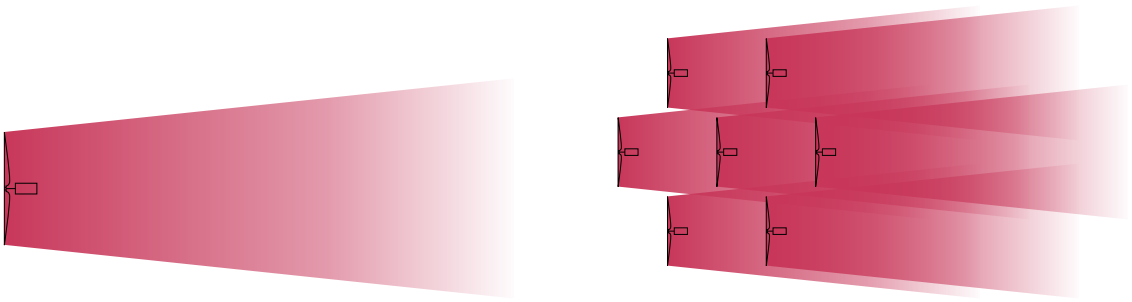
The task of wind farm monitoring is in essence that of PBSHM. A wind farm demonstrates a localised homogeneous population, where a population of nominally-identical structures are located in close vicinity to each other, allowing them to share the same environment. According to the wake effect, spatiotemporal correlations exist in the environmental and operational conditions across a wind farm, which in turn affect turbine behaviour. Such spatiotemporal variations may cause some turbines to be more prone to damage than others in the same farm, e.g. a turbine subject to more turbulence fluctuations may be more severely affected by fatigue loading. However, to the best of the author's knowledge, many existing methods for wind farm monitoring adopt a single-structure approach, which detects future anomalies based on the past behaviour of the same turbine, without considering the interactions between turbines; whereas the recent advances in PBSHM focus on normalising/homogenising the difference between structures in a population (e.g. the benign variations across a wind farm caused by different environmental and operational conditions), without modelling directly the spatiotemporal correlations across the farm. A spatiotemporal approach is therefore motivated for the problem of population-based wind farm monitoring, which will be the focus of this thesis.

1.3.1 Wake Effect across a Wind Farm

The wake effect within a wind farm, in short, describes the phenomena that involve reduced wind speed and more intensified turbulence after the wind passes through a turbine rotor. The physics behind such phenomena include, how the wind drives the motion in a turbine rotor and, subsequently, the rotation in gearbox and/or generator. This process is dependent on factors such as atmospheric properties (e.g. air density), wind flow characteristics (e.g. wind speed), turbine design specifications

(e.g. blade aerodynamics and drivetrain losses), and control settings (e.g. blade angle of attack). The effect of the wind on the turbine, in turn, results in disruption to the wind flow, forming a wake behind the rotor, which is composed of slower and more turbulent wind. The flow field immediately downstream of the turbine is also significantly influenced by the interaction between rotor and tower/nacelle, which is yet to be fully understood [23]. Further downstream, the wake expands, meanders (having a fluctuating wake trajectory at any time instance), and finally dissipates. Apart from the operating conditions of its parent turbine, the wake progression is also affected by ambient environment and the existence of other wakes [23, 24]. These wakes can severely affect the power output of downstream turbines, e.g. up to 80% of the power from the free stream wind may be lost in compactly spaced wind farms such as Lillgrund [25].

One of the simplest descriptions of a wind turbine wake is shown in Figure 1.1a for illustrative purposes, where the wake region behind a turbine rotor is assumed to be cone-shaped [26]; the superposition of multiple wakes is illustrated in Figure 1.1b. The level of complexity involved in the creation and progression of a single wake is already very difficult to model physically, not to mention the combined effect of multiple wakes in a wind farm. Nonetheless, the development of physics-based models contributes significantly to our physical understanding of the wake phenomena; for example, simulated turbine response obtained from a fluid-structure interaction method has demonstrated that the structural response of a turbine is primarily a result of rotor thrust instead of torque, which complicates the downstream wake vortices accordingly [27]. A more detailed review on physics-based wake models will be given in Chapter 2.



(a) Single turbine wake. (b) Wake superposition.

Figure 1.1: Cone-shaped turbine wakes (top view).

1.3.2 Confounding Influence Caused by Environmental and Operational Variations

In the context of SHM, data variations caused by the wake effect are parts of the *environmental and operational variations* (EOV), which are confounding factors that influence both the damage-sensitive features and structural response in a causal relationship [2]. The EOV found in the data collected from operating structures tends to be more complex than that found in laboratory data. EOV-induced changes in structural response can be comparable to, or sometimes more significant than, damage-induced changes [28]. It is almost impossible to find a damage-sensitive feature that is not influenced by EOV [29]; hence, handling EOV-induced changes in data becomes a necessary step to implement SHM in reality.

Much effort has gone into studying how EOV may affect the response of aerospace and civil structures. Some EOV directly affects how a structure vibrates; this is demonstrated by the flutter or buffeting in long-span bridges as a result of wind-induced vibrations [28, 30]. There are also factors that tend to alter the material properties and boundary conditions of a structure, which then lead to a different dynamic response; for example, [28] summarised that thermal-induced vibrations in bridges may stem mainly from changes in stiffness and boundary conditions (e.g. thermal contraction). In the case of wind turbines, different environmental and operational conditions may jointly affect the dynamic response of turbine components; an example discussed in [31] showed that the turbine modal frequencies might be negatively correlated with temperature (given specific operating ranges), and that this negative correlation could be moderated by the wind-induced aerodynamic damping.

The existing data-based methods for handling EOV in SHM data can be divided into two approaches. The first approach removes the confounding influence from the data using methods such as (PCA-based) minor principal component extraction or cointegration, so that the normalised data are only sensitive to structural damage (i.e. changes in the mass, stiffness or damping properties); the second approach models the EOV-induced changes in data, e.g. capturing the relationship between EOV and structural response, with an additional aim of understanding the effect of EOV [2, 30]. The work presented in this thesis focusses on the second approach — EOV modelling — to understand and, thus, account for the EOV across a wind farm.

1.3.3 Spatiotemporal Confounding Influence

When an individual structure is assessed, as in traditional SHM, the changes in EOV are often mainly thought of as temporal. For example, in the monitoring campaigns of bridges, the EOV can be considered as the daily or seasonal variations in temperature, wind loading, and traffic loading [30]. However, in PBSHM, the spatial dimension also needs to be considered; as well as the temporal trends, the structures in a population can also be subject to different EOV based on where they are located. The wake effect across a wind farm is an ideal example of such spatiotemporal EOV.

Similar to the single-structure case, there are two approaches to address the confounding influence caused by spatiotemporal EOV in PBSHM data. The first approach is to remove the EOV-induced changes by homogenising the data of various structures, which is in line with the recent advances in PBSHM — knowledge transfer. Such homogenisation of data tends to treat all benign variations alike, i.e. regardless of whether the difference between structures in a population is spatiotemporal or not. It can be achieved through methods based on domain adaptation [15, 18–20] or by interpreting the population data using ideas from modern theoretical physics [17].

This thesis aims to develop an alternative approach, which models the correlation between EOV and structural performance in a spatiotemporal setting. To the author's knowledge, the utilisation of spatiotemporal analysis has not been properly discussed in the field of SHM, primarily because most methods were developed to assess damage for individual structures. As PBSHM extends the subject of SHM from an individual structure to a population of structures, an unprecedented opportunity is provided to approach SHM from a spatiotemporal perspective.

Another difference between the two approaches is the scale of the relevant modelling methods. Most existing methods for PBSHM remove the confounding influence during the process of knowledge transfer, which is achieved at the structure scale, i.e. focussing on one target structure at a time instead of making inference collectively. In the context of a wind farm, the dynamics of each turbine is typically assumed to be the same process; thus, the data collected across the wind farm represent multiple realisations of the same process. Alternatively, the approach developed in this work concerns the modelling of the spatiotemporal confounding influence, which, within the scope of this thesis, requires a population of structures to be considered

collectively, i.e. at the population scale. In this case, the entire population is assumed a spatiotemporal process, and the population data represent a realisation of the process at various spatiotemporal points.

The two approaches are also motivated differently. The first approach is motivated to address the scarcity of damage-state data, such that the assessment of structure conditions can be achieved with limited availability of data. The second approach is motivated by the existence of spatiotemporal correlations across a population, which can be captured either in the EOV data of the population (if available) or in some kinds of physical laws (if well understood); hence, the motivation here is to make full use of the available knowledge (data or physics). Note that the use of both approaches assumes that the effect of (the same type of) damage is the same across a population of similar structures.

1.4 Spatiotemporal Correlations in SHM Data

The data collected for SHM purposes are ordered according to spatial and temporal indices. For example, in the case of a single structure, a bridge may have multiple accelerometers installed to take measurements at various locations [30]; in the case of a homogeneous population, considering a wind farm, for example, each structure is likely to be equipped with the same sensors, thus, the same set of measurements are typically recorded at each structure location. As a result, spatiotemporal correlations can be found in a significant portion of the available SHM data sets.

The general goal of data analysis is to extract information (through statistical summaries) and knowledge (via statistical inferences), which ultimately facilitate decision/policy making. So far, the majority of studies in SHM (including PBSHM) are concerned with developing automated approaches to extract knowledge and information from monitoring data to accomplish the five-step task of SHM, with automated decision/policy making being next in line. If spatiotemporal data are available during this process, then opportunities arise to develop prediction methods that incorporate the physical understanding of the phenomena of interest. Space and time are fundamental in our descriptions of the world and keep every observation and measurement recorded in an orderly manner; therefore, they are the fundamental factors that contribute to a wide range of phenomena. Understanding the spatiotemporal correlations in data may allow one to better grasp the dynamic evolution of

the phenomena and, thus, to better infer causal relationships [32, 33].

Simpson's paradox implies that causal models might not be generalisable across different spatial or temporal scales [32]. Hence, the awareness of spatiotemporal correlations in data also allows researchers to anticipate the potential difference at various spatiotemporal scales and interactions between scales [32, 33]. To this end, a range of modelling methods are developed to address the different spatiotemporal scales; the background literature on spatiotemporal models across disciplines will be briefly reviewed in Chapter 3.

1.5 Scope of the Thesis

This thesis aims to investigate a spatiotemporal approach to the PBSHM task of wind farm monitoring, with a focus on developing data-based modelling methods.

In the context of SHM, methods are developed with a primary focus on identifying damage, where the full potential of spatiotemporal correlations in data might not be unleashed; this lack of attention on spatiotemporality is more discernible in PBSHM. The inter-structure spatiotemporal correlations are evident across a wind farm, where the spatiotemporally changing environment induced by wakes affacts the dynamic response of turbines and poses a challenge to wind farm monitoring. A spatiotemporal approach to wind farm monitoring is therefore motivated. With the aim of developing diagnostic tools for a population-based assessment of wind farm performance, this work first establishes an anomaly detection framework in Chapter 4 that provides a spatial map of the turbine performance across a wind farm, by accounting for the spatiotemporal correlations between the environment and turbine performance. At the first attempt, an anomaly detector based on Gaussian process regression is chosen (Chapter 4). The latter half of the thesis (Chapter 5-7) is dedicated to developing a new modelling approach that is tailored to capturing the spatiotemporal variations across a wind farm; this new approach — a Gaussian process spatial autoregressive model — is inspired by the physics of the wake effect. It shows that physics knowledge can provide insight into how data-based models are developed, resulting in physics-inspired models. This new model is then incorporated into the established anomaly detection framework for wind farm performance indication.

1.5.1 Thesis Layout

The chapters that constitute the thesis are briefly summarised as below.

- Chapter 2 reviews the existing methods in the field of SHM for the monitoring of wind turbines and wind farms, with an awareness of the spatial and temporal scales of these methods.
- Chapter 3 reviews the background literature on spatiotemporal models, with more detailed descriptions given to the data-based models utilised later in the thesis.
- Chapter 4 introduces a mapping method for wind farm anomaly detection that accounts for the spatiotemporal EOV across a wind farm. A case study demonstrates the capability of the mapping method in identifying performance anomalies in an operating wind farm.
- Chapter 5 introduces the GP-SPARX model as a physics-inspired method that captures the spatiotemporal variations caused by the wake effect. Connections are drawn between the model structure and the physical understanding of the problem. The switching GP-SPARX method is also introduced to handle the changing wind directions that give rise to multiple wake patterns.
- Chapter 6 demonstrates the application of GP-SPARX to simulation data. Results show that a GP-SPARX model can capture the nonlinear correlations between turbines, and that the switching GP-SPARX method can capture time-varying wake patterns.
- Chapter 7 applies GP-SPARX to an operating wind farm for wind and power modelling.
- Chapter 8 provides the concluding remarks and points on future work.

Chapter 2

SHM for Wind Turbines

This chapter reviews the existing methods for the monitoring of wind turbines and farms. In the first half of the chapter, the SHM systems for wind turbine fault/damage detection are discussed in terms of the spatial and temporal scales. The second half of the chapter summarises the characteristics of offshore wind farms, followed by the reviews of existing physics-based and data-based models that (may) contribute to wind farm monitoring on a population level.

2.1 SHM for Wind Turbines in Various Spatial and Temporal Scales

In this section, a brief overview of the existing attempts at SHM for wind turbines is given in light of the temporal and spatial scales of the commonly-addressed monitoring problems and techniques. An SHM system usually consists of a sensing technology for the collection of damage-sensitive data and a data analysis method that infers the condition of the structure. The spatial and temporal scales of an SHM system are dependent on the type of monitoring task of interest and the type of sensing techniques available. It is to be clarified that, in the following text, the term temporal scale is referred to the smallest increment in measured data instead of the duration of time in which measurements are taken; the reason is that most sensors are designed to be permanently mounted to turbines and continuously record measurements from the first day of operation, in which case the time durations for all

sensor recordings are expected to be the same. Figure 2.1 visualises a range of sensing techniques applicable to wind turbine monitoring and how the data collected are used to infer the conditions at various spatial and temporal scales. Note that Figure 2.1 does not provide an exhaustive list of the SHM systems that have been applied to wind turbines but only includes those that provide data potentially suitable for spatiotemporal analysis.

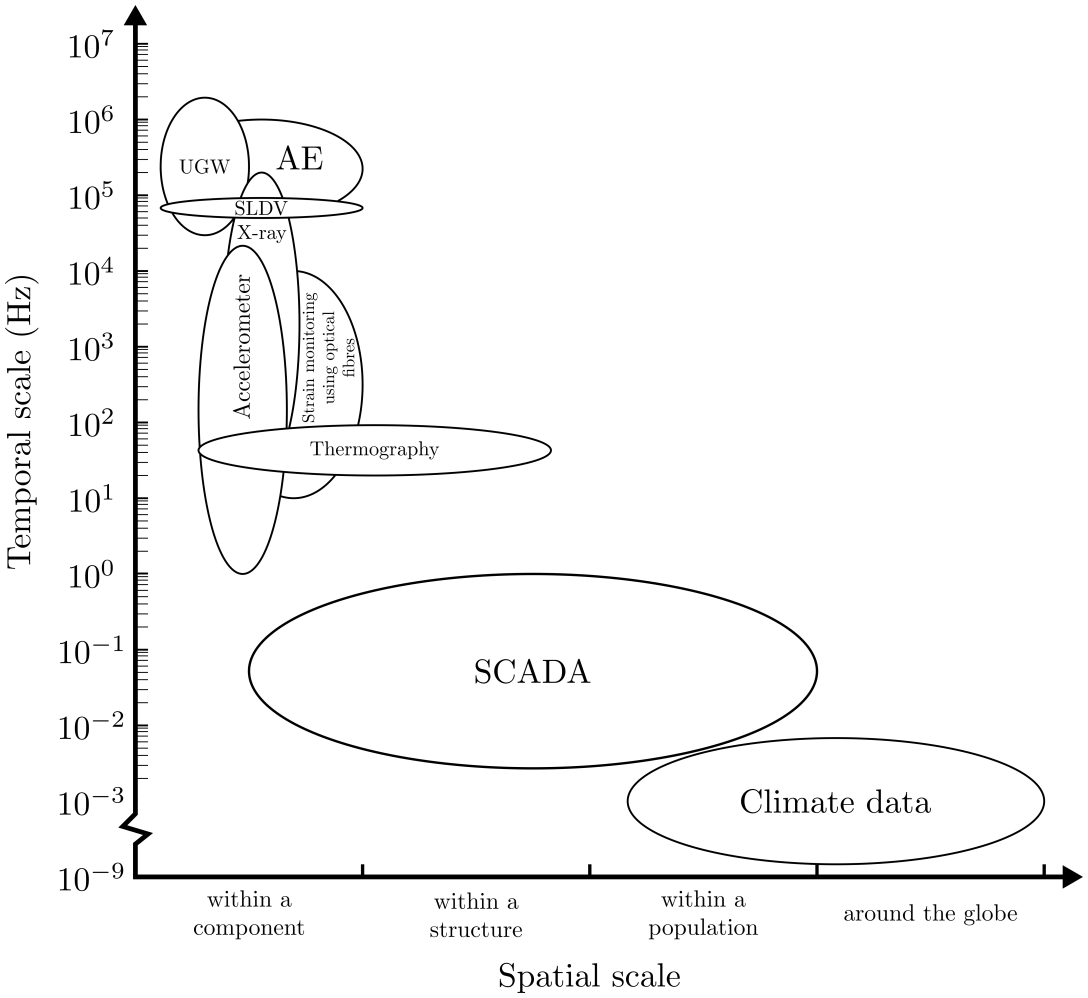


Figure 2.1: Existing methods for wind turbine SHM in terms of temporal and spatial scales.

Since damage is a local phenomenon, many SHM (and nondestructive evaluation) methods are initially developed to investigate the damage-prone component of a structure [2]; such methods are found towards the left side of Figure 2.1. The data collected by sensors — such as accelerometer and fibre optic strain sensor — provide point measurements of the structural dynamic response [34–36]; whereas methods based on acoustic emission (AE), ultrasonic guided waves (UGW), scanning laser

Doppler vibrometer (SLDV), X-radiography, and thermography may be able to "scan" across a component or multiple components [34, 37–41]. The latter type of methods is considered to be capable of both the localisation and detection of damage, e.g. locating a crack in a turbine blade; while the former is more commonly used for detection, i.e. detecting whether the mass, stiffness or damping properties have changed as a result of damage. Nonetheless, a network of point sensors may also have the capability of damage localisation, e.g. a distribution of fibre Bragg gratings can locate and track crack evolution [34, 36]. Note that the data obtained through accelerometers, optical fibre sensors, and thermography are sometimes referred to as condition monitoring system (CMS) data [35]. Absent from Figure 2.1 are methods that explore the electrical, electromagnetic, chemical, and physical properties of a component, such as methods based on electrical resistance, eddy current, corrosion sensor, crack detection, and oil analysis, which can detect damage in certain types of turbine components [2, 34, 35].

The majority of operating wind turbines are equipped with the supervisory control and data acquisition (SCADA) system, which records the statistical summary, including mean, maximum, minimum, and standard deviation, of the raw data stream (usually with a measuring frequency of 1 Hz) every ten minutes. Although a standard measurement nomenclature does not exist, most SCADA systems record the basic environmental and operational quantities, with a detailed summary provided in [42]. Previous studies have used the SCADA data to detect various types of damage and performance anomalies across a wide range of spatial scales. For example, within the scope of the drivetrain system, the continuous monitoring of gearbox oil temperature, power output, and rotational speed in the SCADA data provided early signs of planetary gear failure [43]. SCADA measurements taken at different parts of the same turbine can also be used together for fault diagnosis; for example, the relationship between 22 SCADA variables — describing the ambient environment, operational settings and performance of a turbine — and five different types of faults could be captured by a Gaussian process classifier, which was used to detect and classify turbine faults [44]. Another example in [45] demonstrated the detection and forecast of two types of failure occurred at two different turbines in the same wind farm by means of comparing the processed SCADA information across the farm; this shows an early attempt at PBSHM, but the confounding influence caused by the spatiotemporally varying environmental conditions was not properly accounted for.

Beyond the population level, data are available that describe the atmosphere, land

surface and ocean waves around the globe, which is summarised as the climate data in Figure 2.1. In comparison with SCADA data, climate data cover a larger geographical region (i.e. the globe) with coarser spatial and temporal resolutions; examples of climate data include but not limited to [46–49]. To the best of the author's knowledge, climate data are mainly used for wind farm site planning [50] instead of fault detection. However, the author believes that climate data are going to be beneficial for wind farm PBSHM in the future and, hence, decides to include the term in Figure 2.1.

The spatial and temporal scales of the sensing techniques shown in Figure 2.1 indicate how the collected data may be utilised to detect, locate or assess damage in turbines. The range of spatial scale serves as a rough indicator of the size of detectable defects (on the lower end) and the size of the region covered by the sensing technique (on the upper end). For example, SLDV may arguably provide images of the highest spatial resolution out of all SHM methods mentioned, whereas thermography may scan across multiple components simultaneously with a lower spatial resolution [34]. The range of temporal scale is determined by the measuring/sampling scale of the technique. A small temporal scale is required to capture the dynamics of structures/machines in motion; for example, methods based on UGW may be used to track fatigue crack development under vibration condition [51]. Methods of larger temporal scales focus on capturing the long-term changes; for example, SCADA data can capture the change in structural dynamics or operational conditions caused by damage in the structure.

It should be clarified that anomaly detection using SCADA data has been investigated over the last decade for the purposes of both wind integration studies and predictive maintenance. The former studies the effect of integrating wind power into existing electricity systems, where anomaly detection techniques are used to remove abnormal wind power data in order to obtain an accurate overview of the statistical properties of real turbine production [52]. The latter is a proactive, cost-effective maintenance strategy that predicts when maintenance should be performed, based on data-driven assessment of turbine conditions. In this context, anomaly detection techniques are used to detect turbine performance anomalies, especially as early signs of failure [42]. Anomaly detection in predictive maintenance is indeed another term for the first step of the SHM task (Section 1.1), where performance anomalies are the primary focus of detection. Nonetheless, the detection techniques are the same across various application domains [53].

2.2 SHM for Offshore Wind Farms

The requirements and considerations for offshore turbine operation and maintenance are different from those for the onshore counterpart because of the difference in characteristics between offshore and onshore environments. Firstly, the offshore environment is more suitable for siting large wind farms. There is generally a greater area available offshore, and it is easier to find a site offshore than on land that is far away from the dense residential areas while close to the electricity load centres [54]. Secondly, the offshore wind has higher speed, lower wind shear, and less turbulence than the wind on land. On the positive side, the wind with high speed and low turbulence can contribute to consistent, high power productivity, and low wind shear means low tower height requirement (for the turbine rotor to work in the less turbulent region of the atmosphere) [54]. However, these benefits come with costs. For example, the fact that higher wind speed is likely to be found further away from shore means that longer power transmission cables are required to harvest higher wind speed, which may result in more expensive cable installation, more transmission losses, and higher risk of cable failure. Likewise, the lower atmospheric turbulence offshore allows the formation of longer, more persistent turbine wakes, which place a bigger impact on the productivity of downstream turbines; increasing turbine spacing is a potential solution but may also lead to a higher power transmission cost. Therefore, any existing and future offshore wind farm is (to be) designed with a balance between productivity and cost. As a result, an offshore wind farm is likely to consist of a large number of turbines located moderately close to each other, sharing the same environment; the interaction between turbines thus plays a bigger part in the operation and maintenance of offshore wind farms.

2.2.1 Physics-Based Wake Models

The interaction between turbines across a farm is investigated in the study of the wake effect, which is briefly introduced in Section 1.3.1. In this section, a review is given on the existing physics-based wake models and their main areas of application.

The study of physics-based models that represent the governing laws of wake behaviours, i.e. wake models, can be dated back to almost four decades ago. Depending on the ways solutions are obtained, the existing wake models can be subdivided into

two types — analytical and numerical models. Analytical wake models, also known as kinematic wake models, provide an analytical or empirical solution to the relative drop in wind speed or power of a downstream turbine with regard to its upstream neighbour; such models are usually computationally inexpensive and associated with many simplifying assumptions [24]. Numerical wake models refer to methods based on computational fluid dynamics (CFD) that provide an estimate of the entire vector field of flow velocity across a wind farm, together with estimates of other quantities, such as pressure or temperature, by numerically solving the Navier-Stokes equations [23]; the computational cost of numerical wake models may vary depending on the CFD code of choice but is generally much higher than that of the analytical models. Owing to the complexity of the wake effect, existing physics-based models are often designed by confining the scope of the problem to fit specific applications.

One of the most important applications of wake models is in the optimisation of wind farm layout. The task is to find a farm layout that yields the best overall power production out of all potential candidates, while considering technological, economic, environmental, ecological, aesthetic, and social constraints. The standard approach to this task includes the assessment of all candidate layouts, which would only be practical using computationally efficient analytical models. However, model accuracy tends to deteriorate once turbine layouts and/or environmental conditions differ from the benchmark scenarios. The major types of analytical wake models were evaluated for operational wind farms of a range of sizes, terrains and layouts in a review given by Archer et al. [24].

Given the trend towards increased turbine sizes to meet the higher demand for wind energy, a second important application of wake models is the improvement of blade design. A larger turbine requires blades that are longer and yet more lightweight. Such new blade designs, however, can be prone to aeroelastic deformation, which inevitably complicates the structure of a turbine wake, especially under the combined effect of individual blade control and rotor-tower interaction [23]. Numerical models, e.g. the combination of CFD-based wake models and finite element (FE) turbine models, are the only physics-based options with the potential to provide sufficient accuracy. Not surprisingly, the extremely-large computational resources required by CFD-based models make them infeasible for most industrial implementations, and even some research cases. Because of cost limitations, such numerical models are only available for the simplest environmental and operational conditions. According to a review by O'Brien et al. [23], the development of advanced 3D numerical

approaches to wake modelling is still "in its infancy", with enormous potential as well as formidable challenges ahead.

Since the physics-based wake models (analytical and numerical) are mainly created for wind farm planning and turbine design, these models need to be adapted or repurposed for application to SHM; that is, to create a model that can assess current structural conditions based on monitoring data. Assuming that it is possible to repurpose the models, difficulties may arise from high computational burden (if using CFD-based models) or from poor accuracy (if analytical models are chosen), especially when the real environmental and operational conditions become different from the ideal scenarios encapsulated in the simplifying assumptions. Hence, given the availability of data collected from operating turbines, data-based models are more likely to produce predictions which supply the accuracy required for monitoring purposes (e.g. short-term turbine performance prediction) at a relatively lower cost [55]. Nevertheless, physics-based models can provide tremendous insight into how data-based models are to be created.

2.2.2 Data-Based Population-Level Models

A major objective of SHM for wind farms is to develop farm-level (or population/fleet-level) damage identification methods/software based on the data collected from the existing sensing technologies; this means extending the areas of the existing methods to cover the right side of Figure 2.1. To the best of the author's knowledge, data-based methods for population-level monitoring are currently under rapid development, especially in the field of PBSHM.

Power-curve monitoring is a method of wind farm performance indication that models the correlation between the environment (wind speed) and turbine performance (power). Previous studies have developed the power-curve method for farm-level monitoring (as the first applications of PBSHM) [12–14] and turbine-level [56] monitoring; the focus here is to review the power-curve method as a farm-level method. A power curve describes the dependency of power generated by a wind turbine upon ambient wind speed. Theoretically, turbines with the same design and configuration should show a similar shape of power curve. It is assumed that a power curve significantly distorted from the normal shape indicates anomalous turbine performance. The power-curve methods model the temporal variations of power

in relation to wind speed, while accounting for the spatiotemporal environment, manufacturing tolerances, and varying operational settings as the population variance; any variation outside of the population variance is deemed anomalous [12–14].

As mentioned previously in Section 1.2, the current trend of PBSHM is to focus on knowledge transfer between structures via abstract spaces, with population-level monitoring for wind farms being one of the most significant areas of application [15, 17].

2.3 Conclusions

This chapter provides, firstly, an overview of the existing SHM systems for wind turbines and, secondly, summaries of the physics-based and data-based models for wind farm monitoring. By reviewing the SHM methods for wind turbines in terms of spatial and temporal scales (Figure 2.1), the need for more research is highlighted on methods that allow data from individual turbines to be used collectively for the assessment of wind farm conditions. The unique characteristics of offshore wind farms bring to attention the significance of accounting for the spatiotemporal interaction between turbines when monitoring the farms. The inter-turbine interaction is commonly described using physics-based wake models, which are not the best options for SHM; whereas capturing spatiotemporal correlations is not the primary focus of the existing population-based methods for wind farm monitoring. With this thesis aiming to explore the usability of spatiotemporal models for wind farm monitoring, the next chapter is dedicated to reviewing the existing spatiotemporal methods in other research fields.

Chapter 3

Methods for Handling Spatiotemporal Data

This chapter provides a review of the existing approaches to spatiotemporal analysis across disciplines. Since space and time are the fundamental factors of most physical phenomena, data-based models that account for spatiotemporality are physics-inspired to various extents; this demonstrates one way to utilise physics/mathematical models in data-based algorithms, which is a current research trend in the communities of SHM and machine learning. More detailed descriptions of Gaussian process, spatial autoregressive models, and Gaussian process nonlinear autoregressive model with exogenous inputs are given in Section 3.2 to Section 3.4; these models will be adapted to account for the spatiotemporal wake effect for the purpose of wind farm monitoring in the following chapters of this thesis.

3.1 Existing Methods for Spatiotemporal Modelling

Spatiotemporal data analysis is relevant to various problems in the fields of environment, climate, ecology, health, energy, food, finance, etc. The dynamic evolution of many phenomena can only be captured with the spatiotemporal correlations properly accounted for, e.g. the formation of natural hazards. A representation available only in the spatial or temporal domain is insufficient for such phenomena

[32]. The arbitrary combination of spatial and temporal models is unlikely to result in a valid spatiotemporal model [33]. Hence, an appropriate spatiotemporal approach is necessary when handling data that describe variations across space and time.

The definitions of space and time are to be clarified before reviewing the existing spatiotemporal methods. In Newtonian physics, space and time are distinct from one another in that time moves forward linearly and continuously in an irreversible direction regardless of anything that happens in the world. Although, in Einsteinian physics, space and time are demonstrated to be interdependent and inseparable, this relativistic view barely affects phenomena at the scale of the solar system [33]. Currently, in SHM, the phenomena associated with the structures of interest stay within the scales of space and time where Newtonian physics is assumed.

Most existing models that describe spatiotemporal dependence assume persistency, i.e. nearby values (in space and time) are more alike [32]. This spatiotemporal persistency is most-commonly described using spatiotemporal covariance functions; they characterise the statistical dependencies in space and time for a stochastic spatiotemporal process. Heine demonstrated that spatiotemporal correlation functions can be derived from stochastic partial differencial equations (PDEs) [57]. Christakos and Hristopulos developed separable spatiotemporal covariance functions for applications in environmental health monitoring, i.e. to monitor how the natural environment affects the state of human health [33]. Cressie and Wikle categorised the space-time covariance functions derived from moments of probability distributions as part of a descriptive approach to representing spatiotemporal dependence, whereas they preferred a dynamical approach where the spatiotemporal dependence is motivated by physical understandings [32]. Stein summarised three theoretial aspects of spatiotemporal interactions implied by existing covariance functions and developed covariance functions to address each aspect [58]. These spatiotemporal covariance functions were later adapted and applied to crime event modelling [59], wind field forecasting [60], and ozone modelling [61]. Note that a large portion of the existing covariance functions are developed assuming spatiotemporal stationarity.

The spatiotemporal correlations depicted as covariance functions provide information about the probability density function of the underlying spatiotemporal stochastic process and can be used to infer the behaviour of the process [32, 33]. Kriging, which is equivalent to Gaussian process regressio, is one of the commonly used methods to make spatiotemporal predictions from specified covariance models [32]. Alternative methods such as Bayesian maximum entropy can also combine prior knowledge

(contained in covariance functions) with data for mapping spatiotemporal random fields [33].

Several modelling approaches have been developed to address the cases when the assumption of spatiotemporal stationarity no longer holds. One of the simplest approaches assumes that the nonstationarity lies in the mean of the process. The nonstationary mean (in space and/or time) can be captured through a separate regression model; the results obtained from the covariance-based methods can then be combined with the nonstationary mean to represent the spatiotemporal process [33]. Note that the nonstationary mean may be described by a physics-based model, which, when combined with a data-based covariance model, leads to a physical-statistical approach — also known as grey-box modelling or physics-informed machine learning in various fields [75, 85]. In particular, residual modelling refers to the use of a data-based model to capture the residual error between data and a physics-based model; this approach is demonstrated by combining Gaussian process-based models with Morison's equation to model the time series of wave loading [62].

When the nonstationary behaviour (or dynamic evolution) of a spatiotemporal process is relatively well understood, an alternative approach can rely primarily on this physical understanding. The results are physics-based models, which are likely based on PDEs, such as stochastic PDEs. The predicted behaviour of the process is then given as the solutions to the stochastic PDEs, and the spatiotemporal correlations may be found in the covariance derived from the stochastic PDEs [32, 57].

Given the availability of physics-based models for a spatiotemporal process, the observed data are unlikely to be perfectly in line with the theory because of (a) the uncertainties associated with data collection, storage and processing, (b) the difficulty to measure directly the underlying process in certain situations, or (c) the limited applicability of the physics-based models due to simplifying assumptions. In this context, a hierarchical state-space framework can utilise the observations (potentially from multiple sources) together with known physics (possibly in the forms of PDEs or integro-difference equations) to make inferences that better represent the process of interest [32]; this is also an example of the physics-informed data-based model.

In cases when there is limited knowledge about the dynamic evolution of a process or when the PDEs describing a process give intractable solutions, the spatiotemporal correlations are frequently represented using models of the autoregressive (AR) forms. The use of AR models is motivated by that the discrete approximation of

a continuous process is often described in terms of difference equations, which, for example, satisfy the Markov property. Discretisation also allows a spatiotemporal process to be approximated as a time series of spatial processes or a spatial process of time series [32]. In the former case, a vector autoregressive model (VAR) can be used by assuming autoregression in time and specifying the spatial relationships in a weight matrix, which is considered as a spacial case of the space-time autoregressive moving-average (STARMA) model; the STARMA model can also be equivalent to the simultaneous autoregressive model and/or the conditional autoregressive model under certain conditions [32]. It is worth mentioning that many studies on spatiotemporal environmental processes assume autoregression in time domain only, with the spatial relationships modelled in a case-specific manner. Hence, the spatially evolving time series is less likely to be modelled explicitly as an AR process; instead, it can be modelled as a Gaussian process with the covariance being the convolution of an isotropic spatial element and an autoregressive temporal element [63], or it can be modelled as a Gaussian Markov random field when Markov assumption is appropriate [64].

The spatiotemporal correlations found in data are dependent on the underlying process; thus, the choice of modelling method is determined by the type of spatiotemporal process under consideration and is largely case-specific.

The rest of this chapter will describe three data-based models in more detail, as they will be adapted to capture the spatiotemporal relationships for the wind farm monitoring problem in later chapters.

3.2 Gaussian Process

As an extension to multivariate Gaussian distributions that describe random variables of finite dimensions, a Gaussian process (GP) is held over functions, with the function values $f(X)$ specified at a potentially infinite number of inputs X. The fact that a GP can provide consistent inferences over any finite number of points makes it a powerful tool to learn functional relationships [65]. A GP is specified by a mean and a covariance function,

$$f(X) \sim \mathcal{GP}\left(m(X), k(X, X')\right) \tag{3.1}$$

Conventionally, the mean function is assumed to be zero to simplify the formulation, and the covariance function takes the form of *squared exponential* (SE), which is the most commonly-used kernel in describing smooth data variations,

$$k_{SE}(X, X') = \sigma_f^2 \ \exp\left(-\frac{1}{2l^2}|X - X'|^2\right) \tag{3.2}$$

The process of model training involves the optimisation of hyperparameters, which, according to Eq. 3.2, include process variance σ_f^2 and input lengthscale l. Note that characteristic lengthscales are used in the current model, meaning that each input dimension is assigned a different lengthscale; for input data with a dimension D, there is a total of D lengthscales $l_1, \ldots, l_D$. These hyperparameters are optimised here by maximising the marginal likelihood via gradient-based optimisation methods [65].

With the optimised hyperparameters, a predictive distribution corresponding to the unseen inputs, X^*, can be given by specifying the predictive mean vector, $\bar{\mathbf{f}}^*$, and covariance matrix, $\text{cov}(\mathbf{f}^*)$:

$$\bar{\mathbf{f}}^* = K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1}\mathbf{y} \tag{3.3}$$

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

where $K(\cdot, \cdot)$ denotes a covariance matrix obtained from Eq. 3.2, and σ_n^2 is the noise variance associated with the noisy observations $\mathbf{y}$. To account for the noise in the predictive distribution, the covariance for the noisy target $\mathbf{y}^*$ is $\text{cov}(\hat{\mathbf{y}}^*) = \text{cov}(\mathbf{f}^*) + \sigma_n^2 I$.

For more details on GP implementation, readers are referred to [65].

3.3 The Spatial Autoregressive Model

Methods based on the spatial autoregressive (SAR) model have been applied to many problems across disciplines such as spatial econometrics and geostatistics over the past few decades. Most variations of the SAR model are developed based on the

linear spatial autoregressive process by Cliff and Ord [66],

$$y_i = \alpha \sum_{j=1}^{n} w_{ij} y_j + \epsilon_i \tag{3.5}$$
$$\epsilon_i \sim N\left(0, \sigma^2\right) \quad i = 1, \ldots, n$$

where, for a total of n locations, the observation y at each location i is modelled as a linear combination of the observations at the neighbouring locations. The information about spatial contiguity is contained in a $n \times n$ weight matrix W, where the element $w_{ij} \neq 0$ when the locations i and j are contiguous. In Equation 3.5, each observation is modelled as the sum of an underlying mean (obtained from the spatial lag term) and a random component ϵ_i. In this case, ϵ_i is assumed to be a Gaussian white noise with a variance σ^2. Equation 3.5 can also be presented in a matrix form,

$$\mathbf{y} = \alpha W \mathbf{y} + \boldsymbol{\epsilon} \tag{3.6}$$
$$\boldsymbol{\epsilon} \sim N\left(0, \sigma^2 I_n\right)$$

Variations of the SAR model are created by including exogenous input variables, with the general form given by [67],

$$\mathbf{y} = \alpha W \mathbf{y} + Z\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3.7}$$

where Z may include the matrix of exogenous inputs X, the weighted sum of the spatially lagged exogenous inputs WX, and/ or a vector of ones $\mathbf{1}$, and $\boldsymbol{\beta}$ is a vector of parameters; for example, if $Z = [\mathbf{1}\ X\ WX]$, $\boldsymbol{\beta} = [\beta_0\ \beta_1\ \beta_2]'$.

Cliff and Ord highlighted two main types of analysis associated with the SAR model: (a) estimating model parameters and (b) detecting spatial autocorrelation within data [66]; following them, Griffith developed methods of scientific visualisation to explore the spatial autocorrelation in georeferenced data [68], whereas Lee [69] and LeSage [67] investigated the methods to estimate and interpret model parameters, respectively. The general form of a standard SAR model remains the same throughout these studies (as indicated in Equation 3.7), which is characterised by predefined functional forms and weight matrices. Similar to the case of all other deterministic models, substantial work is required to decide which functional form of the SAR model is best for a specific problem. The predefined weight matrices in the SAR model specify prior knowledge in spatial correlations [67], which provides a parsimonious

solution but might introduce subjective bias. The standard SAR model do not account for temporal correlations without a spatiotemporal extension, however, the SAR-based spatiotemporal models inherit the same traits [67] and, thus, constraints.

3.4 Gaussian Process Nonlinear Autoregressive Model with Exogenous Inputs

An autoregressive model based on GP is one of the methods to capture nonlinear relationships while circumventing the limitations caused by predefined functional forms. Such a model also overcomes a major drawback of GP — that it can only interpolate between data points — by adopting an autoregressive model structure; the resulting model is able to extrapolate by learning a static map between values of the target variable itself and thus capture the behaviour of dynamical systems [70–72].

The Gaussian process nonlinear autoregressive model with exogenous inputs (GP-NARX) is given as,

$$y_i = f\left(\mathbf{x}_{i-l_\mathrm{x}}, \ldots, \mathbf{x}_i, y_{i-l_\mathrm{y}}, \ldots, y_{i-1}\right) + \epsilon \tag{3.8}$$

where $\mathbf{x}$ is the exogenous inputs. The numbers of lags in the exogenous term and in the autoregressive term are denoted by l_x and l_y, respectively. A Gaussian process prior is assumed over functions of $f(\cdot)$, as indicated in Equation 3.1. The noise ϵ is assumed Gaussian, $\epsilon \sim N(0, \sigma_n^2)$, with the noise variance σ_n^2.

The GP-NARX model is able to make two types of predictions. The one step ahead (OSA) prediction is made when true measurements are used as model input, and the model is only predicting the next (temporal or spatial) step ahead; the model predicted output (MPO) prediction is made when the model predictions from the previous step(s) are used as input to compute the next step [72]. Using the superscript $*$ to indicate model predicted values, the OSA prediction may be expressed as,

$$y_i^* = f\left(\mathbf{x}_{i-l_\mathrm{x}}, \ldots, \mathbf{x}_i, y_{i-l_\mathrm{y}}, \ldots, y_{i-1}\right) + \epsilon \tag{3.9}$$

and the MPO prediction as,

$$y_i^* = f\left(\mathbf{x}_{i-l_x}, \ldots, \mathbf{x}_i, y_{i-l_y}^*, \ldots, y_{i-1}^*\right) + \epsilon \tag{3.10}$$

More details on the implementation and application of GP-NARX can be found in [70–72].

3.5 Conclusions

This chapter reviews the existing methods of spatiotemporal modelling, which are physics-informed to various extents. If the phenomenon is well understood, a physics-based model, such as a stochastic PDE, can be solved to predict the spatiotemporal process. When partial physics knowledge is available, e.g. a physics-based model only describes the ideal situations, the physics in the form of PDEs may be incorporated in a hierarchical state-space framework; alternatively, the physics knowledge may be converted into spatiotemporal covariance and/or mean functions and applied to data-based frameworks such as GP (kriging), Gaussian random fields, and AR models. In cases of limited physics understanding, the data-based models can be used with generic priors to capture spatiotemporal variations.

This thesis mainly concerns the last scenario — (physics-inspired) data-based methods. In Chapter 4, a purely data-based GP (with generic priors on mean and covariance) will be developed into a wind farm performance indicator that accounts for the spatiotemporality due to the wake effect. Chapter 5 will provide the details on how the physics of the wake effect motivates the development of a GP-based spatial autoregressive model, with its application to simulated and real wind farm data given in Chapter 6 and Chapter 7, respectively.

Chapter 4

The Mapping Method for Wind Farm Anomaly Detection

This chapter introduces a mapping method for anomaly detection across a wind farm. As the first attempt in this thesis to develop a spatiotemporal approach to wind farm monitoring, the mapping method is developed based on a purely data-based model, GP regression. Since data play an important role in data-based modelling, this chapter begins with a brief overview of an operating wind farm (Section 4.1), from which data are collected and used for the case studies presented later on in this thesis. Section 4.2 is a description of the mapping method, including details on the algorithm of choice, the (spatiotemporal) functional relationships captured, and the metrics for anomaly detection. Section 4.3 provides case studies that demonstrate the capability of the mapping method as a performance indicator for operating wind farms.

4.1 SCADA Data from the Lillgrund Wind Farm

This section provides a brief summary of an operating offshore wind farm, Lillgrund, and the SCADA data collected from the wind farm.

Lillgrund is located in the middle of the Öresund region between Denmark and Sweden; it consists of 48 turbines rated at 2.3MW, with the farm arrangement as seen in Figure 4.1. The turbine spacing is $8.6R$, where R represents the turbine rotor

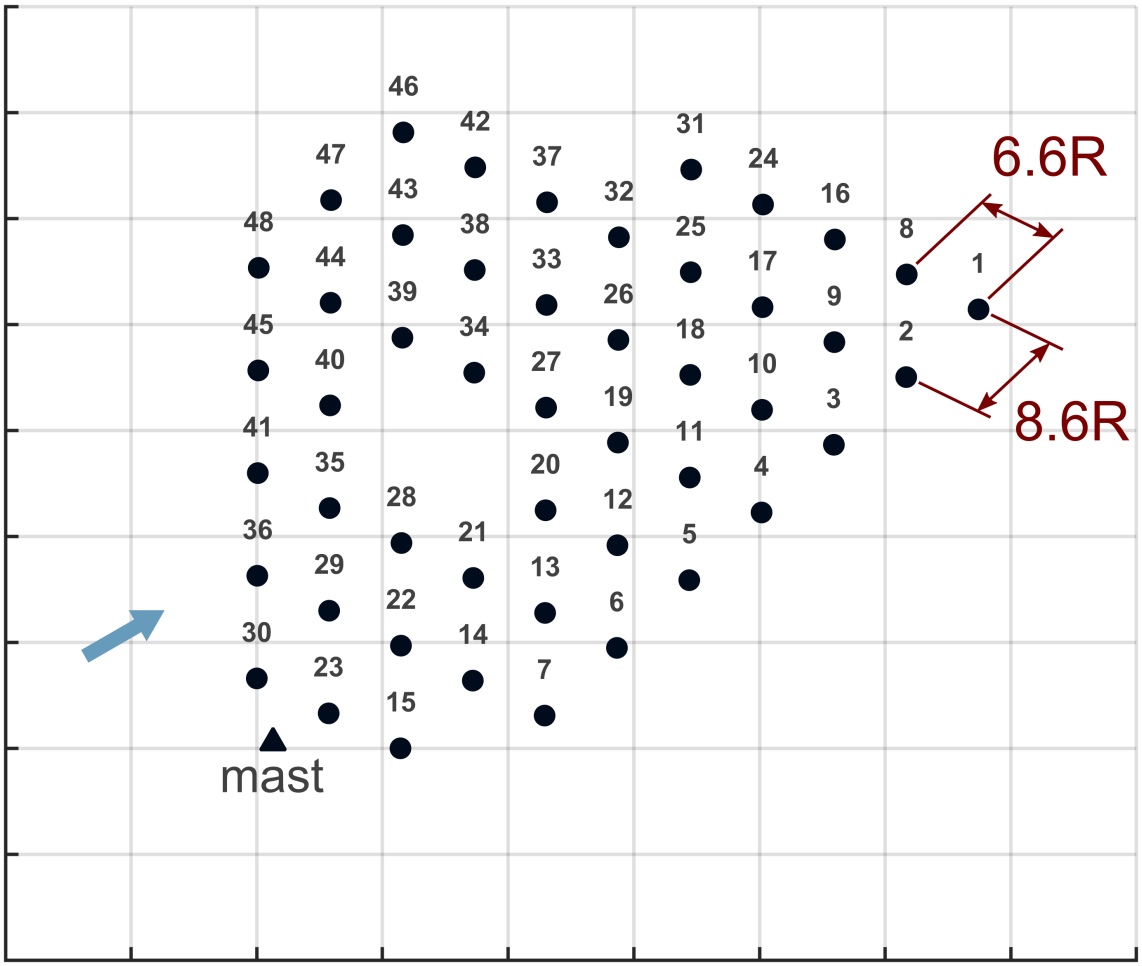


Figure 4.1: Layout of the Lillgrund wind farm. The dots represent the turbines, and the triangle represents the weather mast. Turbine spacings are shown in terms of rotor radius. The prevailing wind direction is indicated by the blue arrow. The spacing between ticks along both axes is 10R.

radius, in the prevailing wind direction as indicated by the blue arrow in Figure 4.1 (i.e. the column-wise direction of the wind farm); the spacing in the row-wise direction is slightly smaller ($6.6R$) [73]. Such spacings are considered small because the layout was originally designed for smaller turbines; this compact layout was preserved so as to facilitate the investigation into how turbine wakes can affect power production [25]. Detailed description and analysis of the Lillgrund wind farm can be found in [25, 73].

The data used in the following analysis were collected during a year of operation from the SCADA system in Lillgrund. As introduced in Section 2.1, the SCADA data are presented as a statistical summary, including the mean, maximum, minimum,

and standard deviation, of the actual recorded data every ten minutes. In this case study, *time series data* refer to a series of mean values summarising every ten-minute section of the raw data streams.

One may argue that the ten-minute interval is too coarse to represent the temporal variations, as the data fluctuations (e.g. due to turbulence) may be lost through the averaging process. However, in fluid dynamics, a fundamental assumption in the Reynolds-Averaged Navier-Stokes approach to modelling turbulent flow is that a statistically steady flow can be thought of as the sum of a time-averaged term and a fluctuation term (i.e. Reynolds decomposition) [74]. Hence, the ten-minute mean values provide an ample opportunity to explore the spatiotemporal variations in the mean wind, given that turbulence is notoriously difficult to model and usually requires a separate step/model to capture. This thesis focusses on the spatiotemporality expressed through the mean wind and is a prerequisite for developing a more complex model that captures the effect of the turbulent wind.

4.2 The Mapping Method as a Spatiotemporal Anomaly Detection Method

The mapping method aims to detect performance anomalies by mapping the spatiotemporal variations in EOV onto an EOV-affected feature that represents structural performance. It is in essence a novelty detection method with the assumption that the spatiotemporal pattern in the chosen feature will be maintained unless performance anomalies (including damage) occur. The mapping method consists of two steps: modelling and detection, with the details given as follows,

1. Baseline modelling refers to the training of a model that describes the normal variations of a spatiotemporal pattern, based on which the normal structural performance is predicted during a testing period.

2. Anomaly detection involves the quantification of the discrepancy between predictions and measurements. Anomalies are detected when the discrepancy value exceeds a preset threshold.

4.2.1 A Baseline Model Based on GP Regression

To model the spatiotemporal variations across a wind farm, a flexible regression algorithm is required. The model of choice in this chapter is GP regression, which has been used as a novelty detector in place of the control chart since [30]. The first advantage of GP as a probabilistic model is its ability to provide a predictive distribution, from which the mean values and confidence intervals can be computed [65]. In the context of PBSHM, the confidence intervals also account for the population variance, i.e. the difference between individual structures within a population that arises from benign sources such as turbulent boundary conditions and manufacturing tolerances, as a means of handling confounding influence [14]. Another advantage of GP lies in the fact that it is nonparametric and, thus, well suited for learning complex relationships, as opposed to a parametric model whose complexity may be restricted by a predefined functional form. Hence, GP regression is used here to construct an anomaly detector that accounts for the spatiotemporal EOV as part of the normal population behaviour.

4.2.2 Functional Relationships Captured by the Model

The mapping method aims to capture the spatiotemporal correlation between the EOV and turbine performance. Power is chosen as the target feature since it is a direct measurement of turbine performance. At the same time, power is a direct result of wind variations as well as operational control; thus, the power variation across a wind farm is a spatiotemporal process. Input features include the spatial coordinates at all turbine locations and the wind speed at a subset of locations. A resulting GP model is trained to predict the power time series at all turbine locations in a wind farm, from the corresponding spatial coordinates and wind speed at a subset of locations. In other words, the model intends to capture the wind-power relationship and how it is affected by the wake-induced correlation across space and time.

The wind-power correlation can be summarised by a power curve, with an example shown in Figure 4.2. Note that here the data are normalised so that the values of power and wind speed vary in-between -1 and 1. It is seen that power is strongly correlated to wind speed, forming an approximately sigmoidal (or cubic within a specific range) relationship. The shape of the curve is piecewise as a result of the turbine control

strategy; thus, the modelling of a full power curve is a complicated task in itself. The use of a simple GP model, without incorporating physical knowledge or partitioning the input space, might result in reduced accuracy in regions of control intervention (e.g. above rated power) [56]. Therefore, it is unlikely that a simple GP is able to capture the full wind-power correlation as well as the spatiotemporal correlation induced by wakes. As a result, the data used in the following analysis correspond to a restricted set of operating conditions, in which the wind speed variations are kept within the range in-between the cut-in and rated values (Figure 4.2). An extension to modelling the full wind-power correlation may be useful once the current model is proven valid.

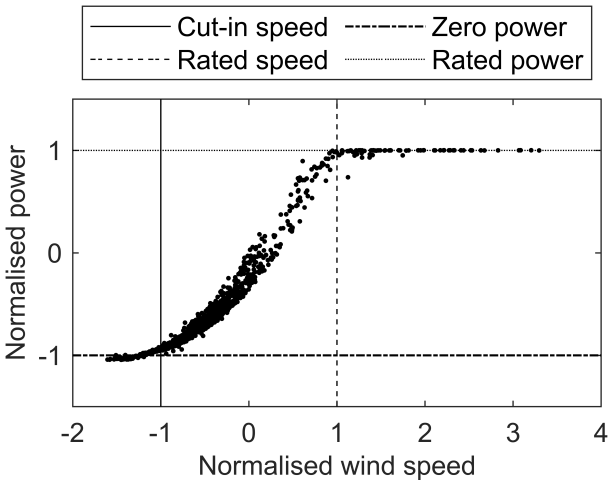
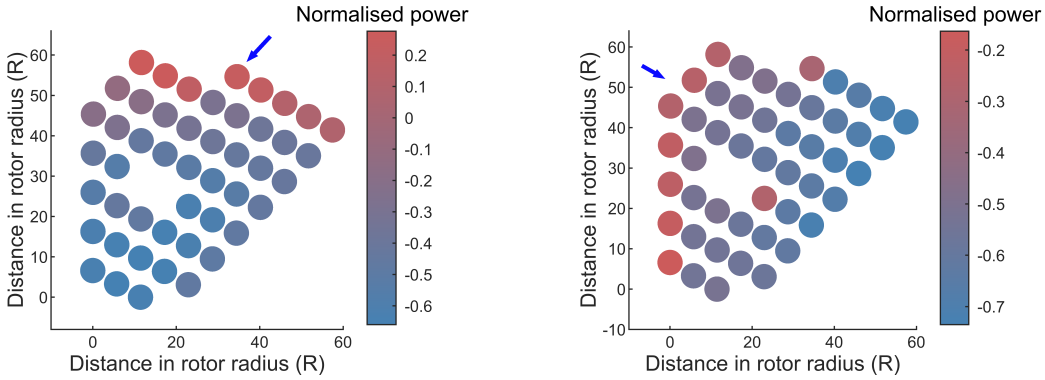


Figure 4.2: An example of a power curve

At the risk of repetition, it should be clarified that power-curve monitoring is not part of the aim of the mapping method and that the turbine-by-turbine wind-power correlation is not modelled directly. Instead, wind speed and power are the chosen features that represent EOV and turbine performance, respectively. The aim of the mapping method is to achieve EOV modelling in a spatiotemporal setting, and the aim of this chapter is to demonstrate the potential of this method as a performance indicator for a wind farm.

The GP model not only captures how power changes with wind speed temporally, but how this correlation varies spatially. The spatial aspect comes from the wake effect in the wind farm, which affects wind speed and power in a similar manner. Figure 4.3 shows the spatial distribution of turbine power output across the Lillgrund wind farm during two different time periods. Here the maximum power output in a farm is found at the front row of turbines, and the power levels decrease as

the wind progresses further downstream across the farm. This spatial correlation is modelled by using as input variables (a) the spatial coordinates of all turbine locations and (b) the wind speed time series at a fixed set of reference locations. The reference locations are selected to ensure that the model learns the spatial and temporal correlations through interpolation (i.e. to avoid extrapolation); the choosing of reference turbines will be discussed later in Section 4.3.1. Figure 4.3 also shows that the spatial pattern is sensitive to wind direction, which changes with time; this time-varying spatial pattern is also captured by the model.



(a) 42°. (b) 300°.

Figure 4.3: Example power variations across the Lillgrund wind farm corresponding to different wind directions, which are indicated by the blue arrows

To conclude, the GP model captures (a) the correlation between wind speed at reference locations and power output at each turbine, (b) how the wind-power correlation varies spatially, and (c) how the spatial pattern changes with wind direction (which is time-varying).

4.2.3 Discrepancy Measures and Benchmark Values

After establishing a model of normal condition, the expected normal turbine performance during a test period can be predicted, and the discrepancy between the prediction and measurement are computed. As a population-based method, the mapping method detects anomalies via comparisons across the structures in a population. Therefore, a single discrepancy value is assigned to each turbine, summarising the difference between the predicted and measured power time series, so as to facilitate the comparison between turbines. Two discrepancy measures are used for the purpose

of this study, namely *normalised mean-squared error* (NMSE) and *mean standardised log loss* (MSLL), which are error metrics commonly used for model quality evaluation in the literature of GP [65].

The NMSE is given by

$$\mathrm{NMSE}_s = \frac{100}{n_\mathrm{T}\sigma_*^2}\sum_{t=1}^{n_\mathrm{T}}\left(y_{t,s}^* - \bar{\mathrm{f}}_{t,s}^*\right)^2 \tag{4.1}$$

where y and $\bar{\mathrm{f}}$ denote the measured target variable and the GP predicted mean, respectively. The super- or subscript $*$ indicates the test data set, and the subscripts t and s denote time instance $t = 1 \ldots n_\mathrm{T}$ and turbine index $s = 1 \ldots n_\mathrm{S}$. The mean-squared error term is normalised by the term $100/\sigma_*^2$. σ_*^2 represents the variance of the measured target variable in the testing data across all turbines, in order to remove error sensitivity to the scale of testing data. The scaling factor 100 converts the error values into percentages. A score of 100% means the predictions are equivalent to the target mean of the test data. Hence, 100% is considered a benchmark value for NMSE; an NMSE value below 100% means that the model can provide better predictions than the mean of test data and is assumed to have captured some correlations between input and output. Note that the calculation of NMSE does not include information about GP predictive variance.

By contrasts, MSLL includes both the mean and variance of the GP predictive distribution in its formulation [65],

$$\mathrm{MSLL}_s = \frac{1}{2T}\sum_{t=1}^{n_\mathrm{T}}\left(\log(2\pi\hat{\sigma}_*^2) + \frac{\left(y_{t,s}^* - \bar{\mathrm{f}}_{t,s}^*\right)^2}{\hat{\sigma}_*^2}\right) - \frac{1}{2T}\sum_{t=1}^{n_\mathrm{T}}\left(\log(2\pi\sigma^2) + \frac{\left(y_{t,s}^* - \bar{y}\right)^2}{\sigma^2}\right) \tag{4.2}$$

The first term in Equation 4.2 is the negative log probability that indicates how likely the target measurements are to be predicted using the trained model. Here, the GP-predicted variance for a noisy target is $\hat{\sigma}_*^2 = \mathrm{diag}(\mathrm{cov}(\hat{\mathbf{y}}^*)) = \mathrm{diag}(\mathrm{cov}(\mathbf{f}^*) + \sigma_n^2 I)$. The second term in Equation 4.2 represents the (negative log) probability obtained by treating the training mean $\bar{y}$ and variance σ^2 as the model. It standardises the first term such that MSLL > 0 if the GP predictions are worse than the trivial model of training mean and variance. The benchmark value of MSLL is therefore 0, below which the model prediction is considered acceptable.

In the analysis that follows, the benchmark values mentioned above are used directly

as detection thresholds. Usually, in SHM, thresholds are determined by assuming a Gaussian generating distribution for the normal condition; for non-Gaussian features, thresholds can be obtained through methods such as extreme value statistics and receiver operating characteristics [2]. However, the Gaussian assumption is not appropriate for the wind speed or power distribution, and the available data do not contain the necessary information (i.e. there is no labelled data) for the advanced threshold generating methods. As a result, the benchmark values are used as thresholds in this thesis to demonstrate how the mapping method can be used to detect performance anomaly across a wind farm. Before implementing the mapping method in an actual wind farm, labelled data are needed to further validate the model and to generate realistic thresholds.

4.3 A Case Study on an Offshore Wind Farm

In this section a case study is presented to demonstrate the potential applicability of the mapping method to the problem of wind farm anomaly detection using SCADA data. Section 4.3.1 summarises the data used for model training and testing. It is followed by the testing results which demonstrate (a) the model's capability to predict power production with adequate accuracy (Section 4.3.2), and (b) the potential to use the model as a performance indicator for wind farms (Section 4.3.3).

4.3.1 Training and Testing Data

A GP-based model of normal condition is trained to predict power from spatial coordinates and wind speed, as mentioned in Section 4.2.2. One of the input features to the GP is the spatial locations, which are given as the Cartesian coordinates of each turbine position. Another input feature is the wind speed at a number of reference turbines. A set of ten reference turbines are randomly selected across the wind farm, with the locations indicated by the blue rectangles in Figure 4.5. The number of reference turbines and their locations are chosen to (visually) ensure that they are spread across the wind farm in order to avoid extrapolation in space, since a standard GP model (without incorporating physics) is unable to extrapolate [75]. The effect of reference turbines on model predictive accuracy will be discussed later in Section 4.3.4. More advanced methods for reference selection such as Latin

hypercube sampling [76] may be considered for future studies.

The training and testing data sets are selected based on the following criteria. First, the data on wind speed and power used in the case study are confined to the range between cut-in and rated values (Figure 4.2), for reasons mentioned in Section 4.2.2. Second, the training data cover a range of wind directions as the normal condition of the wind farm includes the variations across multiple wind directions. The available data are sectioned into many time windows according to the wind direction. The training data consist of time windows that correspond to four different wind direction ranges — according to the layout of the wind farm; each testing data set corresponds to a time window from one of the four wind direction ranges. Note that since the weather mast data are unavailable to the author, the turbine nacelle angle is used to estimate the local wind direction around each turbine, which has been proven acceptable in [25]. The wind (nacelle) direction ranges covered by training and testing data are illustrated in Figure 4.4.

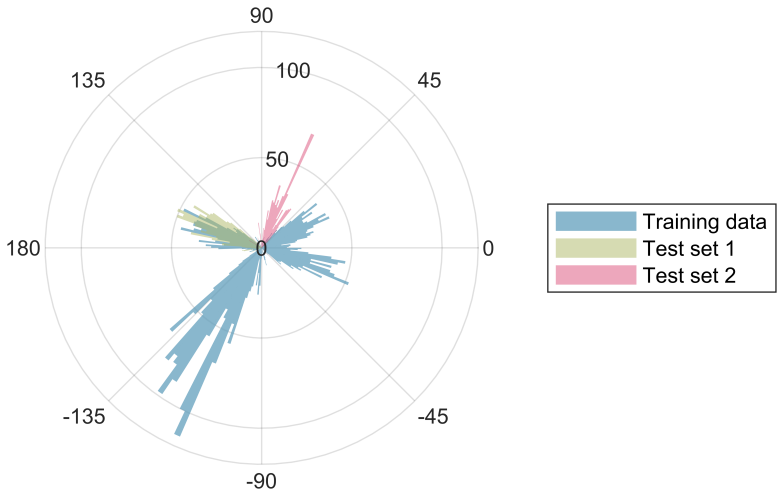


Figure 4.4: A polar histogram showing the wind direction ranges of training and testing data.

A summary of the training and testing data sets used in the case study is given in Table 4.1. Note that the numbers of data points for the training period in Table 4.1 correspond to the numbers obtained after subsampling the original SCADA data by a factor of two (in order to reduce computation time), whereas the testing data are not subsampled. Thus, to make test predictions, the model interpolates not only in space but in time. The total number training points is 3672, which corresponds to about 25.5 hours' worth of data for all 48 turbines in the Lillgrund wind farm. The training of the GP model using the selected data set takes about 30 minutes using 6

logical cores.

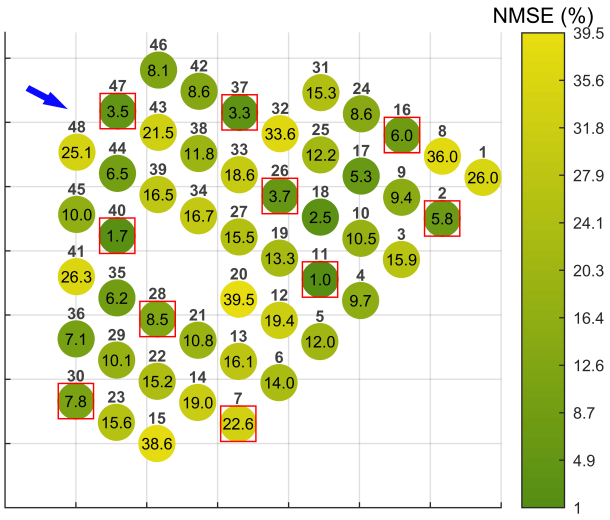
	Wind direction (θ) range	Normalised wind speed range	Normalised power range	Number of data points	Time period covered (hours)
Training sets	$12.2° \leq \theta \leq 56.2°$	0.18-0.81	0.15-0.91	336	2.3
	$10.1° \leq \theta \leq 55.4°$	0.24-0.91	0.21-0.95	312	2.2
	$-33.3° \leq \theta \leq 13.0°$	0.06-0.90	0.09-0.93	408	2.8
	$-40.5° \leq \theta \leq 7.6°$	0.28-0.98	0.18-0.95	360	2.5
	$-141.5° \leq \theta \leq -94.6°$	0.05-0.94	0.08-0.94	1080	7.5
	$-140.0° \leq \theta \leq -93.2°$	0.09-0.88	0.10-0.84	408	2.8
	$149.3° \leq \theta \leq 180°$ and $-180° \leq \theta \leq -165.0°$	0.02-0.88	0.07-0.81	456	3.2
	$140.4° \leq \theta \leq 180°$ and $-180° \leq \theta \leq -169.6°$	0.21-0.89	0.19-0.95	312	2.2
Test set 1	$140.4° \leq \theta \leq 180°$ and $-180° \leq \theta \leq -170.0°$	0.00-0.68	0.08-0.59	816	2.8
Test set 2	$50.3° \leq \theta \leq 97.9°$	0.09-0.72	0.00-0.64	624	2.2

Table 4.1: A summary of training and testing data sets.

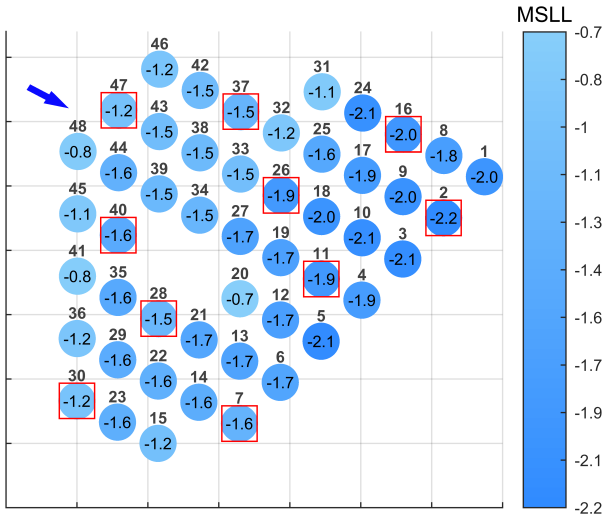
4.3.2 Capability to Predict Normal Power Production

The trained GP model is used to predict the power output during a testing period, when all turbines are operating under normal conditions. This example refers to test set 1 in Table 4.1. The prediction errors — NMSE and MSLL — are shown as error maps in Figure 4.5. Here each data point represents a turbine in its geographical location, with the turbine numbers (1 to 48) indicated next to each point. The data points are coloured according to their respective error values; the error values are also presented inside the circles. The incoming wind direction is indicated by the blue arrow, and the reference turbine locations are illustrated using blue rectangles.

It is seen that no specific spatial pattern can be found in the NMSE map (Figure 4.5a), whereas the MSLL results tend to get better towards the downstream side of the farm (Figure 4.5b). Such a spatial trend in the MSLL values is arguably caused by the calculation of the MSLL requiring an indirect comparison of the mean and variance between training data and GP predictions, i.e. the two terms in Equation 4.2 correspond to the log loss obtained by comparing the testing data with GP predictions and training statistics, respectively. Since the predicted mean power levels tend to



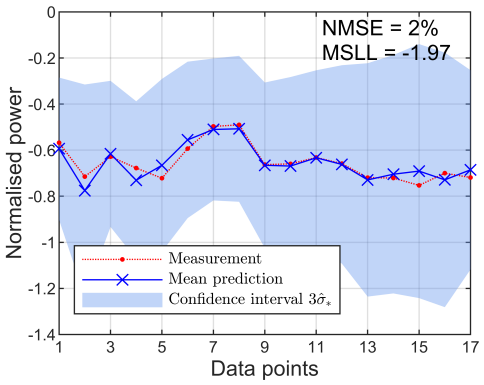
(a) NMSE.



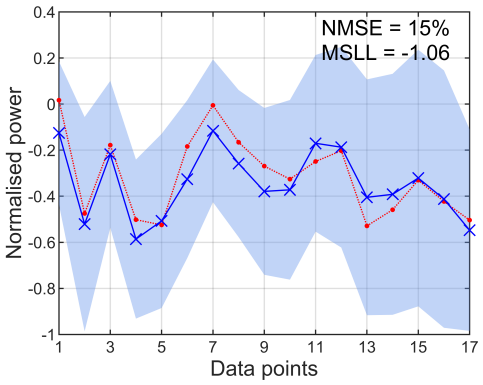
(b) MSLL.

Figure 4.5: Maps of (a) NMSE and (b) MSLL averaged across a testing data period (of about 2 hours 50 minutes) that describes normal operational conditions. The spacing between ticks along both axes is 10R.

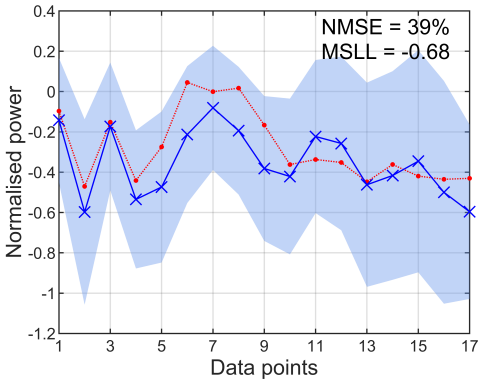
decrease in the direction of the wind, the MSLL values change in the wind direction as well, but the way the MSLL changes depends on how far away the predictions are from the training mean. Note that the predicted confidence intervals remain roughly constant throughout the farm. Overall, the error values are well below the corresponding thresholds (100% for NMSE; 0 for MSLL). At reference turbine locations, i.e. where wind speed measurements are provided as inputs to the model, there is a tendency to obtain relatively-better NMSE results; while whether reference inputs are provided does not seem to affect the MSLL values. The difference between



(a) Turbine 18.



(b) Turbine 31.



(c) Turbine 20.

Figure 4.6: Examples of the predicted and measured power time histories for the normal testing set, in cases of (a) small, (b) medium, and (c) large errors

the two error metrics is, of course, associated with the fact that only the mean predictions are used in the NMSE, while both the predictive mean and variance are included in the calculation of the MSLL.

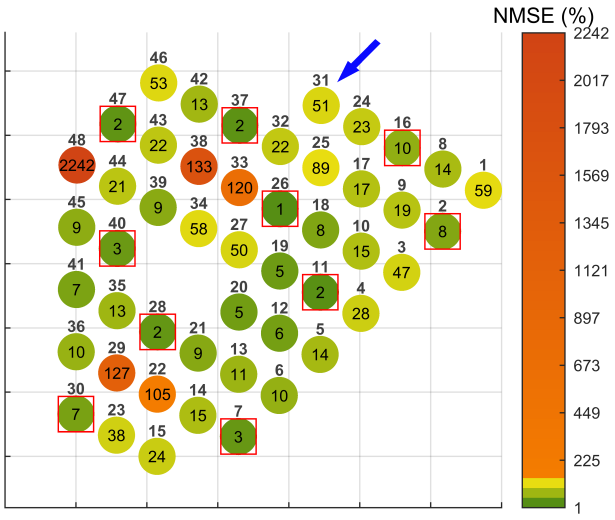
Examples of time series predictions are visualised in Figure 4.6, where the GP prediction constitutes a mean and a confidence interval ($3\hat{\sigma}_*$). Figure 4.6a illustrates one of the best predictions, excluding those at reference turbines. It is seen that the mean prediction follows the measured trend almost exactly, even at regions of relatively-wider confidence intervals (from data point 11 onwards). Figure 4.6b demonstrates the model prediction with a medium level of errors, where there are periods of mismatch between the mean prediction and the measurement, but the amount of deviation is considered small compared to the confidence interval. The

largest-error prediction is illustrated in Figure 4.6c; it can be seen that the mean prediction deviates from the measurement throughout the time window, while still loosely following the trend. In regions of large deviation (data points 5-9), the measured data are close to the boundary of the confidence interval, which leads to the relatively large MSLL as well as NMSE. In summary, the model is able to provide mean predictions that follow (at least roughly) the measured trends, as well as confidence intervals that capture all the benign power variations in the given testing data set.

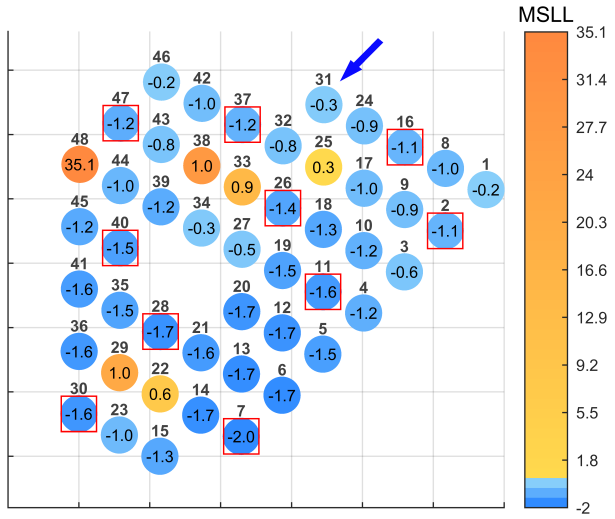
4.3.3 Performance Indicator for Wind Farms

Having established an appropriate model of normality, the model is used to predict the power production in a test set that contains potential performance anomalies, which refers to test set 2 in Table 4.1. Figure 4.7 shows the discrepancy between predictions and measurements in terms of maps of NMSE and MSLL. The colour schemes of the error maps are designed such that the turbines with an error value above the threshold are highlighted (in warm colours). The thresholds for NMSE and MSLL are 100% and 0, respectively, with reasons given in Section 4.2.3. It is seen that five turbines are highlighted as candidate anomalies in the NMSE map, i.e. turbines 22, 29, 33, 38, and 48, and there is one additional candidate anomaly, Turbine 25, highlighted in the MSLL map. Among these highlighted turbines, Turbine 48 is disqualified from being a potential anomaly, since it was manually switched off during the time period being considered. This power-down results in extremely high values of NMSE and MSLL. Although this is a known control action, it provides very basic validation that the model is able to detect anomalous behaviour. The other highlighted turbines will now be investigated. Note that the dataset available is currently unlabelled, i.e. the true reason for suspected anomalies is unknown — here some possible reasons for the observed behaviour will be explored.

The time histories of the predicted and measured power for the candidate anomalies (turbines 22, 25, 29, 33, and 38) are shown in Figure 4.8. For all candidates, there are obvious discrepancies between mean predictions and measurements throughout the time window. However, the trends of measured power are still captured by the confidence intervals most of the time. With respect to the error values, the further away the mean predictions are from the measurements, the higher the NMSE values; whereas higher MSLL values tend to result from scenarios when the measured data



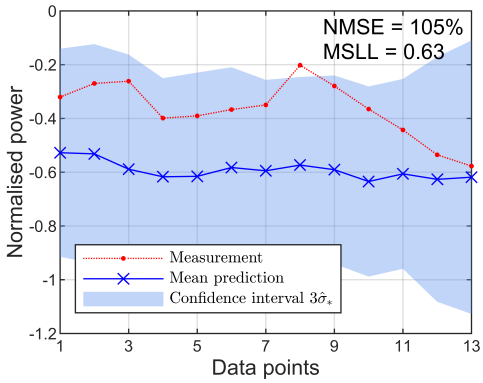
(a) NMSE.



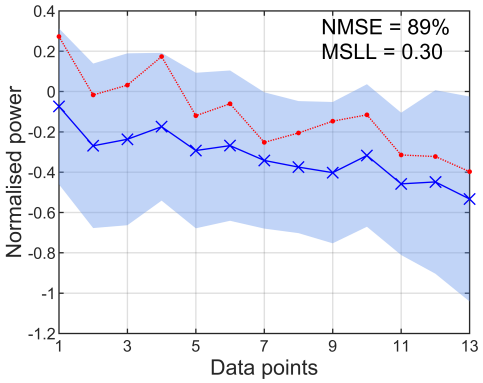
(b) MSLL.

Figure 4.7: Maps of (a) NMSE and (b) MSLL averaged across a testing period (of about 2 hours 10 minutes) with potential anomalies. The spacing between ticks along both axes is 10R.

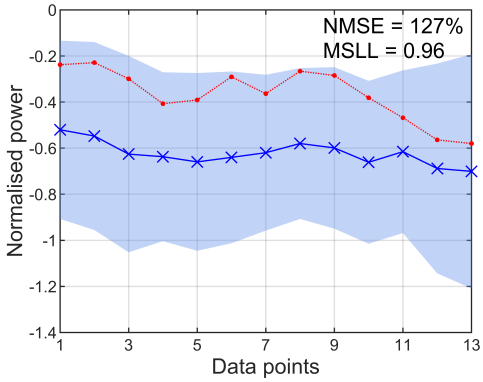
trends are consistently close to or beyond the uncertainty boundaries. In general, none of the candidate anomalies exhibits large fluctuations in power, and the relatively high errors seem to result from predictions with shifted means. Interestingly, the model underpredicts power in all five cases, although there is no consistent trend of underprediction across the entire wind farm; this may still imply the existence of a common factor that affected all five turbines during this period of time. Intuitively, underprediction in power may suggest that the model has predicted a wake effect stronger than that in reality. A possible reason for the overestimated wake effect is



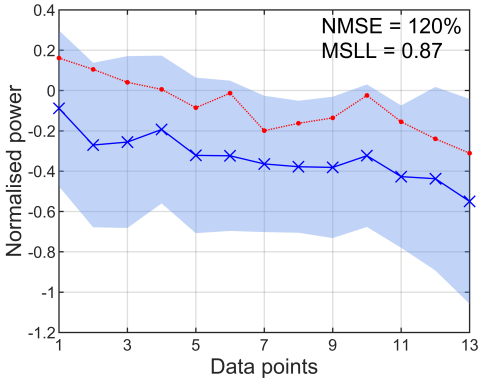
(a) Turbine 22.



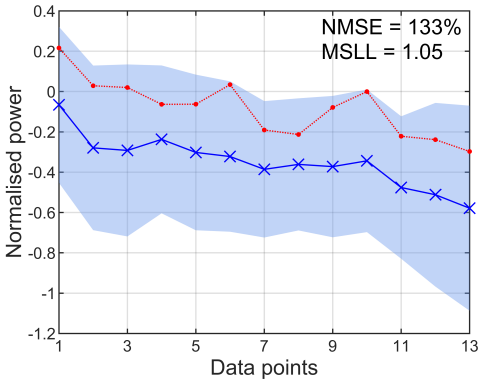
(b) Turbine 25.



(c) Turbine 29.



(d) Turbine 33.



(e) Turbine 38.

Figure 4.8: Time histories of the predicted and measured power for five highlighted candidate anomalies

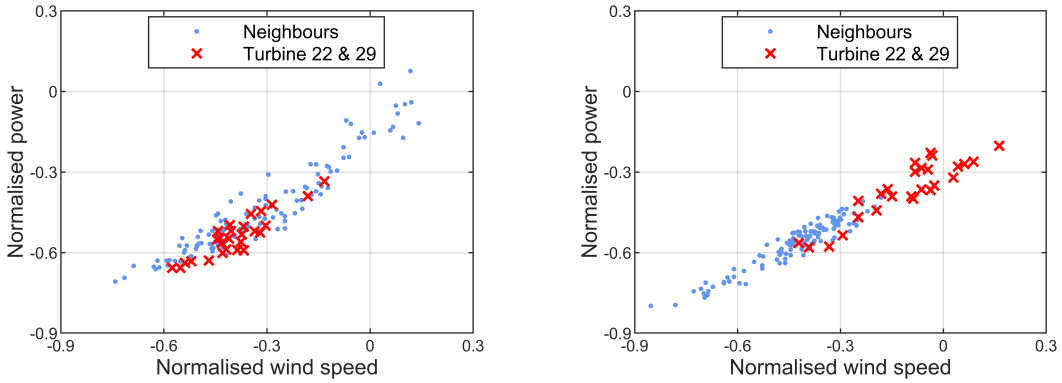
that the wind (nacelle) direction corresponding to test set 2 is not covered by the training data (Figure 4.4). A shifted global wind direction may give rise to an unseen spatial pattern, to which the model has struggled to extrapolate; this means that, although the model has captured the wake patterns across multiple wind directions (hence is able to make accurate prediction for test set 1), it is unable to extrapolate to unseen wind directions. In the following text, the (potential) reason for the high prediction error at each of the five candidate anomalies will be discussed in detail.

One reason for the high prediction error is an unexpected local nacelle direction. In particular, if a turbine turns away from the wake(s) of the upstream turbine(s), the power produced by the turbine may be underestimated by the model. Such an unexpected nacelle direction is found at Turbine 33. Normally, Turbine 33 should be affected by the combined wakes of the two upstream turbines (turbines 31 and 32). However, during the testing period, this turbine is constantly facing Turbine 37, a region likely to be affected by fewer wakes. Therefore, Turbine 33 ends up with more power output than predicted, as seen in Figure 4.8d, and this overprediction has been successfully highlighted by both maps in MSLL and NMSE.

Another cause of the high prediction error is a counter-intuitive environmental condition. Turbines 22 and 29 stand out from their neighbours as they experience unexpectedly-high wind speed and, thus, produce more power. Figure 4.9 shows the wind-power correlations of Turbine 22 and 29 in relation to those of the neighbouring turbines. As part of the normal condition acquired by the model, turbines 22 and 29 generate relatively-low power in comparison to their surrounding turbines (Figure 4.9a). However, in the testing period (Figure 4.9b), the wind and power at the two turbines extend far beyond the ranges covered by their neighbours, while maintaining the gradient of the wind-power correlations. A comparison between the testing and training data shows that extrapolation is required for the model to accurately predict the performance at turbines 22 and 29, explaining the underprediction seen in Figure 4.8a and Figure 4.8c.

The fact that the wind-power correlations follow the same gradient in Figure 4.9b means that the unexpected high power at turbines 22 and 29 is likely to stem from changes in the environment rather than changes in the turbine systems. Upon further examination of the data under similar environmental and operational conditions, the high wind at turbines 22 and 29 are unlikely to be caused by unexpected nacelle direction or ambient temperature, nor are there seasonal or diurnal patterns in the occurrence of such phenomena. On a closer look at the wind farm layout (shown in

Figure 4.1), one can notice the existence of a central gap in the wind farm next to turbines 28 and 35, where the water depth is insufficient for turbine construction [73]. Given the central gap, more wind eddies may be generated in the surrounding area, bringing about more short-term turbulence that causes the occasional high wind around turbines 22 and 29. The fact that these infrequent environmental variations are not described by the training data has led to seemingly sensitive detection results. However, the occasional exposure to wind eddies that give rise to counter-intuitive wake patterns may subject turbines 22 and 29 to higher fatigue loads and, thus, a higher risk of anomalous performance. The model results therefore provide insight into this potential risk.



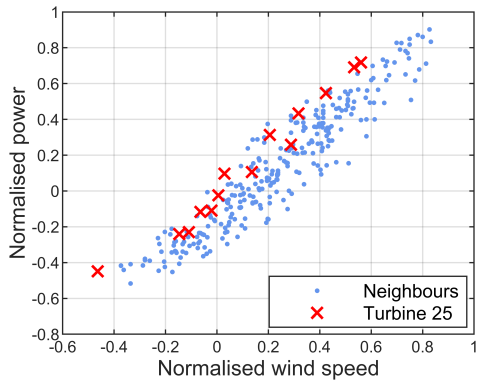
(a) Training data.

(b) Testing data.

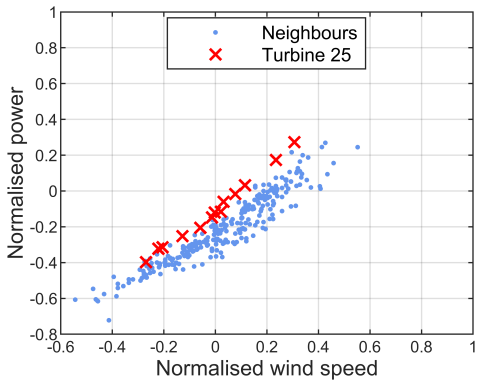
Figure 4.9: The wind-power correlation for a subset of turbines towards the back of the farm, i.e. turbines 13, 14, 15, 21, 22, 23, 28, 29, 30, 35, and 36, obtained from (a) training and (b) testing data

Lastly, the high prediction error can be the result of an anomalous spatial pattern in the wind-power correlation. This difference in spatial correlation is the main cause of the candidate anomalies at turbines 25 and 38. In Figure 4.10, the wind-power correlation for the first three turbine rows, i.e. the neighbourhood of turbines 25 and 38, are illustrated, with comparisons made available between training and testing data. A comparison between Figure 4.10a and Figure 4.10b illustrates how the spatial correlations between Turbine 25 and its neighbours change from training to testing data. It is seen that, although Turbine 25 tends to generate relatively high power, given the same wind speed compared to its neighbours during training (Figure 4.10a), the amount of extra power it generates in the testing period becomes more distinct from the neighbourhood (Figure 4.10b). However, what is shown in Figure 4.10b can (arguably) be merely a demonstration of the population variance; perhaps that is why Turbine 25 is flagged as a candidate anomaly by only one of the

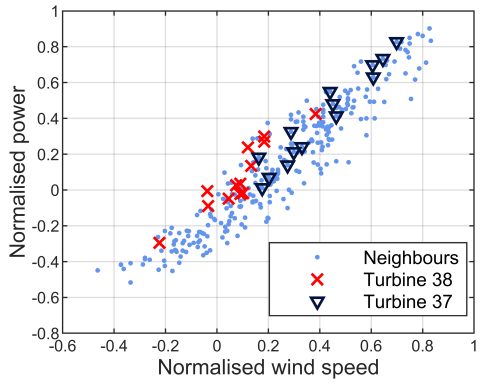
error metrics. For Turbine 38, the difference between training and testing data lies mainly in its relationship with the upstream Turbine 37. During training, the effect of wake shielding is apparent as the wind and power at Turbine 38 are significantly less than those at Turbine 37 (Figure 4.10c); however, this wake shielding effect is no longer seen in the testing period (Figure 4.10d). Wake deviations due to yaw misalignment may be a reason for such difference, which remains hypothetical since the relevant data for verification are unavailable to the author. The difference in spatial patterns around Turbine 38 is captured by both error metrics.



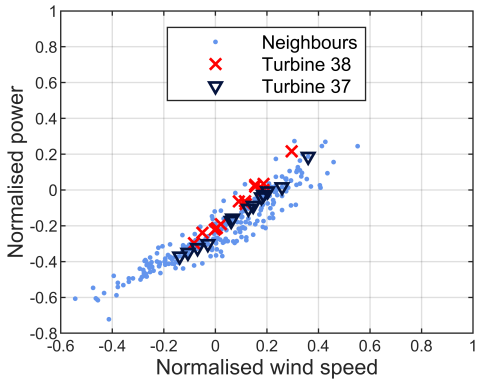
(a) Training wind-power correlations (with turbine 25 highlighted).



(b) Testing wind-power correlations (with turbine 25 highlighted).



(c) Training wind-power correlations (with turbines 37 & 38 highlighted).



(d) Testing wind-power correlations (with turbines 37 & 38 highlighted).

Figure 4.10: Wind-power correlations associated with (a)-(b) Turbine 25 and (c)-(d) Turbine 38. The "Neighbours" are referred to turbines in the neighbourhood of turbines 25 and 38, in this case the first three rows of turbines in the wind direction (excluding other candidate anomalies).

4.3.4 Effect of Reference Wind Speed Inputs

As previously stated, the GP model is trained to predict power across a wind farm, as a function of turbine locations and wind speed at a fixed set of reference turbines. In the previous case study, ten reference turbines are chosen, with their locations indicated in Figure 4.5 and Figure 4.7. This section focusses on investigating how the number of reference turbines may affect model accuracy.

An experiment was carried out to train the GP model described in Section 4.2.2 using various numbers of reference turbines. For each given number of reference turbines, five different sets of reference locations were selected, where every set of references was spatially distributed. A GP was trained with the same data specified in Section 4.3.1 for each set of reference turbines and tested on a relatively large testing data set that covers a wide range of wind directions, so that the results presented in this section do not correspond to a model tuned to a specific time window. The wind direction range of the test data set used in this section is illustrated in Figure 4.11; the test set corresponds to roughly 8 days' worth of data.

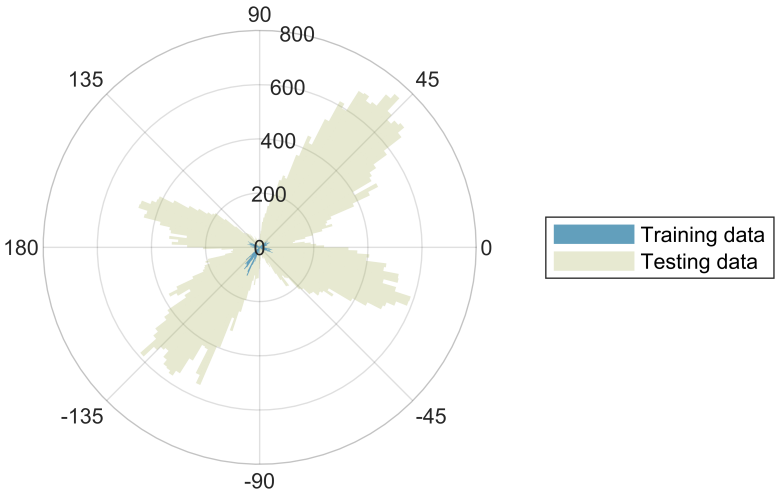


Figure 4.11: A polar histogram showing the wind direction ranges of training and testing data.

The NMSE results corresponding to different numbers of reference turbines are shown in Figure 4.12. It is seen that, as the number of reference turbines increases, there is an approximately exponential decrease in both the NMSE values and the error variances. If the reference locations are to be optimised, the NMSE improvements will be larger at fewer numbers of references. Optimised reference locations are considered to bring significant NMSE reduction when the number of references is smaller than

or equal to eight. The average NMSE level for ten references is around 30%. For the set of reference turbines used in the case study, the overall NMSE is around 20% across all normal testing data sets. Hence, in the current study, an optimised set of reference locations may not bring significant improvements in predictive accuracy. Nonetheless, an optimisation algorithm, such as a genetic algorithm [77], will be necessary if fewer references are to be selected in future studies.

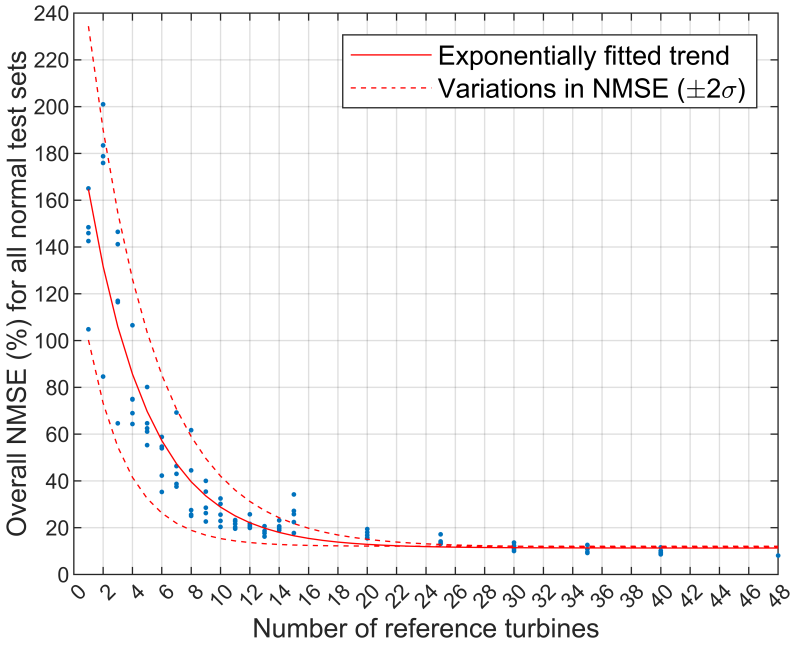


Figure 4.12: The NMSE for various numbers of reference turbines

4.4 Conclusions

The aim of this chapter is twofold. First, a population-based anomaly detection method is proposed here that accounts for the spatiotemporal correlations in environmental and operational conditions for the purpose of anomaly detection across a wind farm. The proposed mapping method is a two-step process that includes modelling and detection. In the modelling step, a GP model is trained, which captures the spatiotemporal variations of power across a wind farm under normal conditions. In the detection step, the discrepancy between predictions and measurements is quantified, and the discrepancy values above thresholds are considered (potentially) anomalous. Second, a case study is given to demonstrate the proposed method as a wind farm performance indicator.

The GP-based model of the normal condition has demonstrated its ability to learn the functional relationships including:

1. the correlation between wind speed at reference locations and power at all locations,
2. how the wind-power correlation varies spatially, and
3. how the spatial pattern changes with time-varying wind directions.

The GP model is associated with several assumptions. Firstly, the functional relationships captured by the model are assumed stationary. It indicates that the covariance is a function of the difference between input points rather than their absolutes, which greatly simplifies the complexity of the model [65]. Another key assumption arises from the chosen SE covariance function, which intrinsically assumes *smoothness*. Although some might argue that this function is too smooth to reflect any realistic process, the SE covariance function is still commonly employed across disciplines [65]. The smoothness assumption is acceptable here, as no apparent discontinuity is found in the data. First, the training and testing data are confined to a specific range of wind speed and power that represents the continuous region in the wind-power correlation (Figure 4.2). Second, the spatial variations in power are smooth when all turbines are operating normally, which is supported by the illustration given in Figure 4.3. Third, the training data are selected to avoid extreme fluctuations in wind directions, in an attempt to avoid discontinuities. As seen in Table 4.1, although the training data do not cover the entire range of possible wind directions, they cover the wind direction range of the normal test set (test set 1). The unexpectedly wide wind direction range in test set 2 is a result of anomalous turbine angles as explained in Section 4.3.3. Therefore, the smoothness assumption holds for the given case study. Both mentioned assumptions are the most commonly used and give rise to one of the simplest forms of GP model. Since the aim of the current work is to prove the concept of the proposed mapping method, it is reasonable to start with a relatively simple and easily-accessible model.

Two types of discrepancy measures are used in the analysis, namely NMSE and MSLL. The NMSE focusses on how the mean predictions deviate from the measurements, whereas the MSLL also accounts for the position of the data trends in relation to the boundaries of the confidence intervals. In the case when two turbines return similar NMSE values, the MSLL results might help distinguish which turbine has

the more confident prediction (and thus less likely to be erroneous). In summary, the two types of error metrics contain information that complements one another and should be used in conjunction.

The detection thresholds used in the case study are chosen based on the (mathematical) definitions of NMSE and MSLL (Section 4.2.3). As a preliminary study, the current work focusses on establishing the framework of the mapping method, while keeping the individual elements, such as the form of GP model and detection thresholds, as simple as possible. Thus, this work demonstrates the potential of the mapping method under the simplest possible setup, which justifies further investigation into this method. In future studies, more sophisticated methods of threshold selection, such as extreme value statistics [2, 13] may be considered.

As shown in the given case study, the candidate anomalies detected by the model can be a result of:

- powering off (unexpected wind-power correlation),
- unexpected turbine angles (unexpected spatial pattern),
- spatially counter-intuitive environmental conditions (unexpected spatial pattern), or
- other reasons that give rise to unexpected wind-power correlation in space.

The key assumption of the mapping method is that the spatiotemporal pattern of the target variable changes when anomalies occur. In Section 4.3.3, many of the candidate anomalies (turbines 22, 25, 29, and 38) are justified by comparing their wind-power correlations with those of the surrounding turbines. Such anomalies would be difficult to detect without considering the interrelations between turbines. Hence, the mapping method, as a population-based approach, possesses a unique advantage in detecting anomalies.

On the other hand, the fact that some anomalies result from the difference between training and testing data raises the question of whether the pattern described by the testing data is necessarily an "anomalous" one. To reduce false positives (and negatives), training data should be representative of the conditions expected — as is the case with any data-driven model. In practice, this may be difficult given the range of conditions/benign variations that may occur. This issue suggests that a grey-box approach may be suitable [75] and might be the topic of future work.

In conclusion, the case study has successfully demonstrated the capability of the mapping method to identify performance anomalies across a wind farm, using wind turbine SCADA data.

Chapter 5

The Gaussian Process Spatial Autoregressive Model with Exogenous Inputs

This chapter introduces the theory of the Gaussian process spatial autoregressive model with exogenous inputs (GP-SPARX), a model inspired by the physics of the wake effect to capture the spatiotemporal variation across a wind farm. In the context of PBSHM, the GP-SPARX model is a physics-inspired approach to modelling a population of structures, motivated by the task of wind farm monitoring. The chapter starts with a brief review on the physics-based models for the wake effect (Section 5.1), which motivates the use of spatial autoregressive methods for wind farm modelling. Section 5.2 discusses in detail how a wake pattern can be modelled spatial autoregressively. Section 5.3 introduces a GP-based spatial autoregressive method, in comparison with the parametric SAR model and the GP-NARX model for handling dynamic data. In Section 5.4, the two-level switching GP-SPARX method is introduced for modelling time-varying wake patterns.

5.1 The Spatial Autoregressive Nature of Wake Propagation

The current understanding on how turbine wakes propagate across a wind farm is established through the existing physics-based wake models (including analytical and numerical models), with a brief overview provided in Section 2.2.1. This section reviews three of the oldest and most widely used analytical wake models in detail, before deciding how this physics knowledge can be utilised to improve the modelling method of wind farm monitoring. The reviewed wake models include Jensen [78], Frandsen [79], and Larsen [80] models, all of which are approximated solutions of the Navier-Stokes equations, with varying levels of simplification. More details on the assumptions involved in these wake models will be discussed later in this section.

The notations used across Chapter 5-7 are defined in Figure 5.1. v_i and u_i are the wind speed immediately upwind and downwind of the i^{th} turbine rotor. Free-stream (or reference) wind speed is denoted as v_∞. Note that the term wind speed always refers to the wind velocity that is perpendicular to a specific rotor plane towards the downwind direction. The yaw direction is denoted as θ, and the angle between turbine yaw and the path connecting two turbines — the off-path angle — is denoted as $\Delta\theta$, where $\Delta\theta_i = |\theta_i - \theta_{\text{path}}|$. The distance between the i^{th} turbine and its upstream neighbour is denoted as d_i.

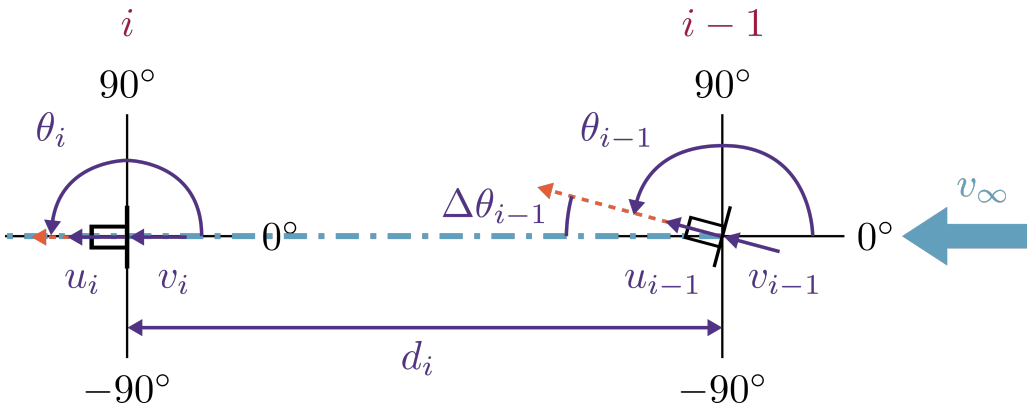


Figure 5.1: A schematic diagram of wake models.

The Jensen wake model is arguably the most simplified solution of the Navier-Stokes equations for the wake problem. It is based on the conservation of momentum across a control volume behind a turbine rotor. The major assumptions include that (a) the wake expands linearly downwind, (b) the ratio of wind speed reduction is only

dependent on the distance behind the rotor, and (c) the effect of vortex shedding is negligible in the wake region of consideration [24, 78, 81]. The wind speed at a wake-shadowed turbine, v_i, is modelled as

$$\begin{aligned} v_i &= v_\infty - \frac{\alpha_1}{(\alpha_2 + \alpha_3 d_i)^2} \left(v_\infty - a(v_{i-1})v_{i-1}\right) \left(\alpha_4 + \alpha_5 \cos(\alpha_6 \Delta\theta_{i-1})\right) \\ &= f_{\text{Jens}}\left(v_{i-1}, v_\infty, d_i, \Delta\theta_{i-1}\right) \end{aligned} \tag{5.1}$$

where $a(v_{i-1})$ is the axial flow induction factor, which is a function of the wind speed-dependent thrust coefficient $C_{\text{T}}(v_{i-1})$ and, thus, a function of v_{i-1}. α_1–α_6 denote the calibrated parameters of the model.

Similar to Jensen's approach, the Frandsen model is derived from applying Newton's Second Law (momentum equation) to a control volume, but in this case, a control volume across the rotor plane. The Frandsen model, like the Jensen model, assumes the dependency between wind speed reduction and distance behind rotor. Self-similarity of the axial wind speed profile and constant wind speed profile across the rotor area are also assumed [24, 79, 82]. The correlation between the wind speeds before and after a turbine rotor is indicated in

$$\begin{aligned} v_i &= v_\infty - \left(\frac{A_{\text{w}}(d_{i-1})}{A_{\text{w}}(d_i)} \left(v_\infty - v_{i-1}\right) + \frac{\beta}{A_{\text{w}}(d_i)} C_{\text{T}}(v_{i-1}) v_{i-1} \right) \\ &= f_{\text{Fran}}\left(v_{i-1}, v_\infty, d_{i-1}, d_i\right) \end{aligned} \tag{5.2}$$

where $A_{\text{w}}(d)$ is the wake area that is dependent on the distance behind rotor d, and β is a model constant. It is noted that the original Frandsen formulation is only applicable to rows of aligned turbines. To account for the off-path angle difference, a modification term containing $\Delta\theta_{i-1}$ may be added.

The derivation of the Larsen model starts from the incompressible, stationary Navier-Stokes equations, i.e. the differential form as opposed to the integral form used in the control volume approach in Jensen and Frandsen models. It is assumed (a) negligible pressure gradient, (b) circular symmetric wake profile across the rotor, and (c) self-similarity of axial wind speed. The Reynolds stress term in the governing equations is modelled according to Prandtl's mixing length theory [24, 80]. The wake

speeds are correlated in the following way,

$$\begin{aligned} v_i = v_{i-1}\Bigg\{ & \gamma_1 - \gamma_2\Big(\gamma_3 C_{\mathrm{T}}(v_{i-1}) d_i \cos(\Delta\theta_{i-1})\Big)^{\frac{1}{3}} \\ & \left[d_i^{\frac{3}{2}} \sin^{\frac{3}{2}}(\Delta\theta_{i-1}) \Big(\gamma_4 c_1(v_{i-1}) C_{\mathrm{T}}(v_{i-1}) d_i \cos(\Delta\theta_{i-1})\Big)^{-\frac{1}{2}} - \gamma_5 c_1(v_{i-1}) \right]^2 \Bigg\} \\ = {} & f_{\mathrm{Lars}}\left(v_{i-1}, d_i, \Delta\theta_{i-1}\right) \end{aligned} \tag{5.3}$$

where $c_1(v_{i-1})$ is a wind speed-dependent non-dimensional mixing length, and γ_1–γ_5 represent the model constants.

Albeit derived from different forms of the Navier-Stokes equations, these three models can arguably be used interchangeably; the choice of wake model is often a subjective opinion of the modeller.

Across the Jensen, Frandsen and Larsen models (and other fluid mechanics-based analytical wake models), the wind speed at a wake-shadowed turbine is depicted as a function of the wind input to its upwind turbine, i.e. v_i is a function of v_{i-1}. Since i denotes the spatial indices along a row of turbines, these wake models suggest that the wind variations across a wind farm are spatial autoregressive, with the remaining variables, v_∞, d_{i-1}, d_i, and $\Delta\theta_{i-1}$, serving as exogenous inputs to the autoregressive formulation. It is noted that in all three models a highly nonlinear correlation is demonstrated between the wake speeds (and the exogenous inputs).

5.2 Modelling a Wind Farm in a Spatial Autoregressive Manner

The physical understanding of wake propagation inspires the use of models based on spatial autoregression to capture the correlations across a wind farm. To the author's knowledge, spatiotemporal methods, including spatial autoregressive models, are rarely, if at all, used for wind farm monitoring. In the field of spatiotemporal modelling, models that assume autoregression in space are not frequently developed, since most models are designed to provide forecast in time instead of off-site predictions (i.e. autoregressive in time and stationary in space) [32]. However, in this work, the use of the spatial autoregressive formulation is physics-inspired, with the ability to

extrapolate in space being an additional benefit of the approach.

This section explores how the spatiotemporal variation across a wind farm can be modelled using a spatial autoregressive approach. A thought experiment is presented in this section that starts with the simplest scenario — modelling a row of turbines — before proceeding into a more complicated/realistic scenario — modelling an array of turbines. Different potential variants of the spatial autoregressive approach are also discussed, with the aim of selecting the most suitable variant for the wind farm monitoring problem.

5.2.1 Modelling a Row of Turbines

Given a row of turbines, if the direction of the incoming free-stream wind is parallel to that of the turbine row, then each downwind turbine is shadowed by the wake of the upwind neighbour (e.g. Figure 5.1). According to the summary on wake models given in Section 5.1, the wind speed in front of turbine rotors, v, is a spatial autoregressive process,

$$v_i = f\left(v_{i-1}, \mathbf{z}_i\right) \tag{5.4}$$

where $f(\cdot)$ is a nonlinear autoregressive function. The vector $\mathbf{z}_i$ represents the exogenous input variables, which may include v_∞, d_{i-1}, d_i, and $\Delta\theta_{i-1}$.

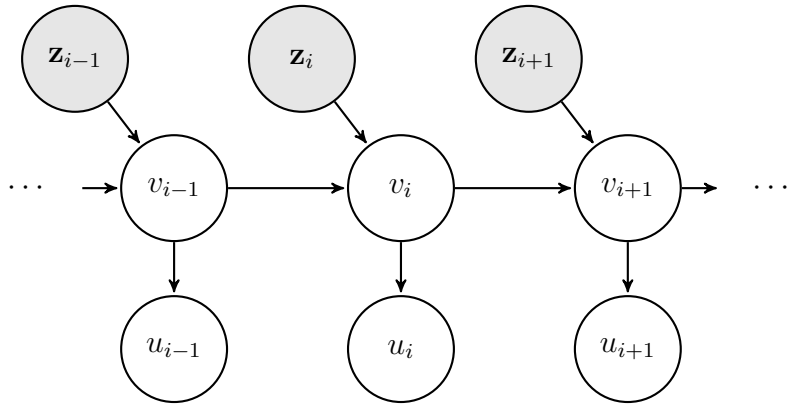


Figure 5.2: Graphic representation of spatial autoregressive wake models with exogenous inputs

According to the wake models, the spatial autoregressive description is only applied to the wind speed in front of turbines v_i. In operating wind farms, the values of v_i are not measured directly; wind speed measurements are taken immediately behind each rotor, giving rise to a new variable u_i. Figure 5.2 illustrates how the wind

speeds across turbine rotors and the exogenous variables are correlated in a spatial autoregressive manner. In theory, u_i can be calculated from v_i using the wake models by assuming $d_i \approx 0$ and $\Delta\theta_i \approx 0$,

$$u_i = f\left(v_i, \mathbf{z}_i\right)$$

and v_i can be expressed in terms of u_i,

$$v_i = f^{-1}\left(u_i, \mathbf{z}_i\right)$$

Hence, the observable variable u_i can also be described as spatial autoregressive process,

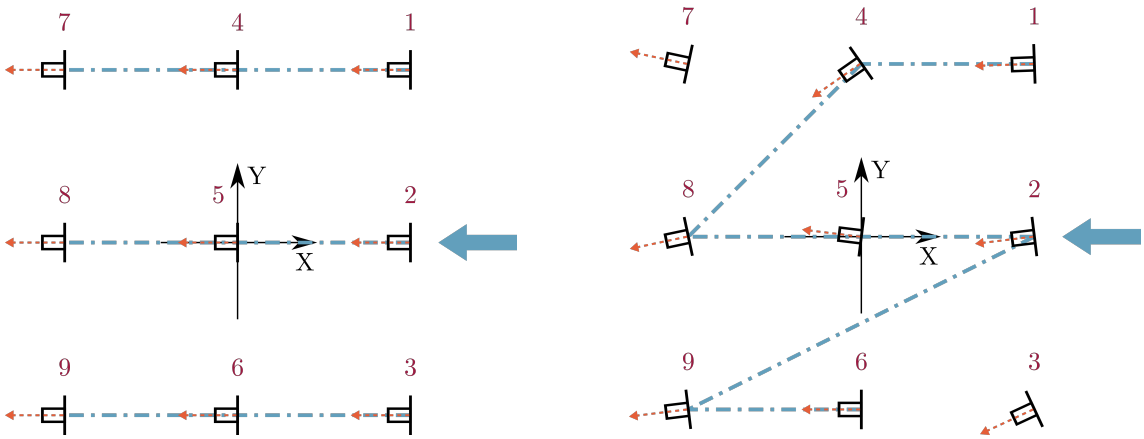
$$u_i = g\left(u_{i-1}, \mathbf{z}_{i-1}, \mathbf{z}_i\right) \tag{5.5}$$

5.2.2 Modelling an Array of Turbines: Wake Paths and Wake Patterns

In a spatial autoregressive process, dependence is conventionally found in all directions [66]. That is, the behaviour of turbine i is dependent on those of both turbine $i-1$ and turbine $i+1$ in the one-dimensional example given in Figure 5.1. However, the wake models implicitly define a direction of spatial dependence, in which dependence is only found between a wake-shadowed turbine and the turbine producing the wake, i.e. the behaviour of turbine i is only affected by that of turbine $i-1$.

Wind farms are often designed as a regular or irregular array of turbines. It is therefore important to identify the paths of wake propagation in a turbine array, along which the spatial autoregressive description applies. A *wake path*, therefore, refers to a path of wake propagation, which is determined by both the free-stream wind direction and the directions of local air vortices. All the wake paths aross a wind farm at a given time form a *wake pattern*. Figure 5.3 shows two wake patterns, where each wake path is indicated as a blue dash-dotted line. It is seen that, given the same turbine layout and free-stream wind direction, a wind farm may still show different wake patterns because of changing local air vortices and/or inconsistent control settings.

In the analysis that follows, it is assumed that the predominant wake effect is caused by the nearest upstream turbine on a wake path. For example, in Figure 5.3a,



(a) Aligned yaw directions (b) Varying yaw directions

Figure 5.3: Examples of wake paths in an array of wind turbines.

turbine 7 is subject to the wakes of both turbine 4 and turbine 1, but the effect of turbine 4 is assumed predominant. It is also an assumption commonly used in wake models, known as negligible wake superposition; this assumption applies to all three wake models reviewed in Section 5.1. In the language of autoregressive models, this assumption implies the use of one spatial lag; if considering the wake propagation as a stochastic process, this assumption implies the satisfaction of Markov property. A spatial autoregressive model assuming Markov property tends to focus more on the local wake effect/spatial pattern rather than the large-scale variations.

An *upstream neighbour* of a turbine has a wake that directly shadows the turbine, providing the predominant wake effect. Every pair of wake-affected neighbours is connected by a wake path. A turbine may have more than one upstream neighbours. Table 5.1 provides an exhaustive list of the upstream neighbours for each turbine in the wind farm example shown in Figure 5.3.

In the case of Figure 5.3b, not all the turbines have yaw directions aligned with those of the wake paths. The term *off-path angle* refers to the angle between the yaw direction of a turbine and a wake path associated with the turbine. A wake may be assumed as a roughly cone-shaped region behind a rotor (Figure 1.1) [26]. Hence, when the off-path angle of an upstream turbine is within a certain limit, its wake can still shadow a downstream neighbour along the wake path. The example given in Figure 5.3b assumes an upper limit of 20° [78] for the off-path angles.

In a spatial autoregressive setting, it is also useful to divide the turbines in a wind farm into various *layers* based on their positions along the wake paths, i.e. how many

Turbine	Upstream neighbour(s)	
	Aligned yaw directions (Figure 5.3a)	Varying yaw directions (Figure 5.3b)
1	-	-
2	-	-
3	-	-
4	1	1
5	2	2
6	3	-
7	4	-
8	5	4, 5
9	6	2, 6

Table 5.1: Upstream neighbours for the example turbine array.

times the free-stream wind is transformed to obtain the local wake environment. For example, turbine 7 is in the third layer in Figure 5.3a but in the first layer in the case of Figure 5.3b. Lists of the turbine members in each layer for both cases illustrated in Figure 5.3 are provided in Table 5.2.

Layer	Aligned yaw directions (Figure 5.3a)	Varying yaw directions (Figure 5.3b)
First	1, 2, 3	1, 2, 3, 6, 7
Second	4, 5, 6	4, 5, 9
Third	7, 8, 9	8

Table 5.2: Turbine layers for the example turbine array.

5.2.3 Modelling an Array of Turbines: Assumptions on the Variation across a Wind Farm

The spatial autoregressive model for the wake behaviours shown in Equation 5.5 can be implemented differently based on different assumptions about the spatial variation across a wind farm. In this section, three assumptions are discussed.

Firstly, the wake characteristics for all turbines across a wind farm can be assumed to be governed by the same spatial autoregressive process. It implies that Equation 5.5 holds true for $i = 1, \ldots, n_{\mathrm{L}}$, where n_{L} is the total number of turbine layers in the wind farm. For both wake patterns illustrated in Figure 5.3, $n_{\mathrm{L}} = 3$, with the constituent of each layer listed in Table 5.2.

The second assumption is that each wake path can be modelled as a separate spatial autoregressive process. For a wake pattern with n_{P} wake paths, a set of n_{P} autoregressive models will be used to fully capture the variation across the region,

$$u_{i,j} = g_j \left(u_{i-1,j}, \mathbf{z}_{i-1,j}, \mathbf{z}_{i,j} \right) \tag{5.6}$$

for $i = 1, \ldots, n_{\mathrm{L}_j}$ and $j = 1, \ldots, n_{\mathrm{P}}$. n_{L_j} indicates the number of turbine layers along the j^{th} wake path. The numbers of wake paths are $n_{\mathrm{P}} = 3$ and $n_{\mathrm{P}} = 5$ for the examples in Figure 5.3a and Figure 5.3b, respectively.

Lastly, it can be assumed that the correlation between each pair of wake-affected neighbours is characterised by a different model. This assumption aims to account for the nonlinearity stemming from individual turbine control settings, through the use of spatially varying model parameters. The resulting model is,

$$u_i = g_i \left(u_j, \mathbf{z}_k, \mathbf{z}_j \right) \tag{5.7}$$

for $i = 1, \ldots, n_{\mathrm{S}}$, where n_{S} is the total number of spatial locations (i.e. turbines) across the farm. In this case, i represents the turbine index. We define a function $e(\cdot)$ that finds all the upstream neighbours of a turbine, such that $H_i = h(i)$ is a list of the upstream neighbours of turbine i. Turbine j is an upstream neighbour of turbine i, i.e. $j \in H_i$. Similarly, $k \in H_j$, where $H_j = h(j)$.

In summary, these assumptions demonstrate various levels of simplification of the problem. The first assumption leads to the most parsimonious model by compromising generalisability or applicability, while the last model is flexible at the cost of a large parameter space. As discussed in Section 1.3.1 and Section 2.2, the problem of wind farm modelling is notorious for exhibiting complex behaviour, resulting from the structure-environment interactions. In such a context, it seems luxurious to opt for a highly parsimonious model, therefore, an autoregressive model with spatially varying parameters is likely to be the most desirable option for the problem at hand.

5.2.4 Multiple Spatial Lags

Based on the assumption that the predominant wake effect is provided by upstream neighbours, all variants of spatial autoregressive models derived in Section 5.2.1 and Section 5.2.3 can be regarded as AR(1) models in space. As the need arises, it

is possible to extend these models to include multiple spatial lags, which may be interpreted differently in the context of wind farm modelling.

Using multiple spatial lags in an autoregressive model can simply refer to adding more autoregressive terms that correspond to the upstream neighbours of the upstream neighbours. For instance, an autoregressive model with two spatial lags for turbine 8 in Figure 5.3b would contain as inputs the wind speeds from turbines 1, 2, 4, and 5.

Alternatively, only the upstream turbines whose wakes shadow the turbine of interest are accounted for. In this case, turbine 8 in Figure 5.3b is only correlated with turbines 2, 4, and 5 in a spatial autoregressive manner, since turbine 1, albeit being an upstream neighbour of turbine 4, does not provide wake shadowing effect to turbine 8. This option may lead to a model more readily interpretable from a physics point of view, since the resulting autoregressive model allows the effect of both the nearest and superimposed wakes to be considered. Nonetheless, both options release the assumption of the AR(1) wake process at the cost of an increased number of model parameters to be estimated.

Instead of using a physics-inspired model structure, the selection of spatial lags can be treated as an optimisation problem, where methods such as exhaustive searches can be used to determine the most appropriate lags [72].

5.2.5 Temporal Lags

The spatial autoregressive models discussed previously imply a temporally static relationship between spatial locations and the exogenous variables, which means that Equation 5.7 is a short form of,

$$u_{i,t} = g_i\left(u_{j,t}, \mathbf{z}_{k,t}, \mathbf{z}_{j,t}\right)$$

for $t = 1, \ldots, n_{\mathrm{T}}$, where n_{T} is the total number of time instances in a given time series data set. It is to be stressed again that the model parameters are spatially varying but constant in time.

This spatial autoregressive formulation can be extended to a spatiotemporal form by introducing temporal lags. Adding l_{T} temporal lags to the model with one spatial

lag results in a spatiotemporal autoregressive model,

$$u_{i,t} = g_i\left(u_{j,t-l_\mathrm{T}}, \ldots, u_{j,t-1}, \mathbf{z}_{k,t-l_\mathrm{T}}, \ldots, \mathbf{z}_{k,t}, \mathbf{z}_{j,t-l_\mathrm{T}}, \ldots, \mathbf{z}_{j,t}\right) \tag{5.8}$$

It is worth noting that the temporally static form of the spatial autoregressive model (Equation 5.7) is an example of approximating a spatiotemporal process as a spatially evolving time series (Section 3.1); such approximation captures the spatiotemporal correlations across a wind farm, without the necessity to extrapolate in time. The extended version of the model (Equation 5.8) is likely to be more capable of handling dynamics in time series data, while retaining the ability to extrapolate in space. In the context of wind farm monitoring, temporal extrapolation is not necessary if the normal (temporal) variations in turbine performance are captured in the training data, as demonstrated in Chapter 4. Hence, the rest of the thesis will be focussed on the spatial autoregressive form of the model, i.e. Equation 5.7, based on which a temporal autoregressive element can be added at a later stage if required.

5.3 The GP-SPARX Model

Section 5.1 has provided the motivation for modelling the spatiotemporal variation across a wind farm in a spatial autoregressive manner, followed by Section 5.2 where a discussion has been given on the different variants of spatial autoregressive models applicable to wind farm modelling. In this section, a GP-based spatial autoregressive model is introduced for capturing the spatiotemporal variation across a wind farm. The reason for using GP-based methods has been discussed previously in Section 4.2.1.

The GP-SPARX version of Equation 5.7 can be given as,

$$\begin{aligned} \mathbf{u}_i &= g_i\left(\mathbf{u}_j, Z_k, Z_j\right) \\ g_i\left(\cdot\right) &\sim \mathcal{GP}\left(\mathbf{0}, K_{SE}\left(\cdot,\cdot\right)\right) \end{aligned} \tag{5.9}$$

where $\mathbf{u}_i$ denotes the time series of wake behaviour at the i^{th} location, i.e. $\mathbf{u}_i = [u_{i,1}, \ldots, u_{i,n_\mathrm{T}}]'$. Z_j is a matrix containing the time series of the exogenous inputs corresponding to the j^{th} location, where $j \in H_i$ represents one of the upstream neighbours of the location i.

A GP with generic priors (zero mean and SE covariance) is shown in Equation 5.9, as an example of a purely data-based GP-SPARX model. Although these assumptions are considered acceptable at this stage (with reasons given in Section 4.4), the GP-SPARX formulation allows modifications by incorporating physics-informed mean and/or covariance functions. At the risk of repetition, Equation 5.9 refers to a set of GPs, each capturing the spatial correlation between a turbine and its upstream neighbours (Section 5.2.3).

5.3.1 Links to SAR and GP-NARX Models

In this section, the links and distinctions between the GP-SPARX model and the related methods in the literature, i.e. SAR and GP-NARX, are discussed.

As reviewed in Section 3.3, the SAR model is a parametric model for spatial autoregressive processes. A main motivation of the SAR model is to capture simultaneous spatial dependence, i.e. when the value at location i depends on that at location j and vice versa [67]. To model the underlying data generating process, the values at all spatial positions are to be solved simultaneously. In contrast, the GP-SPARX is motivated to model a wind farm, where the spatial dependence is assumed directional, with the direction defined through wake paths (Section 5.2.2); this allows the GP-SPARX to evaluate turbine performance sequentially across the wind farm, based on the value at an initial position (e.g. free-stream wind).

One of the main drawbacks of autoregressive models is the large parameter space involved. In the SAR model, a weight matrix is defined in order to reduce the number of parameters (Section 3.3); as for GP-SPARX, parameter reduction is attempted by incorporating known physics (Section 5.2) and adopting a non-parametric framework.

The GP-NARX model, on the other hand, may be regarded as a modification of GP regression to handle dynamic data, i.e. to extrapolate in time; a brief overview of GP-NARX has been given in Section 3.4. On the basis of GP-NARX, the GP-SPARX model contains an additional element of spatial autoregression, extending the scope of the model to space-time. Thus, the considerations made for GP-NARX are also applicable to the GP-SPARX model.

Firstly, the model is able to make two types of predictions based on the autoregressive input used, i.e. OSA and MPO predictions (Section 3.4). In a spatial autoregressive

setting, the MPO prediction implies the use of a spatial reference (values at an initial position) to predict across space; whereas the OSA prediction relies on the measurements at every location to make predictions at neighbouring positions. For the purpose of wind farm monitoring, the MPO prediction allows the correlations across a wind farm to be predicted based on the measurements from a nearby weather station and, thus, is a more desirable way of using the model.

The second consideration concerns the treatment of model uncertainty, especially for the MPO predictions. When the uncertainty in model predictions is significant, the noise from the model output is fed back into the input of the next step; hence, the uncertainty grows with distance away from the spatial reference. The propagated uncertainty can be estimated using various approaches such as a Monte-Carlo sampling approach [72]. Alternatively, when the uncertainty involved in model output is relatively low, this uncertainty can be assumed negligible; thus, the propagation of predictive uncertainty does not need to be considered [72].

The work in later chapters focusses on the latter approach — neglecting predictive uncertainty — for the following reasons. First, it greatly simplifies the implementation of the GP-SPARX model, as the sampling-based methods for estimating uncertainty will require a large amount of computation. Second, from a physics perspective, the most notable source of measurement noise in wind farm data is likely to be atmospheric turbulence. However, most of the case studies presented in this thesis are based on the SCADA data collected from an operating offshore wind farm; the data used in the analysis (i.e. ten-minute mean values) contain only information resulting from the mean wind, and modelling of the effect of turbulence is postponed to the next stage of the study (Section 4.1). It is therefore decided to assume negligible uncertainty at this stage, as the effect of turbulence is not under consideration in this work. Further modifications of the models to account for the accumulated prediction variance can be made afterwards, especially when the modelling of turbulence is of interest.

5.4 The Switching GP-SPARX Method

The wake pattern of a wind farm varies as the turbines turn to face different incoming wind directions. The naturally changing wind environment gives rise to time-varying wake patterns. In this respect, the modelling of a wind turbine array can be

considered a two-level problem. The first level concerns the spatial autoregressive wake progression across the farm given a specific wake pattern; the second level captures how the wake patterns change with time. One way to handle the two-level variations in a wind farm is by using the switching GP-SPARX method.

The idea here is to train a set of GP-SPARX models, each corresponding to a wake pattern, and use a high-level model to switch between the trained models to make predictions for multiple wake patterns. The switching model can be given as,

$$\mathbf{u}_i = \begin{cases} g_i^{(1)}\left(\mathbf{u}_j^{(1)}, Z_k^{(1)}, Z_j^{(1)}\right) & \underline{\theta}^{(1)} \leq \theta_i < \bar{\theta}^{(1)} \\ \vdots & \vdots \\ g_i^{(n_\mathrm{M})}\left(\mathbf{u}_j^{(n_\mathrm{M})}, Z_k^{(n_\mathrm{M})}, Z_j^{(n_\mathrm{M})}\right) & \underline{\theta}^{(n_\mathrm{M})} \leq \theta_i < \bar{\theta}^{(n_\mathrm{M})} \end{cases} \tag{5.10}$$

where the output $\mathbf{u}_i$ is obtained by switching between a set of n_M trained GP-SPARX models according to criteria based on the turbine yaw direction θ_i. $\underline{\theta}$ and $\bar{\theta}$ denote the lower and upper bounds of θ, respectively. The wake patterns (including variables such as $\underline{\theta}$ and $\bar{\theta}$) can be found by processing the data in a way demonstrated in Appendix A.

5.4.1 Model Selection

Various approaches may be used to select the GP-SPARX components to be included in a high-level switching model. This section presents two possible approaches.

The first approach is to work out all of the frequently seen wake patterns (using the method demonstrated in Appendix A) in the wind farm of interest. Training data can then be extracted from each of the patterns, and a full switching model can be formed with each wake pattern captured by a GP-SPARX model. This approach is relatively simple to implement, which only requires an additional step of data processing. However, issues may arise when the information on individual yaw directions is unavailable in a given data set, since there is insufficient information for the wake patterns to be determined. Another issue may be a result of frequently changing wake patterns; in this case, none of the wake patterns has enough data for model training. In these scenarios, a physical understanding of wake patterns may not lead to a feasible switching model.

Hence, an alternative approach is used, which aims to select an optimal set of

wake patterns to best represent the overall variations. The optimisation scheme of choice here is the greedy algorithm, which benefits from being an easy-to-implement algorithm with typically low time complexity. The algorithm solves the problem by selecting the best option from a candidate pool one at a time. In the context of model selection without the physical knowledge of wake patterns, estimated wake patterns are created by (evenly) sectioning the data based on wind directions, from which training data are extracted. A candidate pool then refers to a set of GP-SPARX models, each of which is trained with data that corresponds to an estimated wake pattern. In the first step, the greedy algorithm selects the candidate that gives the best MPO prediction for the testing data. Once a candidate is selected, it is removed from the candidate pool. In each of the following steps, the selected GP-SPARX models are combined with each of the remaining candidates, and the candidate that forms the combination with the best MPO accuracy is added to the switching model, until the required number of GP-SPARX components are selected. The greedy algorithm-based model selection process is described in Algorithm 1.

Figure 5.4 is a graphical illustration of the model selection process based on the greedy algorithm. Given a candidate pool with five training sets $n_{\mathrm{M}} = \{m_1, \ldots, m_5\}$, the goal is to select the best three candidates. The selection process is described as follows,

Step 1 There are five candidates $\{m_1, \ldots, m_5\}$. Each candidate is used to make MPO predictions for the test data set. The candidate with the best predictive accuracy, m_3, is selected as part of the final model, and m_3 is removed from the candidate pool.

Step 2 The selected model m_3 is combined with all four remaining candidates, forming four candidate two-switch models. The model that gives the best MPO predictions, m_3-m_4, is selected as the final model. m_4 is removed from the candidate pool.

Step 3 The selected model, m_3-m_4, is combined with all three remaining candidates, forming three three-switch models. The model that gives the best MPO predictions, m_3-m_4-m_1, is selected as the final model. The selection process is finished as enough candidates have been selected.

Algorithm 1 Greedy model selection

Inputs:

The number of candidates n_{C}
The number of GP-SPARX components required for the switching model n_{M}
Training data sets $\{Z_i, U_i\}$, for $i \in \{1, \ldots, n_{\mathrm{C}}\}$
Testing data set $\{Z^*, U^*\}$
A GP-SPARX model m
A set of candidate GP-SPARX models M_{C}
A set of selected GP-SPARX models M
A switching model made up of a set of GP-SPARX models $s(M, Z)$
A function of error evaluation $\mathrm{NMSE}(\hat{U}, U)$

Initialize:

iteration $= 1$
$M_{\mathrm{C}} = \{\}$
$M = \{\}$

for i $= 1$ to n_{C} **do**

train a GP-SPARX model m_i using training data set $\{Z_i, U_i\}$
$M_{\mathrm{C}} = M_{\mathrm{C}} \cup \{m_i\}$

end for

while iteration $< n_{\mathrm{M}}$ **do**

$m \in M_{\mathrm{C}}$
$M' = M \cup \{m\}$
$\hat{U}^* = s(M', Z^*)$
$m_{\mathrm{best}} = \arg\min_m \mathrm{NMSE}(\hat{U}^*, U^*)$
$M = M \cup \{m_{\mathrm{best}}\}$
$M_{\mathrm{C}} = M_{\mathrm{C}} \setminus \{m_{\mathrm{best}}\}$
iteration $=$ iteration $+ 1$

end while

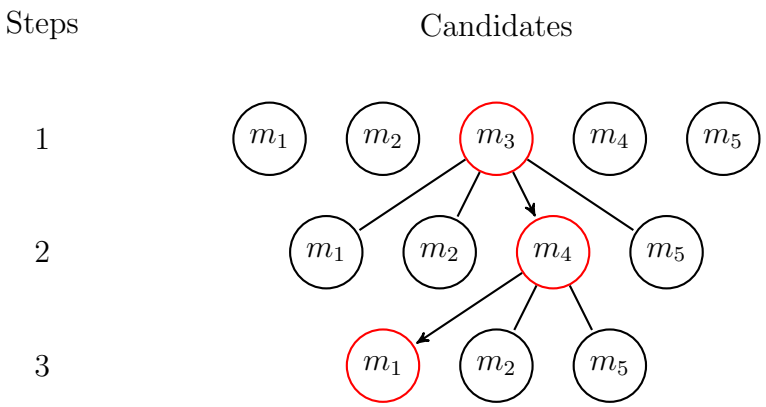


Figure 5.4: Graphic representation of the greedy model selection process with five model candidates.

5.5 Conclusions

In this chapter, the GP-SPARX has been introduced as a data-based model with a physics-inspired model structure, which captures the spatiotemporal variation across a wind farm resulting from the wake effect. Being physics-inspired implies that, to a certain extent, the GP-SPARX reflects our physical understanding of the wake effect while taking the advantage of a stochastic data-driven learner. The studies on the wake effect suggest a spatial autoregressive relationship between the wind inputs to successive turbine rotors (Section 5.1), which leads to a spatial autoregressive way of modelling turbine performance (Section 5.2). Across an array of turbines, wake paths are defined to clarify the spatial autoregressive relationship between turbines that varies with wind directions; given a specific wind direction, all the wake paths across a wind farm form a wake pattern (Section 5.2.2). Various assumptions on the farm-wise wind variation are discussed in Section 5.2.3, based on which a spatial autoregressive formulation with spatially varying parameters is considered appropriate for the wind farm modelling problem. The addition of spatial and temporal lags is discussed in Section 5.2.4 and Section 5.2.5, respectively; justification is given for the temporally static spatial AR(1) model chosen for further analysis. Section 5.3 introduces the GP version of the chosen spatial autoregressive formulation, and the resulting GP-SPARX model is compared with related methods such as SAR and GP-NARX in Section 5.3.1. To model the time-varying wake patterns arising from the time-varying wind directions, the two-level switching GP-SPARX method is suggested in Section 5.4; as a result, the time-varying wake patterns are modelled by switching between multiple GP-SPARX models, each corresponding to a specific wake pattern. In Chapter 6 and Chapter 7, the introduced methods will be applied to simulated examples and real wind farm data, respectively.

Chapter 6

Modelling a Simulated Wind Farm Using GP-SPARX

The GP-SPARX model has been introduced in Chapter 5 as a physics-inspired method for modelling the spatiotemporal variation across a wind farm. This chapter aims to evaluate the ability of GP-SPARX to model the spatiotemporal correlation between turbines simulated using a nonlinear spatial autoregressive formulation, before applying the model to real wind farm data in Chapter 7.

6.1 A Simulated Wind Farm

The focus of this chapter is to provide a case study for the GP-SPARX model in order to verify its capability to model a wind farm. Simulated wind farm data can be generated to describe known physics in normal condition, as opposed to real data that may be subject to performance anomalies, sensor disconnection, and inconsistent sampling rates. Hence, at this stage, simulated data are considered so that unsatisfactory model prediction is unlikely to be associated with actual anomalies, which may be present in operational data, but a representation of the modelling capability.

6.1.1 The Wind Farm Layout

The simulated wind farm contains a regular array of nine turbines, with the layout illustrated in Figure 6.1. The turbines are assumed to be of the same type and configuration, with the same rotor radius R. An equal spacing of $10R$ is assumed between any pair of turbines. The local wind environment at each turbine location (i.e. the behind-rotor wind speed) is computed from a predefined free-stream wind speed and direction.

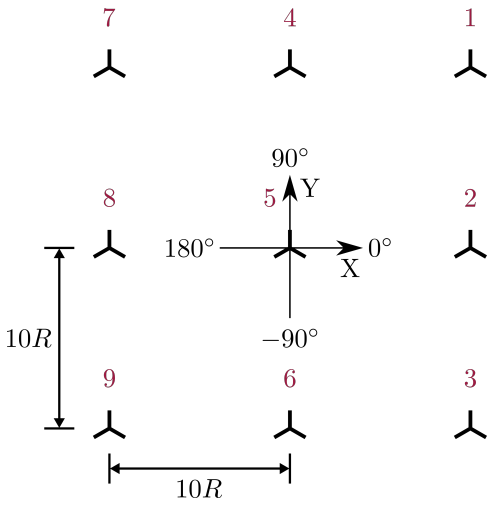


Figure 6.1: The simulated wind farm.

6.1.2 The Free-Stream Wind Environment

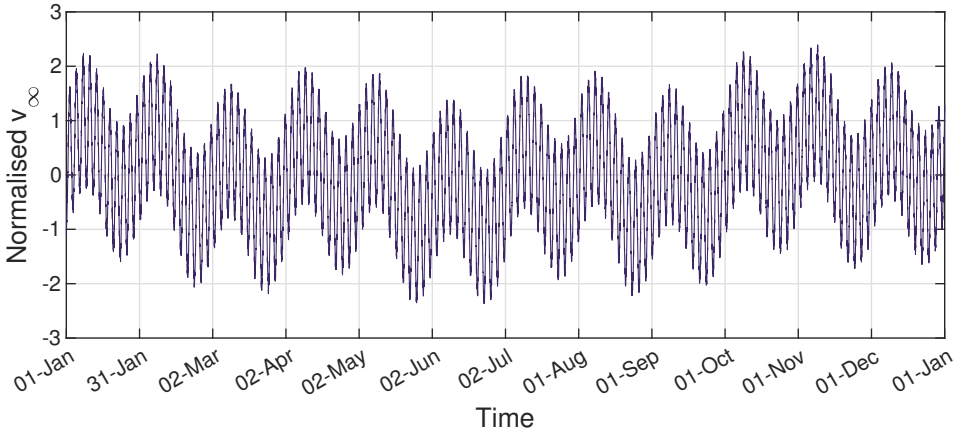
As part of the free-stream wind environment, the temporal variations in both the free-stream wind speed (v_∞) and direction (θ_∞) are simulated. The simulation covers the temporal variations over the period of a year at the SCADA frequency (i.e. a data point every ten minutes). Only the mean variations in wind are simulated in this case, whereas the turbulent effects are reserved for future studies. An independent and identically distributed (i.i.d.) Gaussian white noise — with a standard variation that amounts to 5% of the standard deviation of the noiseless values — is added to account for the natural perturbations of data and to help with model regularisation.

The free-stream wind speed consists of the sinusoidal trends representing the inter-

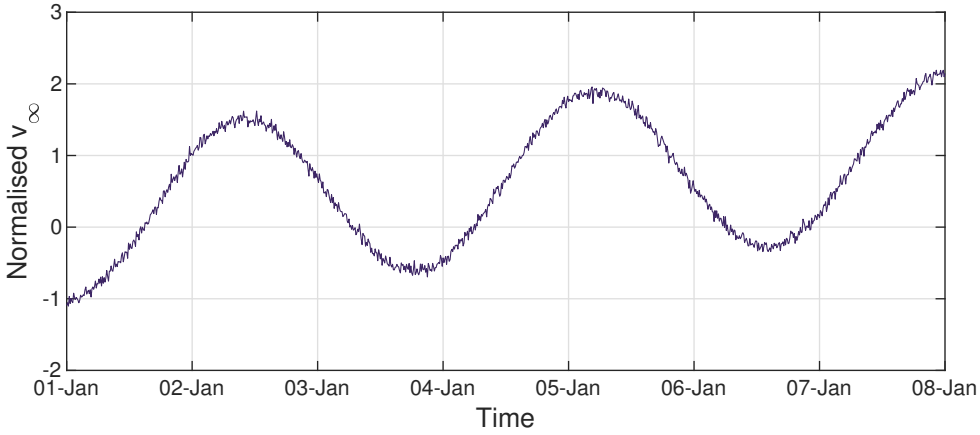
annual, annual, seasonal, monthly, and biweekly variations, as given by

$$v_\infty = 12 + 5 \times 10^{-6} t + 1.5 \sin\left(\frac{\pi t}{131400}\right) + 1.5 \cos\left(\frac{\pi t}{26280}\right) + 1.5 \sin\left(\frac{\pi t}{6570}\right) + 3.75 \sin\left(\frac{\pi t}{2190}\right) - 6 \cos\left(\frac{28\pi t}{1825}\right) \tag{6.1}$$

where the timestamps are $t = 1 \ldots 52560$, and the time increment is equivalent to 10 minutes (i.e. SCADA frequency). The fact that the free-stream speed is modelled as a linear combination of these sinusoidal trends implies that the mean wind variation is a multi-scale phenomenon, even though turbulence is not accounted for. The overall trend is seen in Figure 6.2a, with a closer view provided in Figure 6.2b for the first week of the simulated period.



(a) Over the period of a year



(b) Over the period of the first week

Figure 6.2: The simulated free-stream wind speed.

The free-stream wind direction comprises annual, seasonal, and bimonthly sinusoidal trends, with the combination illustrated in Figure 6.3. The simulated wind direction is

unevenly distributed to show a predominant wind coming from $-180° < \theta_\infty < -135°$ and $-45° < \theta_\infty < 0°$, which provides a more realistic representation of the situation in an offshore wind farm. Note that both predefined trends in wind speed and direction are in line with the theory [54] and the observations of an operating wind farm [25].

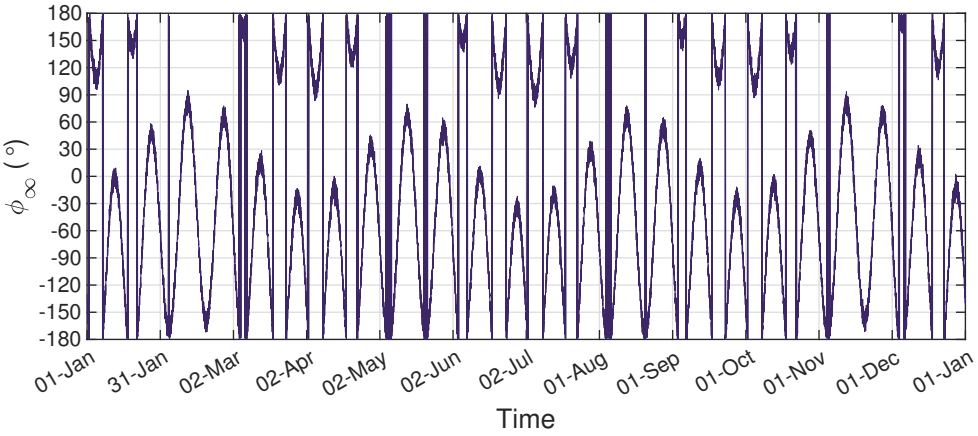
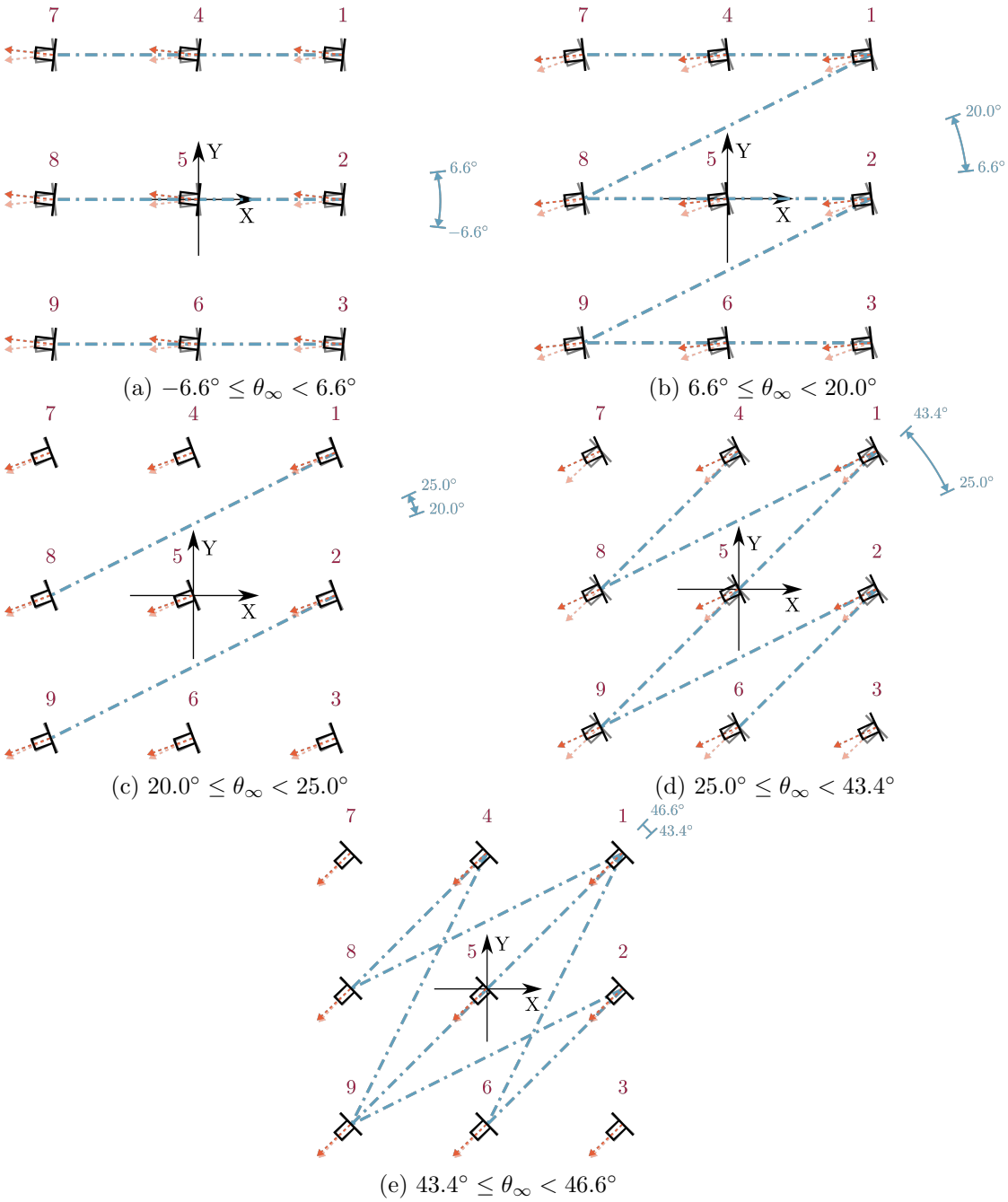


Figure 6.3: The simulated free-stream wind direction.

6.1.3 The Time-Varying Wake Patterns

As defined in Section 5.2.2, a wake pattern refers to all wake paths across a wind farm given a specific free-stream wind direction; different wake patterns result from time-varying wind directions. According to the default control settings, a wind turbine turns to face the incoming wind direction to extract energy from the wind. In a real wind farm, all turbines tend to face the free-stream wind direction, but with occasional deviations due to local vortices; these deviations are relatively small compared to a change in the free-stream wind direction. Thus, in this simulated case, all individual turbine yaw directions are assumed to align with the free-stream wind. The wake of every turbine is assumed to cover a region $\pm 20°$ of its yaw direction [78].

As the free-stream wind changes from $-6.6°$ to $46.6°$, five different wake patterns can be seen in the simulated wind farm, which are illustrated in Figure 6.4. All other wake patterns outside of the wind direction range of $-6.6° \leq \theta_\infty < 46.6°$ are considered similar to the ones shown in the figure (through rotation or reflection). A total of 32 patterns exist in the simulated wind farm.



(a) $-6.6° \leq \theta_\infty < 6.6°$

(b) $6.6° \leq \theta_\infty < 20.0°$

(c) $20.0° \leq \theta_\infty < 25.0°$

(d) $25.0° \leq \theta_\infty < 43.4°$

(e) $43.4° \leq \theta_\infty < 46.6°$

Figure 6.4: All wake patterns for $-6.6° \leq \theta_\infty < 46.6°$.

6.1.4 Correlation between Turbine Locations

Along every wake path, the correlation between turbines is assumed to be a nonlinear spatial autoregressive process with spatially varying parameters and static temporal mapping, i.e. following the form of Equation 5.7. One of the simplest nonlinear autoregressive models with exogenous inputs (NARX) — power-form polynomial — is chosen as the functional form in this simulation. The polynomial NARX model complies with the following empirical laws that are in line with our physical understanding of the wake effect [80]. For a downstream turbine at position i:

1. Four exogenous variables are considered relevant; the free-stream wind speed v_∞, the off-path angle(s) of the upstream neighbour(s) $\Delta\theta_j$, the off-path angle of the current turbine with respect to the wake path corresponding to the j^{th} upstream neighbour $\Delta\theta_{i,j}$, and the distance(s) from the upstream neighbour(s) d_j.

2. v_∞ only exists in a first order term, while $\Delta\theta_j$, $\Delta\theta_{i,j}$ and d_j only exist in terms of order 2 or higher.

3. All the terms of order 2 or higher aim to capture the interactions between variables, therefore these terms should contain more than one variable, e.g. terms such as d_j^2 are discounted.

4. The maximum order of the autoregressive variable u_j in any term is 1.

The result is a reduced polynomial NARX model of order 3,

$$
\begin{aligned}
u_i &= g_i\left(u_j, v_\infty, \Delta\theta_j, \Delta\theta_{i,j}, d_j\right) \\
&= \sum_{j\in H_i} \zeta_{1_j} u_j + \zeta_2 v_\infty + \sum_{j\in H_i}\sum_{p\in H_i} \zeta_{3_{jp}} u_j \Delta\theta_p + \sum_{j\in H_i}\sum_{p\in H_i} \zeta_{4_{jp}} u_j \Delta\theta_{i,p} \\
&\quad + \sum_{j\in H_i}\sum_{p\in H_i} \zeta_{5_{jp}} u_j d_p + \sum_{j\in H_i}\sum_{p\in H_i} \zeta_{6_{jp}} \Delta\theta_j \Delta\theta_{i,p} \\
&\quad + \sum_{j\in H_i}\sum_{p\in H_i} \zeta_{7_{jp}} \Delta\theta_j d_p + \sum_{j\in H_i}\sum_{p\in H_i} \zeta_{8_{jp}} \Delta\theta_{i,j} d_p \\
&\quad + \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{9_{jpq}} u_j \Delta\theta_p \Delta\theta_q \\
&\quad + \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{10_{jpq}} u_j \Delta\theta_p \Delta\theta_{i,q}
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{11_{jpq}} u_j \Delta\theta_p d_q \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{12_{jpq}} u_j \Delta\theta_{i,p} \Delta\theta_{i,q} \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{13_{jpq}} u_j \Delta\theta_{i,p} d_q \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{14_{jpq}} u_j d_p d_q \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{15_{jpq}} \Delta\theta_j \Delta\theta_p \Delta\theta_{i,q} \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{16_{jpq}} \Delta\theta_j \Delta\theta_p d_q \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{17_{jpq}} \Delta\theta_j \Delta\theta_{i,p} \Delta\theta_{i,q} \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{18_{jpq}} \Delta\theta_j \Delta\theta_{i,p} d_q \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{19_{jpq}} \Delta\theta_j d_p d_q \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{20_{jpq}} \Delta\theta_{i,j} \Delta\theta_{i,p} d_q \\
&+ \sum_{j\in H_i}\sum_{p\in H_i}\sum_{q\in H_i} \zeta_{21_{jpq}} \Delta\theta_{i,j} d_p d_q
\end{aligned} \tag{6.2}
$$

Given $|H_i|$ upstream neighbours, the total number of parameters required to specify the wake behaviour at turbine i is $13|H_i|^3 + 6|H_i|^2 + |H_i| + 1$.

6.1.5 The Simulated Wind Variation across the Farm

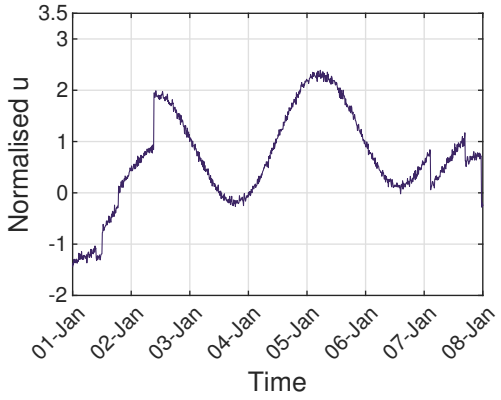
The resulting wind variation across the simulated wind farm is illustrated in Figure 6.5. The wind speed at each turbine follows the temporal trend of the free-stream wind; a shifted mean is seen when a turbine turns in or out of an upstream wake. In summary, the simulated wind variation follows these assumptions:

1. invariant wake shapes
2. constant wind profile across the width of a wake
3. sharp wake boundaries

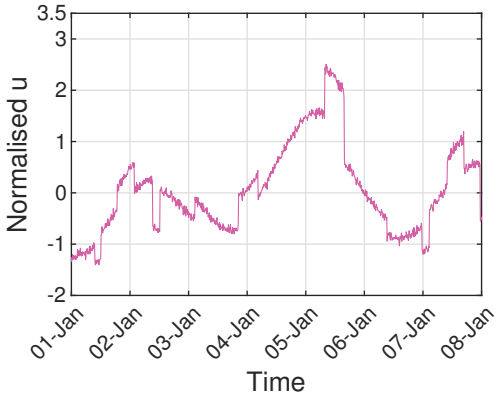
4. negligible wake superposition

5. separable mean and turbulent effects

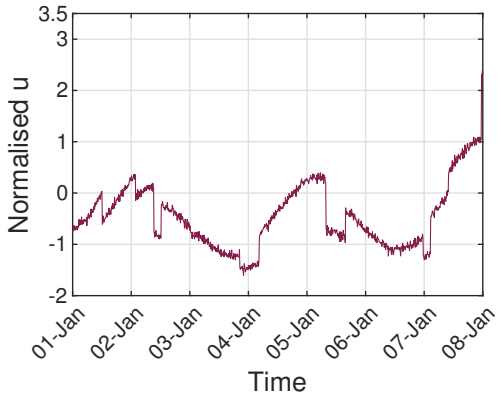
6. independent free-stream wind speed and direction



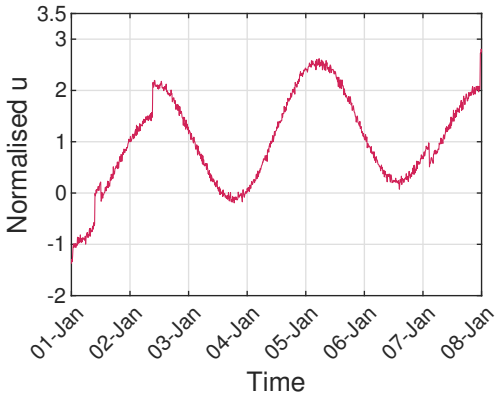
(a) Turbine 1.



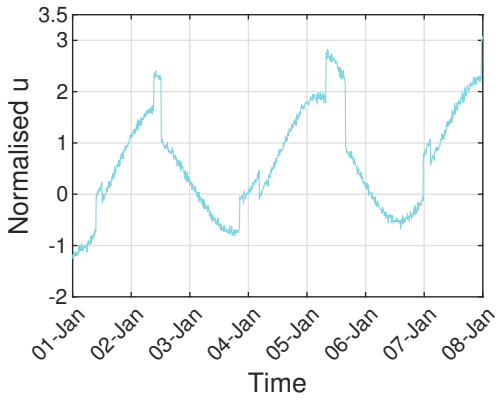
(b) Turbine 2.



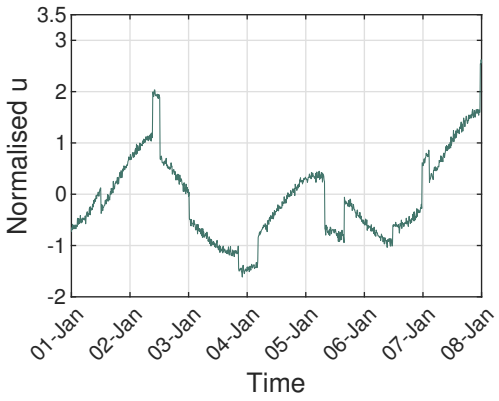
(c) Turbine 3.



(d) Turbine 4.

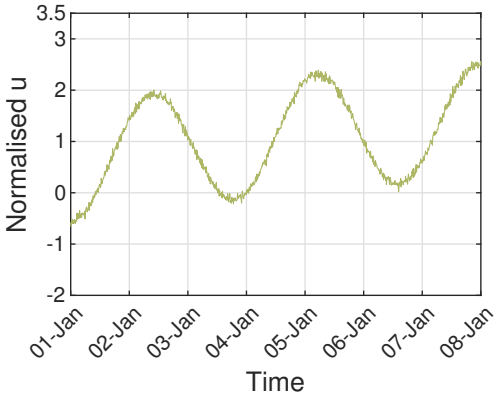


(e) Turbine 5.

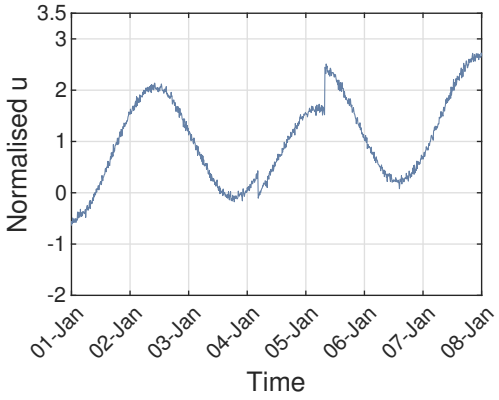


(f) Turbine 6.

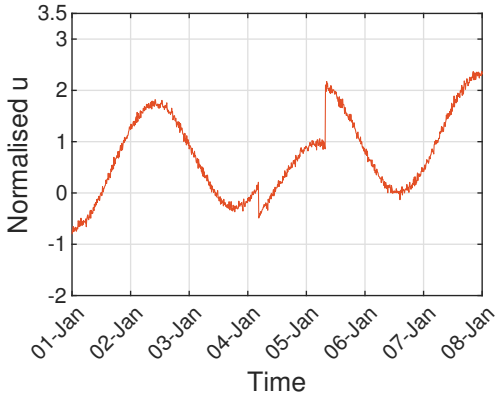
Figure 6.5: The simulated wind speeds at turbine locations over the period of the first week.



(g) Turbine 7.



(h) Turbine 8.



(i) Turbine 9.

Figure 6.5: The simulated wind speeds at turbine locations over the period of the first week.

7. no yaw misalignment
8. all yaw directions perfectly aligned with the global wind direction
9. static mapping in time, i.e. autoregressive only in space
10. the correlation between turbines is of a level of simplification similar to that of the analytical wake models

6.2 Modelling a Wake Pattern

6.2.1 Training and Testing Data

It is decided to use the wake pattern shown in Figure 6.4d to demonstrate the ability of GP-SPARX in modelling a single wake pattern, as Figure 6.4d represents the highest variation in wind directions across the five cases shown in Figure 6.4. The characteristics of the wake pattern are summarised in Table 6.1 and Table 6.2.

Layer	Turbine
First	1, 2, 3, 4, 7
Second	5, 6, 8
Third	9

Table 6.1: Turbine layers for the wake pattern of $25.0° \leq \theta_\infty < 43.4°$.

Turbine	Upstream neighbour(s)
1	-
2	-
3	-
4	-
5	1
6	2
7	-
8	1, 4
9	2, 5

Table 6.2: Upstream neighbours for the wake pattern of $25.0° \leq \theta_\infty < 43.4°$.

Figure 6.6 gives an indication of the training and testing data coverage in terms of wind direction and speed. From the wind variations simulated over the one-year period (Figure 6.3), time instances corresponding to $25.0° \leq \theta_\infty < 43.4°$ are extracted as the testing data, indicated in green in Figure 6.6. The training data (indicated in blue) are selected from the wind direction window $33.3° \leq \theta_\infty < 35.1°$, which is equivalent to one tenth of the size of the testing window. It is seen that the wind speed level drops as the wind proceed downstream. Nonetheless, the selected training data cover the range of wind speed variation in the testing data at each turbine location. Given a fixed wake pattern corresponding to the wind direction window, the chosen testing set assesses the models' ability to interpolate within a learnt range.

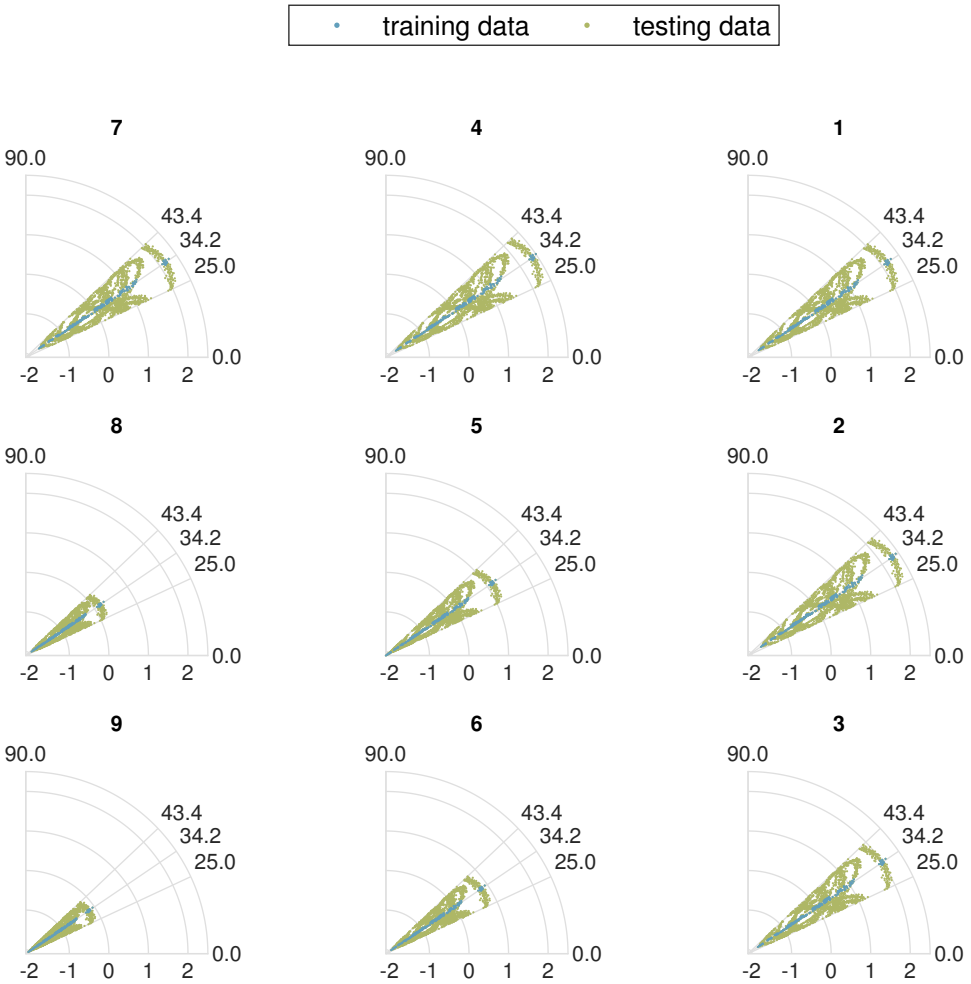


Figure 6.6: Training and testing data coverage. The angular and radial coordinates refer to the yaw direction and the normalised wind speed at each turbine location, respectively.

6.2.2 Models for Comparison

With the aim of assessing model performance, the results obtained from GP-SPARX are compared with those from deterministic spatial autoregressive models.

Selecting Exogenous Variables

According to the physics summarised in Equation 5.5, the relevant exogenous variables include the free-stream wind speed, the distance between turbine locations up to two spatial lags, and the off-path angles up to two spatial lags. Given the assumption of spatially varying parameters, there is one autoregressive function describing the

correlation between each turbine and its upstream neighbours (Equation 5.7). The distance between any pair of turbines stays constant over time and, thus, is unable to contribute any useful information to the model. The yaw directions are assumed to be aligned with the wind, therefore the off-path angles are the same for all turbines along the same wake path. In this respect, only the free-stream speed (v_∞) and one off-path angle (in this case, $\Delta\theta_j$) may contribute as model inputs.

The GP-SPARX Model

Two versions of the GP-SPARX model are tested to model the chosen wake pattern. The first version includes only v_∞ as the exogenous input, whereas the second option uses both v_∞ and $\Delta\theta_j$,

$$\mathbf{u}_i = g_i\left(\mathbf{u}_j, \mathbf{v}_\infty\right) \tag{6.3}$$

$$\mathbf{u}_i = g_i\left(\mathbf{u}_j, \mathbf{v}_\infty, \boldsymbol{\Delta\theta}_\infty\right) \tag{6.4}$$

where g_i refers to a GP such that $g_i\left(\cdot\right) \sim \mathcal{GP}\left(\mathbf{0}, K_{SE}\left(\cdot,\cdot\right)\right)$.

Linear Spatial Autoregressive Model

A simple linear spatial autoregressive model with exogenous input (LSPARX) is used to demonstrate how well (or badly) the simulated nonlinear wind variation may be modelled as a linear process. Similar to the GP-SPARX model, two sets of exogenous inputs are used in this analysis,

$$u_i = \eta_{1_j} u_j + \eta_2 v_\infty \tag{6.5}$$

$$u_i = \eta_{1_j} u_j + \eta_2 v_\infty + \eta_{3_j} \Delta\theta_j \tag{6.6}$$

wheres $j \in H_i$ for $i = 1, \ldots, N_\mathrm{S}$. The model parameters are denoted by η.

Nonlinear Spatial Autoregressive Model

A simple nonlinear spatial autoregressive model with exogenous input (NLSPARX) is created by adding the squared terms to the linear fomulation,

$$u_i = \lambda_{1_j} u_j + \lambda_2 v_\infty + \lambda_{3_j} u_j^2 + \lambda_4 v_\infty^2 \tag{6.7}$$

$$u_i = \lambda_{1_j} u_j + \lambda_2 v_\infty + \lambda_{3_j} + \lambda_{4_j} u_j^2 + \lambda_5 v_\infty^2 + \lambda_{6_j}^2 \Delta\theta_j \tag{6.8}$$

where λ denotes parameters. It is worth noting that the terms of squared wind speeds are not included in the original polynomial NARX model used to simulate the wind farm. Thus, Equation 6.7 and Equation 6.8 present one of the simplest functional forms of nonlinear autoregressive models, without knowledge of the governing physics.

6.2.3 Results

The NMSE results (averaged across the entire wind farm) for all models are presented in Table 6.3. A comparison between the training and testing errors reveals that, although all models tend to give better results for training predictions, they provide testing predictions with a level of accuracy similar to the training result. It shows that the selected training data are a good representation of the wake pattern described in the testing data.

Across all types of predictions (training or testing, OSA or MPO), the GP-SPARX models provide the most accurate result in terms of NMSE. The second best results are obtained from the NLSPARX models, and the LSPARX is the least suitable model out of the three candidates. It implies that the nonlinear relationship between turbines is best captured by the GP-SPARX and partially captured by the NLSPARX, whereas the linear models do not seem to fit this specific task.

For the NLSPARX models, there is still room to improve the predictive accuracy; if the physics of the underlying process is known, the functional form of the NLSPARX models can match with the governing physics. However, such physics knowledge is often incomplete or unavailable, especially for engineering structures operating in a complex environment. In contrast, the GP-SPARX models demonstrate their capability of modelling nonlinearity without incorporating physics.

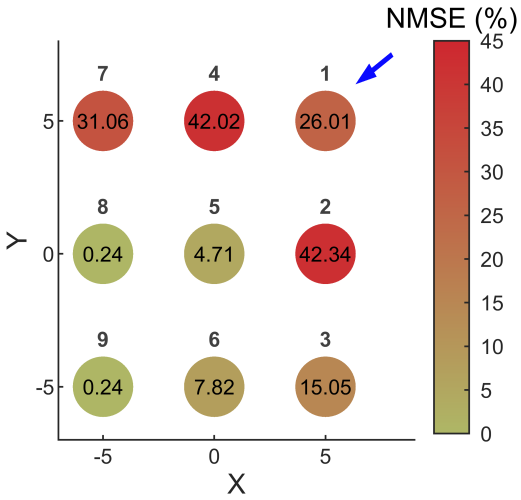
The effect of the exogenous input, $\Delta\theta_j$, differs by model and by prediction type. The

additional input variable brings positive changes to the parametric models (LSPARX and NLSPARX) in the MPO predictions of both training and testing data. The parametric models tend to make very poor MPO predictions with only one exogenous input (v_∞); thus, the addition of $\Delta\theta_j$ is likely to bring in the necessary information for the MPO predictions. In terms of the OSA predictions, the parametric models give slightly better training predictions and slightly worse testing predictions with the addition of $\Delta\theta_j$, where the difference in NMSE is insignificant. It implies that the parametric models are able to make accurate predictions with only the exogenous input of v_∞, given the exact information from the upstream neighbour; the addition of the second exogenous input ($\Delta\theta_j$) does not affect the OSA predictions too much. Lastly, the GP-SPARX model provides good OSA and MPO predictions with or without the second exogenous input ($\Delta\theta_j$); thus, this additional input variable seems unnecessary for GP-SPARX. Hence, in the analysis that follows, the GP-SPARX model refers to the form with one exogenous input (Equation 6.3) unless elsewhere stated.

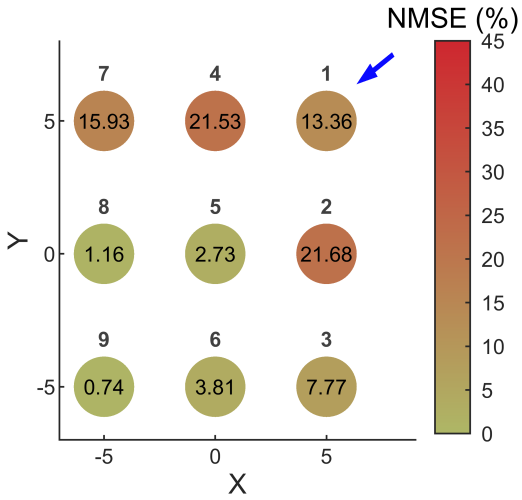
		OSA		MPO	
		v_∞	v_∞ & $\Delta\theta_j$	v_∞	v_∞ & $\Delta\theta_j$
Training prediction	LSPARX	16.16	15.96	38.86	16.00
	NLSPARX	9.20	9.01	19.06	9.05
	GP-SPARX	0.17	0.17	0.20	0.20
Testing prediction	LSPARX	18.07	19.51	44.58	18.83
	NLSPARX	9.33	10.29	20.44	9.86
	GP-SPARX	0.20	0.18	0.21	0.24

Table 6.3: Training and testing results in NMSE (%).

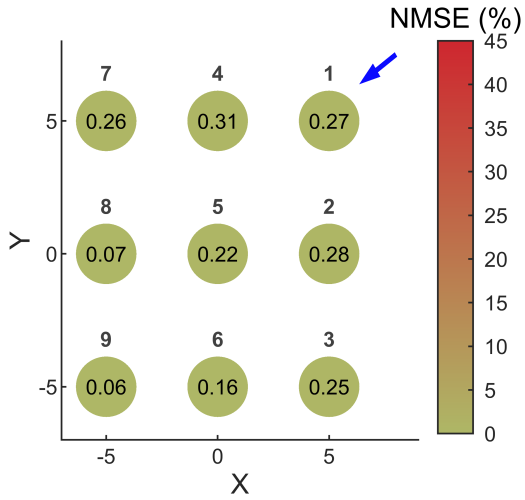
As MPO provides a more rigorous test of the performance of AR-based models [71], the best MPO testing results obtained from the three types of models are shown in Figure 6.7. Here, the NMSE of every turbine location is visualised in a map. What is common across all three NMSE maps is that the error values for the first layer turbines (1–3, 4, 7) are higher than the rest. The fact that these turbines have no upstream neighbours means that the predicted wind speeds at these positions are only based on the exogenous input of v_∞. In this regard, the reduced prediction errors at downstream positions is likely to be a result of providing the model with more input information (i.e. the autoregressive terms). However, the error drop from upstream to downstream locations is almost negligible in the GP-SPARX case compared to the other models. It is also seen that the GP-SPARX provides the most accurate prediction at every turbine location.



(a) LSPARX with two exogenous intputs



(b) NLSPARX with two exogenous intputs



(c) GP-SPARX with one exogenous input

Figure 6.7: NMSE maps of the best MPO predictions of the testing data.

To complete the comparison, the computational complexity corresponding to the three models shown in Figure 6.7 is summarised in Table 6.4. Note that the computational complexity refers to the running time for the computation of each u_i. In the case of both LSPARX and NLSPARX models, the parameters are estimated via the least squares method. For a turbine i with $|H_i|$ upstream neighbours, the inputs to the LSPARX and NLSPARX models are a $T \times (2|H_i|+1)$ matrix and a $T \times 2(2|H_i|+1)$ matrix, respectively. T is the number of training time instances (i.e. training samples). Since $T >> |H_i|$ ($T = 242$ and $0 \leq |H_i| \leq 2$), the computation of least squares estimate is dominated by matrix multiplication. In contrast, GP is notorious for its basic complexity of $O(T^3)$, because of the inversion of a $T \times T$ matrix. However, given the relatively small training data set used in this analysis, the cost difference

is hardly noticeable. One disadvantage of the NLSPARX approach is that the computational complexity will multiply if a higher-order functional form is chosen. On the contrary, the cost of GP-SPARX remains unchanged regardless of the number of exogenous inputs used. It is possible to further reduce the time complexity of GP by methods of sparse approximations, and the complexity can be reduced to $O(TM^2)$ or even $O(M^3)$, given M inducing points; more details on sparse GP can be found in [83].

	LSPARX	NLSPARX	GP-SPARX
Computational complexity	$O\left((2\lvert H_i\rvert+1)^2 T\right)$	$O\left(4\,(2\lvert H_i\rvert+1)^2 T\right)$	$O(T^3)$

Table 6.4: The computational complexity of the models.

6.3 Modelling Time-Varying Wake Patterns

As the wind environment changes naturally over time, a wind farm is subject to time-varying wake patterns. In this section, the switching GP-SPARX method will be used to predict the time-varying wake patterns in cases when (a) the knowledge of wake patterns is known and (b) the patterns are unknown.

6.3.1 With All Wake Patterns Known

Figure 6.8 illustrates the simulated wind variations over the period of a year, which consists of 32 wake patterns (separated by the dotted lines). The simplest method to model these wake patterns is to train a GP-SPARX model for each pattern. The training data are selected from a wind direction window equivalent to one fifth of the size of each wake pattern.

The testing results are shown in Figure 6.9 and Figure 6.10. The NMSE obtained in modelling multiple wake patterns (Figure 6.9) is at a level similar to that in the single-pattern case (Figure 6.7). Since the simulation data are obtained by assuming an instant switch between wake patterns, the abrupt shifts in the simulated wind speed are well captured by the discontinuous switching model when all wake patterns are known. Note that, in real wind farm (SCADA) data, shifts in measured mean wind speeds are likely to be observed when the wakes upwind of a turbine change;

however, such shifts in real data are likely to be more gradual (than those shown in Figure **??**), since the realistic wind environment may be associated with less idealised wake shapes, less defined wake boundaries, more complex wake superposition, and turbulence-induced effect.

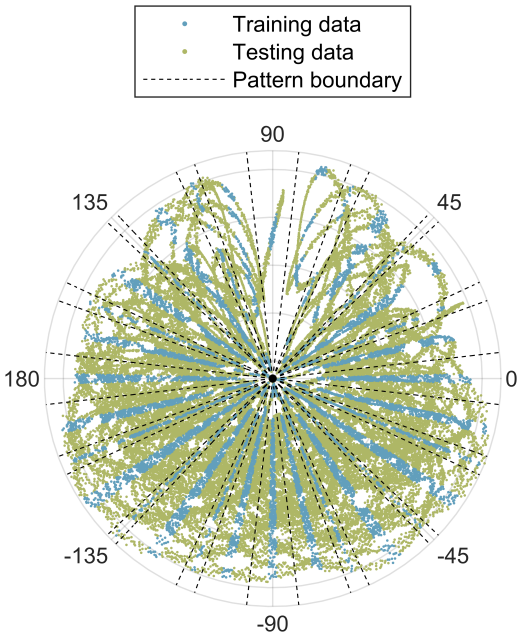


Figure 6.8: Training and testing data with wake patterns indicated. The angular and radial coordinates refer to the free-stream wind direction and speed, respectively.

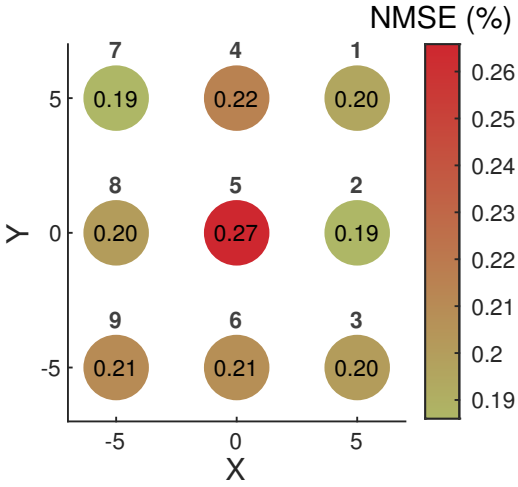


Figure 6.9: The NMSE map for the MPO testing predictions obtained from the full switching model.

Figure 6.10 shows the distribution of prediction errors across wind directions. Although the errors are small across all directions, the occasional high error is seen along the switching boundaries (i.e. wake pattern boundaries), especially in directions when a turbine is shadowed by wakes; this error trend will be discussed in detail in Section 6.3.2.

6.3.2 Without the Knowledge of Wake Patterns

An alternative approach is developed to handle situations when the information on wake patterns (e.g. wind direction measurements) is unavailable or when the patterns change too quickly that there is insufficient data for model training. In this

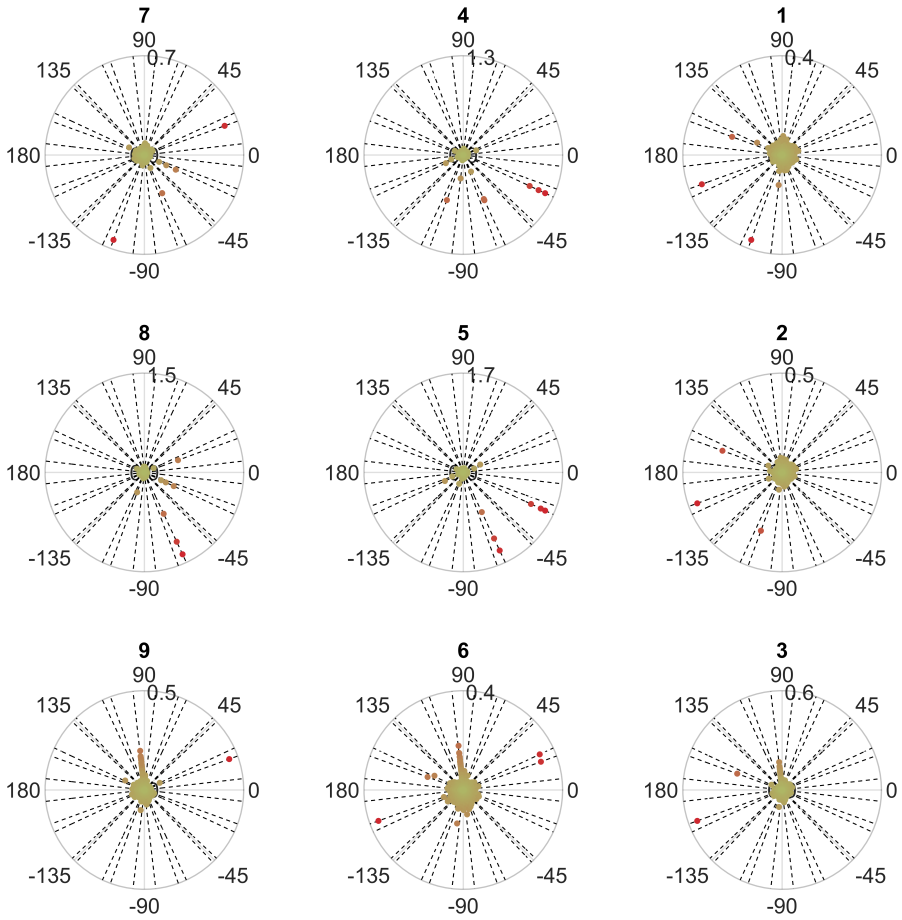
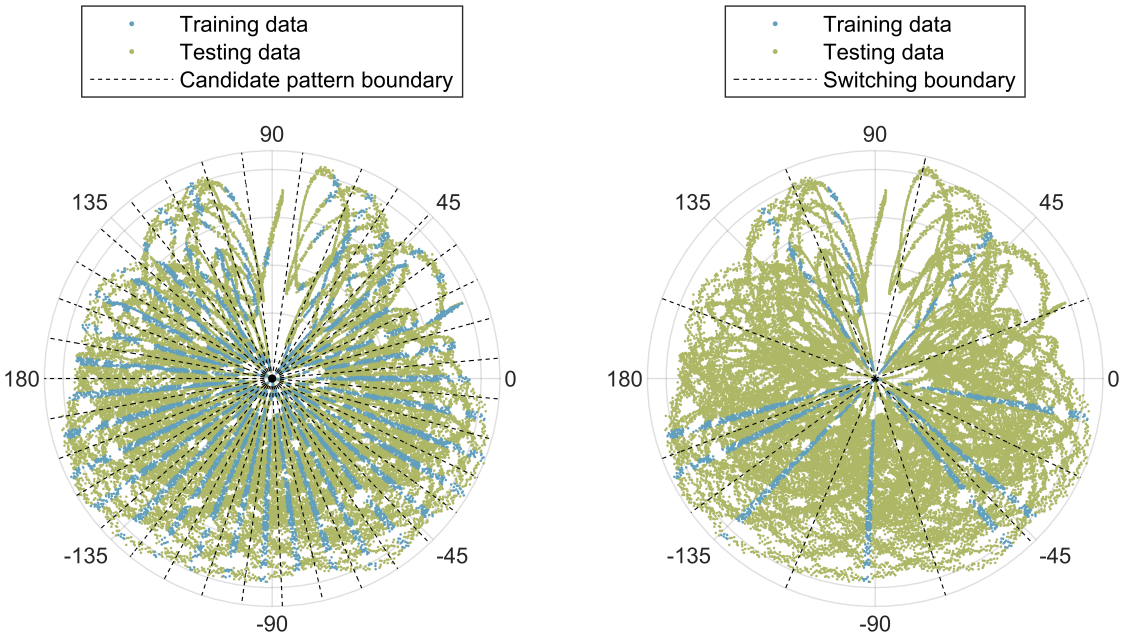


Figure 6.10: The map of squared error distribution across wind directions, obtained from the full switching model.

approach, data are sectioned into even wind direction bins, which serve as estimated wake patterns. Figure 6.11a shows the 34 estimated patterns used in this analysis. In each pattern, data that correspond to one fifth of the wind direction window are selected as a training set.

Amongst these training data sets, the candidate that gives the best MPO prediction for the testing data is chosen one at a time by the greedy algorithm (Algorithm 1), giving rise to a switching method with an increasing number of "switches". The MPO prediction error reduces with an increasing number of selected candidates in a manner shown in Figure 6.12. As more GP-SPARX components are added to the switching model, the MPO performance converges towards an overall NMSE of around 5%. The improvement in MPO prediction becomes negligible after the first ten patterns are chosen. Hence, the minimum number of training patterns needed to represent the wind variations in the simulation data is considered to be ten. The results obtained from the switching model with ten GP-SPARX components are

examined in the following.



(a) All candidate training sets

(b) Top ten training sets selected by the greedy algorithm

Figure 6.11: Training and testing data. The angular and radial coordinates refer to the free-stream wind direction and speed, respectively.

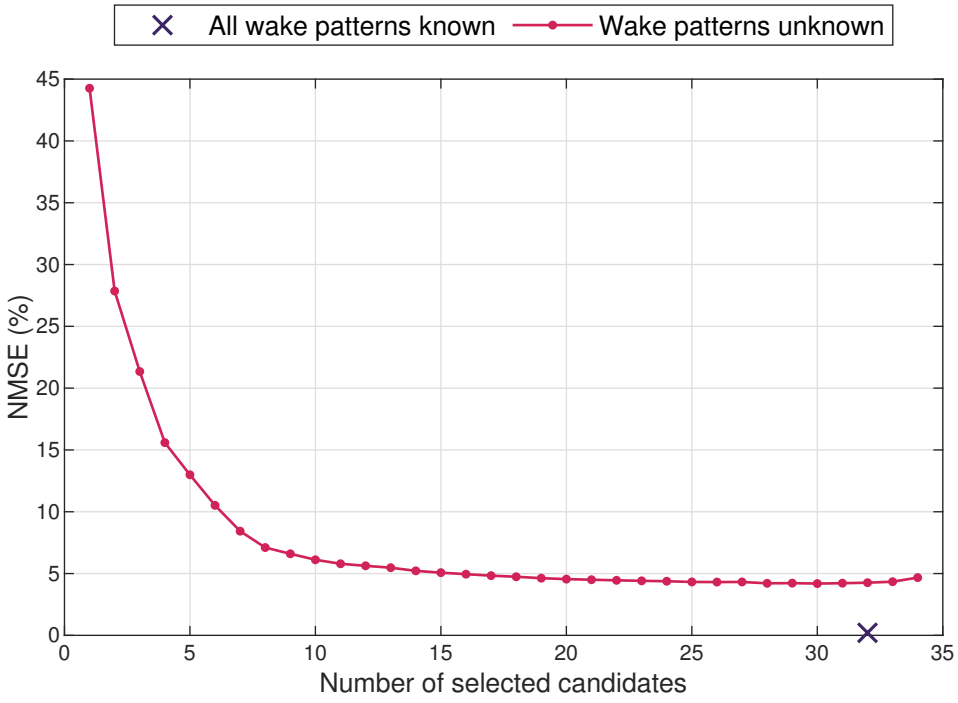


Figure 6.12: NMSE of the MPO predictions for various candidate training sets selected.

The first ten training data sets selected by the greedy algorithm are shown in Figure 6.11b. The training sets are distributed across the entire range of wind directions, with more training data found in the predominant wind directions ($-180° < \theta_\infty < -135°$ and $-45° < \theta_\infty < 0°$). The switching boundaries are set to be the midpoint between the nearest two training wind direction windows, and they refer to the point at which one set of GP-SPARX model is switched to the next. Note that the training regions may not be the midpoint of the neighbouring switching boundaries based on the way the boundaries are defined.

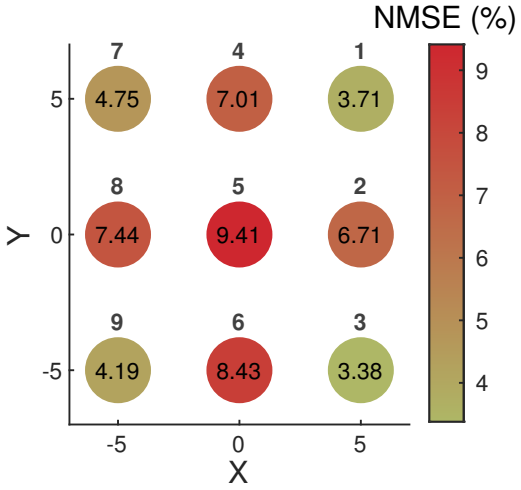


Figure 6.13: The NMSE map for the MPO testing predictions obtained from the optimised switching model.

The NMSE of the MPO predictions obtained from the optimised switching model (Figure 6.13) is considerably higher than the result of the full switching model (Figure 6.9), yet the error values in Figure 6.13 are kept below 10%. Spatially, higher NMSE values are found at locations with more neighbouring turbines, i.e. the centre turbine (turbine 5) has the highest error value, and the turbines at the corners (turbine 1, 3, 7, and 9) are associated with relatively lower NMSE. To understand this, the error distribution across wind directions is inspected in Figure 6.14.

It is seen in Figure 6.14 that the prediction error tends to spike at the switching boundaries. Since a total of 32 wake patterns are approximated with ten estimated patterns, it seems inevitable that the switching boundaries are the most challenging part to accurately predict in. It is also seen that the high error spikes tend to only point in directions from which wakes may come. The more neighbours a turbine has, the wider the wind direction range from which it may be shadowed by a wake, which results in the spatial pattern of NMSE seen in Figure 6.13. Hence, it may

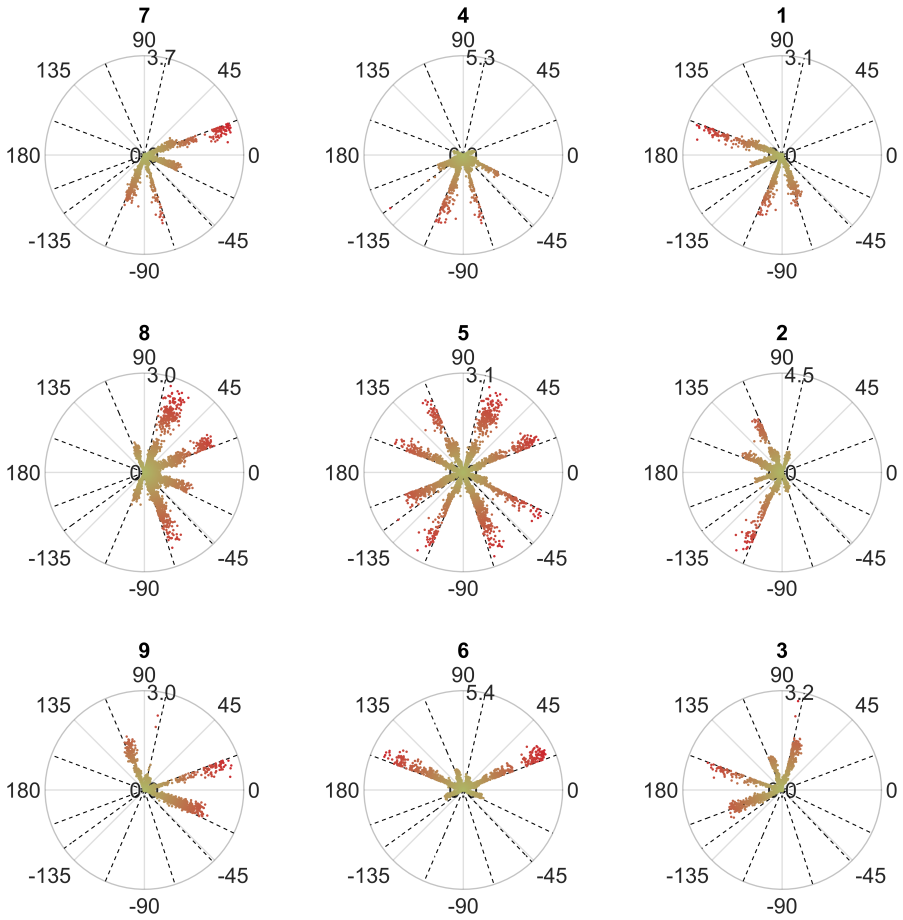


Figure 6.14: The map of squared error distribution across wind directions, obtained from the optimised switching model.

be summarised that the higher prediction errors are more likely a result of the inaccurate approximation of the wake-shadowed turbine behaviour, especially around the boundaries of the switching model.

The error spikes in Figure 6.14 may also be treated as an indication of the positions to add more training data. However, this information is not accounted for in the greedy algorithm used. A different optimisation scheme may be used in an attempt to minimise error spikes at switching boundaries, which will be discussed as part of future work.

The time series predictions made by the full and the optimised switching models are compared in Figure 6.15. The results shown correspond to the turbine location with the highest NMSE in both cases (turbine 5). When all wake patterns are known, the full switching model is able to capture the exact spatiotemporal correlation across the simulated wind farm with high accuracy. When the wake pattern information

is unavailable, the optimised switching model is still able to capture the overall temporal trend but predicts poorly when the turbine turns into or out of a wake.

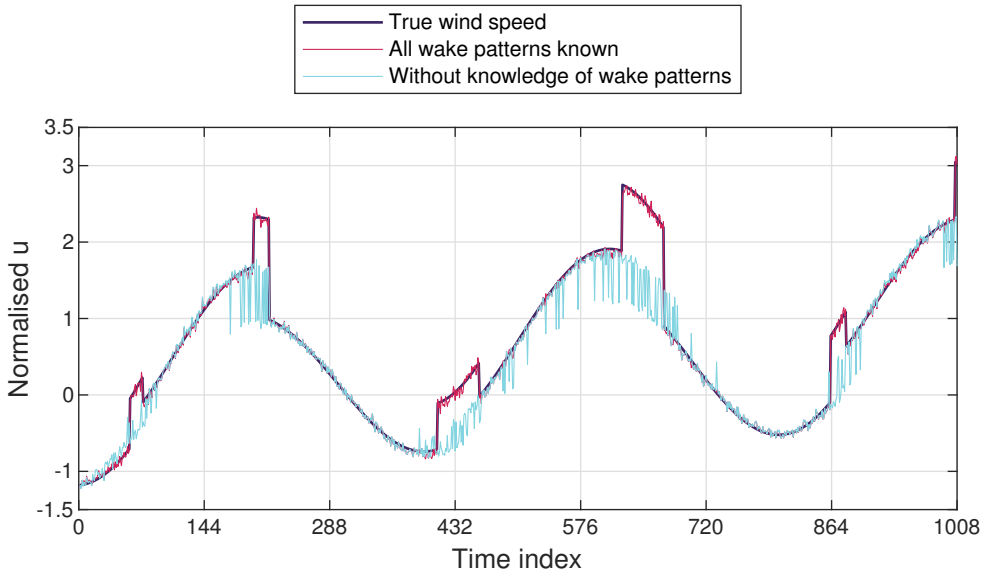


Figure 6.15: The predicted wind speed at turbine 5 over the period of the first week.

6.4 Conclusions

This chapter aims to verify the GP-SPARX model introduced in Chapter 5 with a case study of a simulated wind farm. A simulated data set has been generated from a small, regular wind farm of equally spaced turbines (Section 6.1). The rest of the chapter is devoted to predicting the wind variation across the simulated wind farm using methods based on GP-SPARX. Firstly, a single wake pattern is modelled, and the results obtained from the GP-SPARX are compared with those from other parametric spatial autoregressive models; it is found that the GP-SPARX provides (relatively) the most accurate MPO predictions with minimum input features required. Secondly, time-varying wake patterns are predicted using the switching GP-SPARX method; two switching models are built assuming that the knowledge of wake patterns is (a) known and (b) unknown. The MPO prediction made based on the knowledge of actual wake patterns is understandably better than that obtained from the estimated wake patterns. Nonetheless, in both cases, high errors tend to be found around the switching boundaries, towards wind directions in which wake shadowing is effective.

Chapter 7

Applying GP-SPARX to an Operating Offshore Wind Farm

Chapter 6 has successfully demonstrated the GP-SPARX model on simulated examples. In this chapter, another case study will be provided to demonstrate the model's applicability to real wind farm data and its potential as a low-cost wind farm performance predictor given inputs from weather station measurements.

7.1 Assumptions on GP-SPARX: from Simulation to Reality

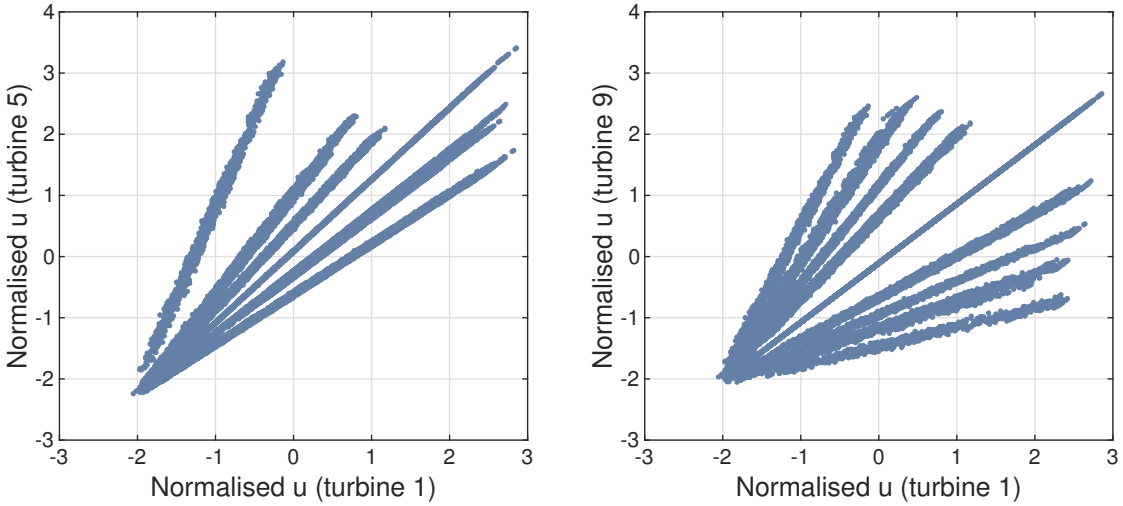
In this section, the assumptions implied through the GP-SPARX model will be discussed. The discussion accounts for the difference and similarities between the simulated and measured wind farm data, to motivate the case study that will be presented later in this chapter as a complement to the one in Chapter 6.

The simulated wind farm described in Chapter 6 provides a simplified representation of the spatiotemporal variation across a real wind farm, with the aim of verifying the ability of the GP-SPARX model to capture known physics. In contrast, data recorded in real wind farms may contain responses to a variety of environmental and/or operational variables, with the causal relationships not fully understood. This can make the measured signals appear more dispersed or "noisy" than the simulated

ones.

In the simulated examples given in Chapter 6, the correlation between turbine locations along any wake path in a wind farm is described using the polynomial NARX model in Equation 6.2. Given that, in the literature of wake modelling, wake propagation has been interpreted as a spatial autoregressive process of various model forms (Section 5.1), the power-form polynomial NARX is chosen as a generalised description of the inter-turbine correlation, for it is a powerful way to model a wide range of systems [84]. The spatial correlation across the simulated wind farm is illustrated in Figure 7.1, where the wind speeds at various turbine locations are plotted against each other. As seen in the figure, the correlation between any pair of turbines appears to comprise many "linear" components. It implies that, in the simulated case, the wind speed at a wake-shadowed turbine is almost equivalent to the wind speed at the wake-producing turbine with a shifted mean; if the wind direction is fixed, the inter-turbine correlations could have appeared linear. The cause of the distinct linear components seen in Figure 7.1 is that all turbines are assumed to face the incoming free-stream wind, which limits the variability of wake patterns. Note that the inter-turbine correlations are nonlinear according to Equation 6.2; they appear linear in Figure 7.1 because the linear terms (i.e. the first two terms in Equation 6.2) dominate the correlations. What complicates the spatial correlation is the time-varying wind directions that gives rise to multiple wake patterns; the more wake patterns exist in a wind farm over a period of time, the more linear components appear on a inter-turbine correlation plot. A comparison between Figure 7.1a and Figure 7.1b shows that the further away two turbines are, the more dissimilar the wake environments the turbines are subject to, making the overall correlation more complicated.

Figure 7.2 compares the wind speeds at various turbine locations in the Lillgrund wind farm. It is seen that, as two turbines become further away, the correlation between them appears less defined, which may be a result of more linear components, as is the case in Figure 7.1. Unlike the simulated data, the real data seem more dispersed with no clear boundary visible between the linear components, the reason being that the turbine angles are free to vary individually in an operating wind farm instead of aligning with the global wind direction as is the simulated case. Other turbulence-induced phenomena such as changing wake shapes and varying degree of wake mixing may also contribute to the data dispersion. Nonetheless, the simplification done in the simulation is a necessary step to help better understand

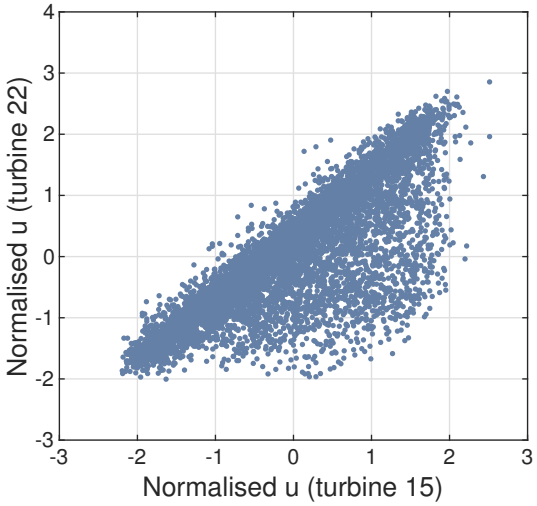


(a) A pair of turbines at a distance of 14.1R (b) A pair of turbines at a distance of 28.2R

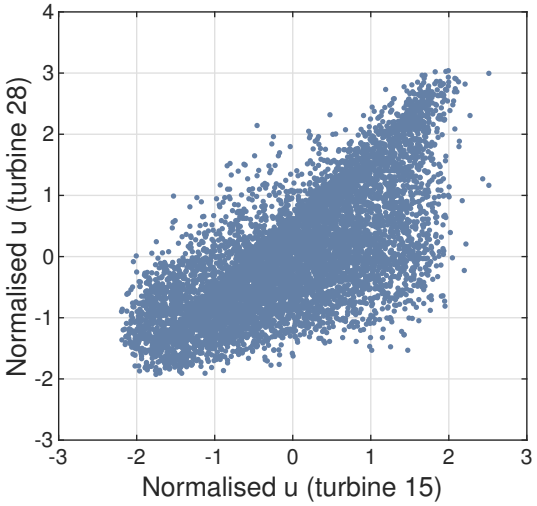
Figure 7.1: Comparing wind speeds at various turbine locations in the simulated farm.

and model the real data.

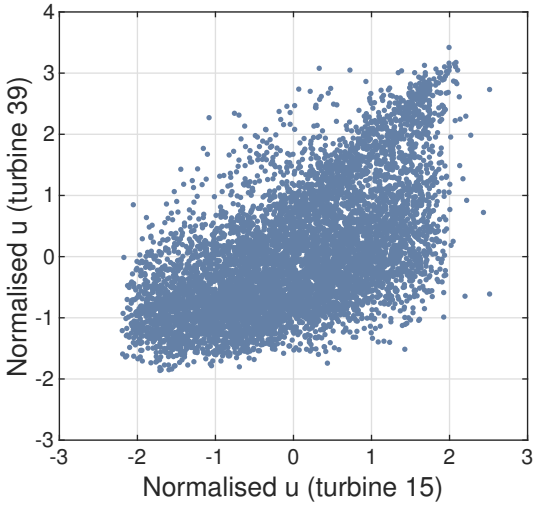
When analysing SHM data, stationarity is often assumed for structural behaviour over long periods of time, whereas nonstationary behaviour is likely to be observed over short-term time windows [30]. To demonstrate wind turbine behaviour in a real wind farm, the wind speed at a turbine in Lillgrund (turbine 15) is illustrated over the period of a year (Figure 7.3a), a week (Figure 7.3b), and 38 hours (Figure 7.3c). Note that the "flat" regions in Figure 7.3 correspond to time periods when (a) data were missing and (b) the wind and power variations exceeded the cut-in and rated range (previously introduced in Figure 4.2). The above assumptions are very relevant to the wind farm monitoring problem, as turbine behaviour is significantly affected by the environmental conditions, primarily wind speed, which appears roughly stationary in the long term (Figure 7.3a) but more and more nonstationary in the short term (Figure 7.3b and Figure 7.3c). This variability in wind is also modelled in the simulated farm but as the sinusoidal variation in wind (Figure 6.2), which demonstrates an simplified way to model the basic properties of wind while circumventing the high cost associated with detailed wind modelling. To this end, the simulation is said to have represented the real wind farm data in terms of the timescales on which (non-)stationarity is determined. However, the actual temporal variation of wind is far more complicated than a linear combination of sinusoids, and the modelling of such a trend ideally requires a nonlinear, nonparametric method.



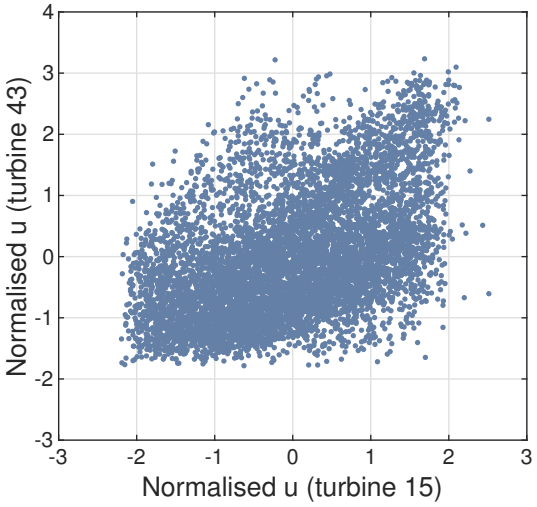
(a) A pair of turbines at a distance of 9.7R



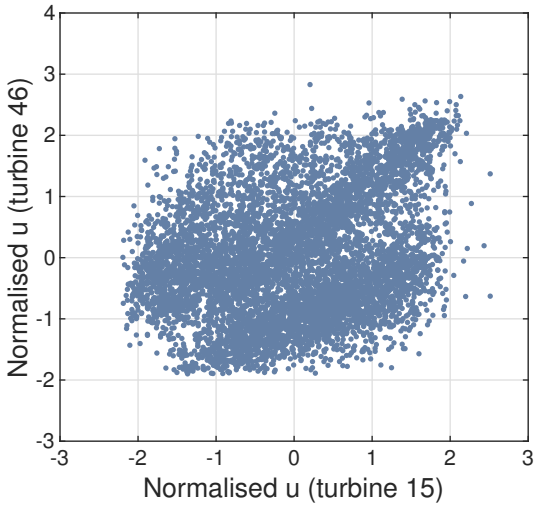
(b) A pair of turbines at a distance of 19.4R



(c) A pair of turbines at a distance of 38.8R

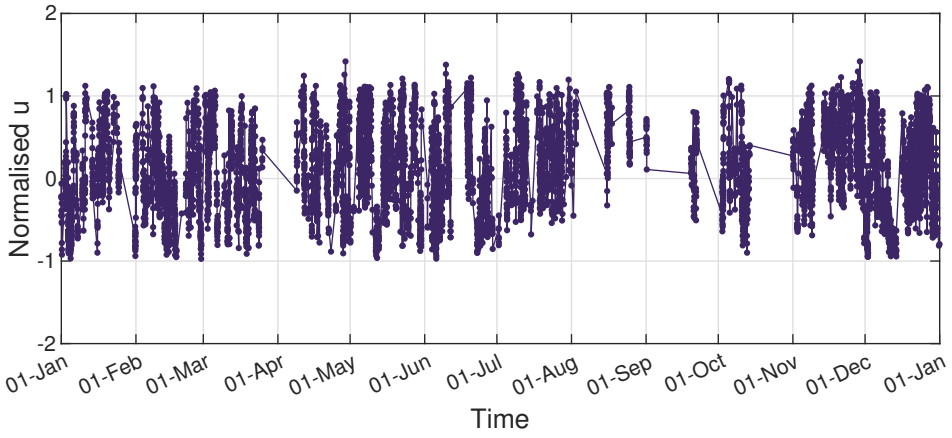


(d) A pair of turbines at a distance of 48.5R

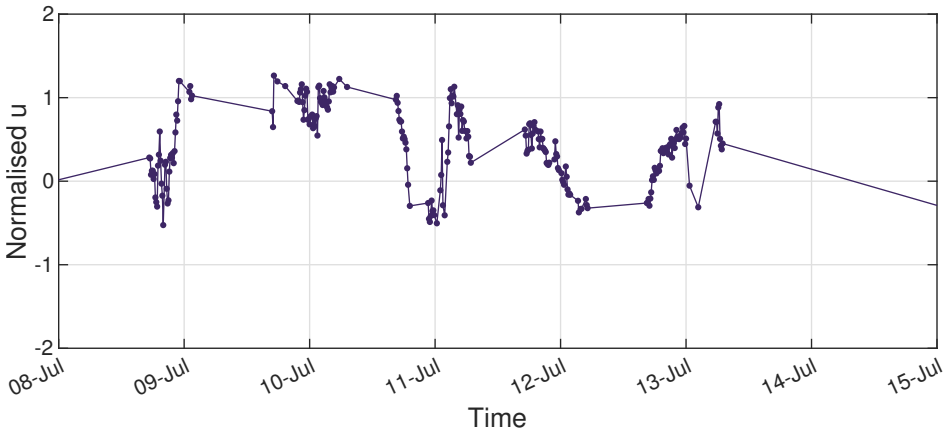


(e) A pair of turbines at a distance of 58.1R

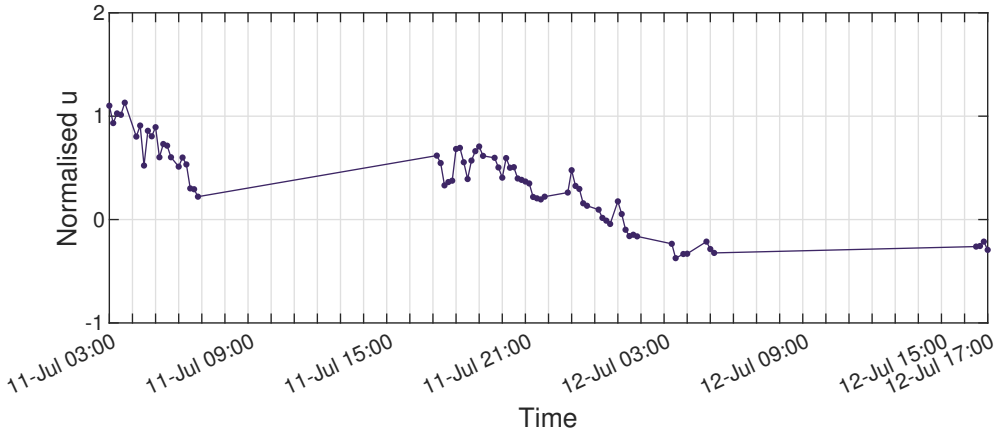
Figure 7.2: Comparing wind speeds at various turbine locations in the Lillgrund wind farm.



(a) Over the period of a year



(b) Over the period of a week



(c) Over the period of 38 hours

Figure 7.3: The measured wind speed at the reference turbine location in Lillgrund (turbine 15).

Since the current GP-SPARX model is based on squared exponential covariance functions (Equation 5.9), stationarity in time is assumed for the input and output of the model [65]. Given the nature of SHM data, the current version of GP-SPARX is best suited for predicting the wind farm variation over a relatively long period of time. With the training data selected to cover all the potential short-term variations, the results in the simulated case have demonstrated the suitability of the stationarity assumption, which is expected to hold for the real wind farm case as well.

Another assumption implied through the choice of the squared exponential kernel is smoothness. However, the use of spatially varying hyperparameters in the GP-SPARX relaxes the smoothness assumptions and allows the model to capture piecewise smooth processes in space. The same piecewice smoothness assumption is used to generate the simulated data, through the choice of the polynomial NARX formulation with spatially varying parameters shown in Equation 6.2; this explains why the GP-SPARX provides accurate predictions for the simulated farm. With regard to real wind farm data, the piecewice smoothness assumption is considered reasonable as wind speeds are expected to vary across a wind farm smoothly except at locations where the wind is interrupted by the existance of operating turbines (and their wakes).

7.2 Modelling Wind Variation Across a Wind Farm

In this section, the switching GP-SPARX method is applied to predict the wind speed variation across an operating wind farm, to demonstrate the predictability of the spatiotemporal variation across an operating wind farm via the use of GP-SPARX.

7.2.1 Model Selection

As introduced in Section 5.4, the switching GP-SPARX method switches between different GP-SPARX models to predict time-varying wake patterns. This section is devoted to the strategies used to create a switching model that is parsimonious and delivers an acceptable level of predictive accuracy. A set of GP-SPARX models that best captures the time-varying wake patterns is selected from a candidate pool of GP-SPARX models, each trained to represent a specific wake pattern. In this case

study, the available wind farm data describe a period of time when the turbine yaw directions are not aligned and (possibly) free to change with local vortices. Given 48 turbines in the farm, the wake patterns change every few time instances. A full switching model becomes infeasible as (a) an large number of GP-SPARX models (in the order of thousands) are required to fully capture the variation in the available data, and (b) the data points in each pattern is insufficient for model training. Therefore, the candidate models are trained based on wake patterns estimated by dividing data into even wind direction windows. The ways data are divided determine the estimated wake patterns and, thus, the constituents in a candidate pool.

In Figure 7.4, the prediction errors obtained from five different candidate pools are compared, to shed light on how the size of a candidate pool may affect prediction accuracy. For candidate pools of size 30 to 60, as more candidates are selected by the greedy algorithm (Algorithm 1) to form the switching model, the predictive NMSE decays in a roughly exponential manner. A larger candidate pool corresponds to a smaller wind direction window for each estimated wake pattern, which means more evenly distributed training data if all candidates are used to form the switching model; thus, improvement in NMSE is seen when more candidates are involved. Across the four candidate pools (i.e. of size 30 to 60), the most significant improvement in NMSE is found up to the point when the first 30 candidates are selected. Although the NMSE continues to improve until about 60 candidates are selected, such improvement comes with the cost of doubled model order. Hence, a switching model with 30 GP-SPARX components appears to be the optimal option that balances computational efficiency with predictive accuracy.

Before deciding an appropriate size for the candidate pool, one has to pay attention to the potential decline in predictive accuracy when the candidate pool exceeds a certain size, as illustrated by the result corresponding to 71 candidates in Figure 7.4. In this instance, the 10^{th} to the 40^{th} selected candidates barely contribute to the overall predictive accuracy of the switching model. A better understanding of the issue is provided through Figure 7.5, where the distribution of training data for the first 30 candidates is illustrated for all five candidate pools. Given the same number of candidates selected (i.e. 30), the larger the candidate pool, the less evenly distributed the training data are across the wind directions. However, what sets Figure 7.5e apart from the rest of the cases is the suboptimal selection of candidates. It seems that the current greedy algorithm cannot arrive at a global optimum until a certain number of candidates are selected. A potential explanation for this problem

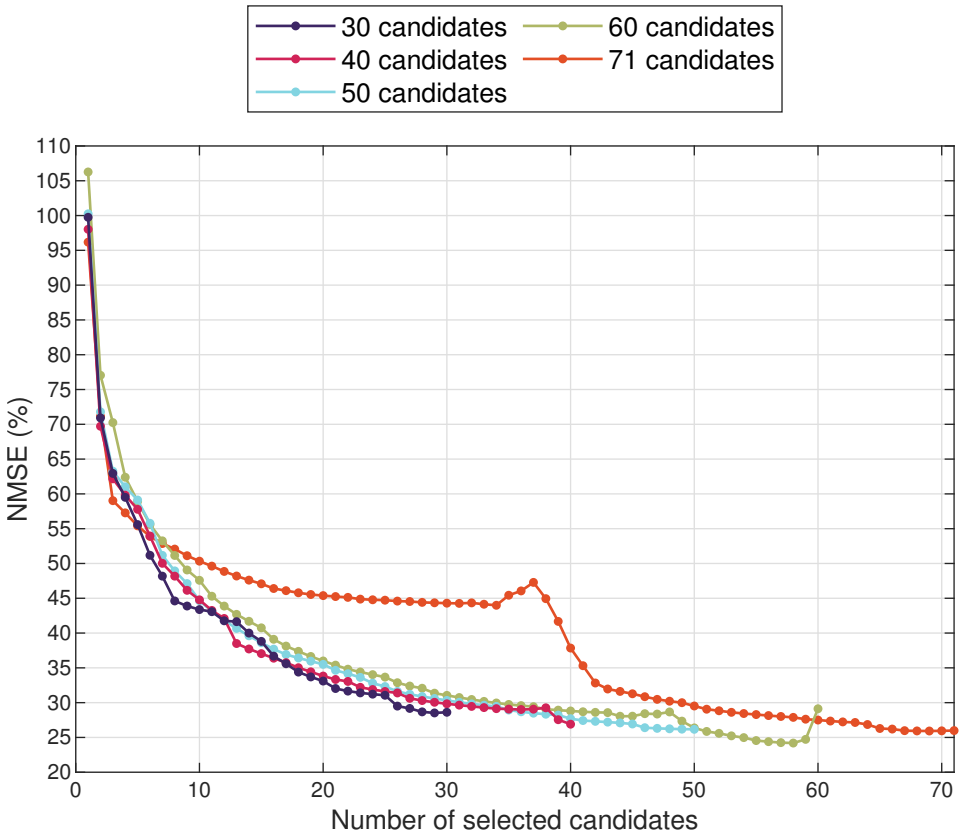
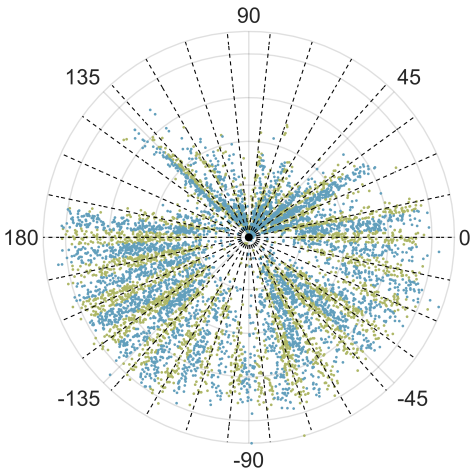


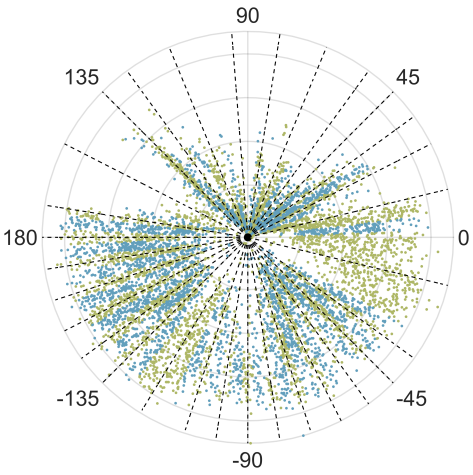
Figure 7.4: NMSE of the validation predictions for various candidate pools.

is that, since a small wind direction window is used to estimate each wake pattern, the amount of data in each window may be insufficient to capture the variation; the greedy algorithm may tend to select multiple patterns in similar wind directions to fully capture the variation over the predominant wind direction (southwest in this case) before moving on to other directions. In the following analysis, controlling the candidate pool size seems to be the most effecient way to avoid suboptimal candidates.

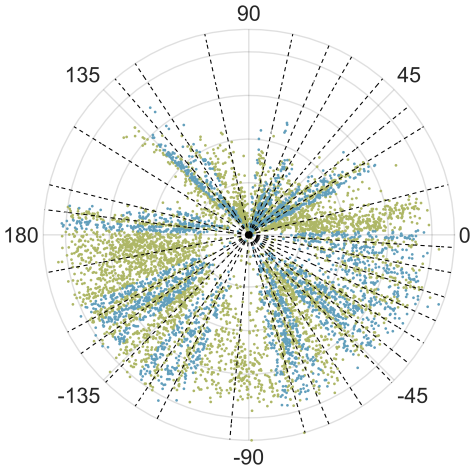
The size of candidate pool used for this study is decided by comparing model performance in terms of the MPO predictions. After the data is divided into even wind direction windows, roughly a half of the data in each window is used as a training set. For a candidate pool size of 30, the training and validation data distribution is illustrated in Figure 7.5a. Given the same number of selected candidates, a larger candidate pool means a smaller amount of training data in total, as shown in Table 7.1. There is no significant rise in prediction errors as the candidate pool size increases, except in the case discussed previously, hence, the tendency here is to select



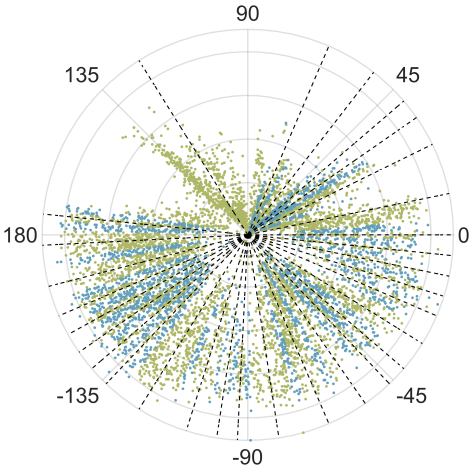
(a) 30 candidates



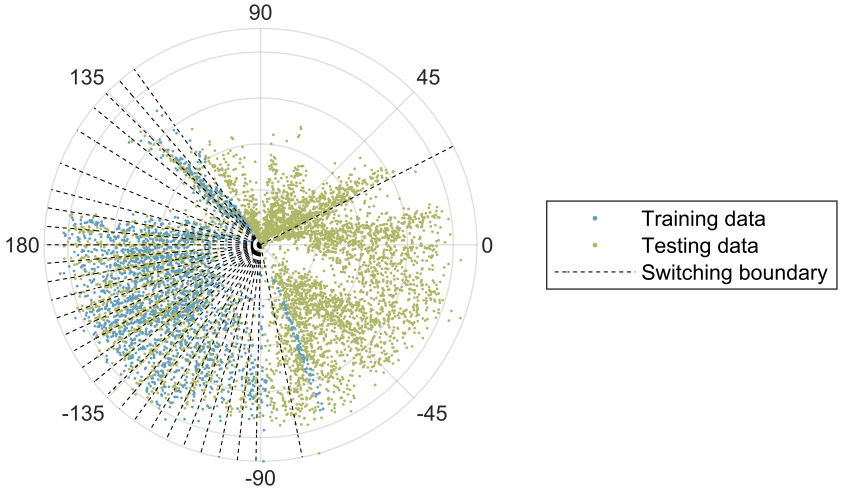
(b) Top 30 selected from 40 candidates



(c) Top 30 selected from 50 candidates



(d) Top 30 selected from 60 candidates



(e) Top 30 selected from 71 candidates

Figure 7.5: The distribution of validation data and the training data corresponding to the first 30 selected candidates. The angular and radial coordinates refer to the free-stream wind direction and speed, respectively.

the candidate pool that provides the least training data usage. A visual inspection of the training data distribution shows that, in the case of 60 candidates, the selected training points provide relatively poor coverage over $90° < \theta_\infty < 180°$ (Figure 7.5d). As a result, a candidate pool of size 50, from which the top 30 candidates are selected to form a switching model, is used for the following analysis; the size and coverage of the relevant training data are considered the most appropriate for this case study.

Size of candidate pool	Data points in the first 30 selected candidates	Equivalent duration (days)	MPO NMSE (%)
30	3798	26.38	28.62
40	2965	20.59	29.84
50	2482	17.24	30.32
60	2260	15.69	31.05
71	1737	12.06	44.30

Table 7.1: The performance of the switching models with 30 GP-SPARX components, given candidate pools of various sizes.

There are, of course, other ways to select training data to ensure even coverage across all wind directions while controlling the training data size. What has been demonstrated above shows that, although uneven coverage of training data may lead to poor prediction accuracy (e.g. Figure 7.5e), even coverage is not the top priority when selecting training data sets (e.g. the error values barely change for Figure 7.5a to Figure 7.5d). The selected candidates also give an idea of how the spatiotemporal variation that occurred during a period of time may be estimated by a set of wake patterns.

7.2.2 Predicting a Wind Farm from a Single Spatial Reference

For the sake of clarity, the results presented and discussed later in this section are obtained from a switching model of 30 GP-SPARX components, with the training data, validation data, and switching boundaries indicated in Figure 7.5c. Each GP-SPARX component models an estimated wake pattern by capturing the spatial autoregressive relationship between turbine locations in the form of Equation 6.3. Since the measurements of free-stream wind speed are difficult to obtain in practice, the wind speed at a reference location, v_0, is used as the exogenous input,

$$\mathbf{u}_i = g_i\left(\mathbf{u}_j, \mathbf{v}_0\right) \tag{7.1}$$

A standard operating wind farm includes several meterological masts for measuring weather conditions, which provide the ideal measurements of reference wind speed. However, such data are unavailable for the current case study. Given that the weather mast in Lillgrund is situated close to the leading edge of the farm for the predominant wind coming from southwest [25], the wind measurements from a turbine on the leading edge (turbine 15) are used as the spatial reference, v_0.

Ideally, the trained model is to be tested on a data set different from the validation data. However, due to the limited availability of data, a unique validation and test set was not possible. Hence, the testing data here refer to the part of validation data not directly used for model training, which contains about 34 days' worth of data. The testing period covers the timescale of a year, excluding, however, the time instances when any turbine is powered off or producing rated power. It ensures that, when extending the model to predict power production from wind measurements (Section 7.3), the focus is on the middle part of the piecewise correlation between power and wind speed (Figure 4.2). The same testing data is used here and later in this chapter such that the prediction results may be interpreted in conjunction.

7.2.3 Results

The testing results obtained from the described switching model are demonstrated in the NMSE map shown in Figure 7.6. Across the wind farm, a spatial difference in errors can be seen as a combination of two trends: (a) prediction errors increase with distance away from the reference turbine 15; (b) higher errors are found in more frequently wake-shadowed positions, i.e. towards the middle of the farm. Together, the two trends indicate a spatial change in predictability.

Since the wind speed at each turbine location is predicted from its upstream speed (the autoregressive term) and the reference wind (the exogenous term), the correlation between the output and each of the input terms is explored. Firstly, the correlation between the exogenous input and the output is spatially varying, as illustrated in Figure 7.2, where the correlation appears weaker with increasing distance between the output and the reference locations. Therefore, it is more difficult to predict the wind speeds far away using the same reference input, resulting in errors that increase with the distance from the reference.

Secondly, predictions at frequently wake-shadowed positions are likely to be less

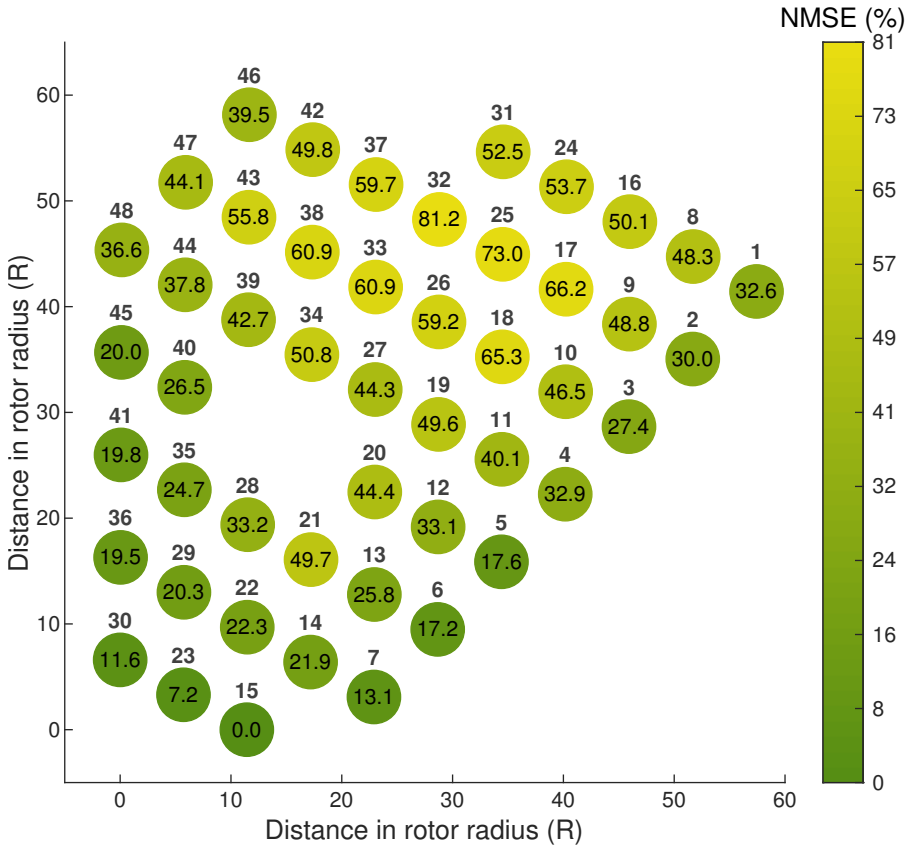


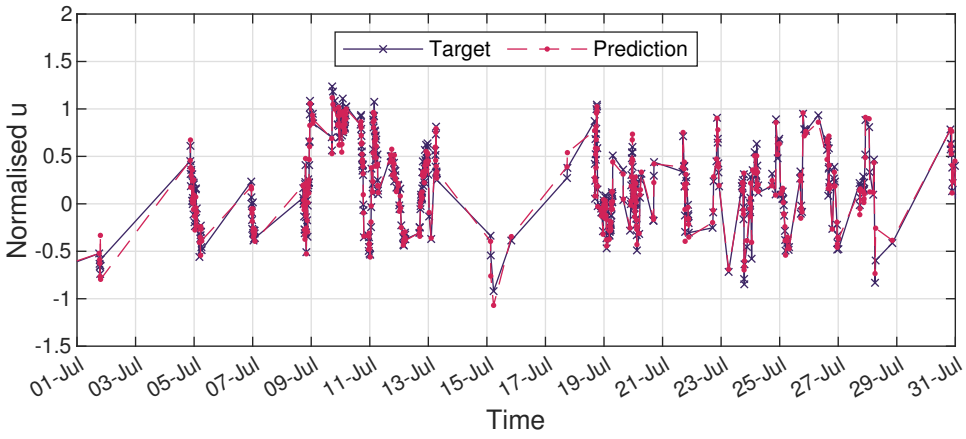
Figure 7.6: The NMSE map for the testing predictions obtained from the optimised switching model.

accurate since the error accumulates through the MPO prediction process, i.e. the GP predicted mean at one location is used as an input to make predictions at the next location. As discussed in Section 5.3, the results in this work are obtained without propagating the predictive variance, thus, a reduction in prediction accuracy is inevitable as more MPO predictions are made.

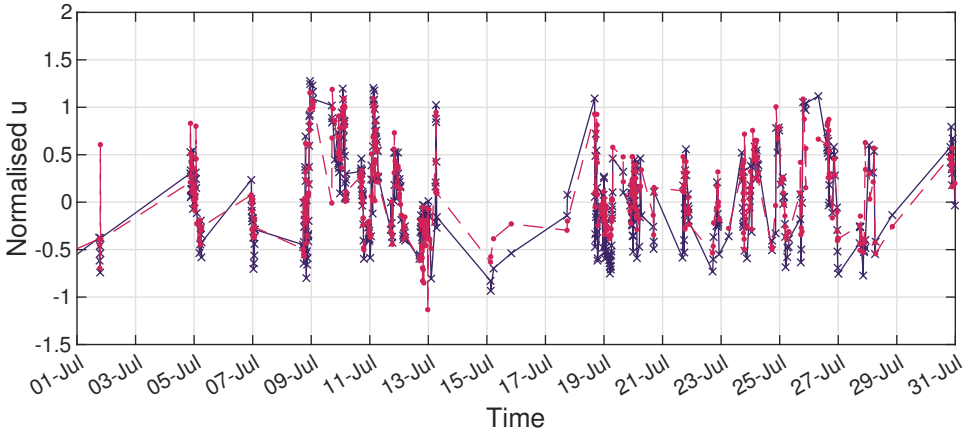
The potential remedies for the spatial difference in error include (a) the selection of a reference location close to the middle of the wind farm or (b) the implementation of an uncertainty propagation method, e.g. Monte-Carlo sampling [72]. That said, the current version of the model predicts with individual NMSE values capped at 81.2% and an overall NMSE of 38.92%, which is encouraging given that only one spatial reference is used and the reference position is not optimised.

The time series predictions made by the switching GP-SPARX method are presented in Figure 7.7. Examples are given at locations with low, medium, and high levels of NMSE, to allow visual assessment and comparison of the model performance. Note

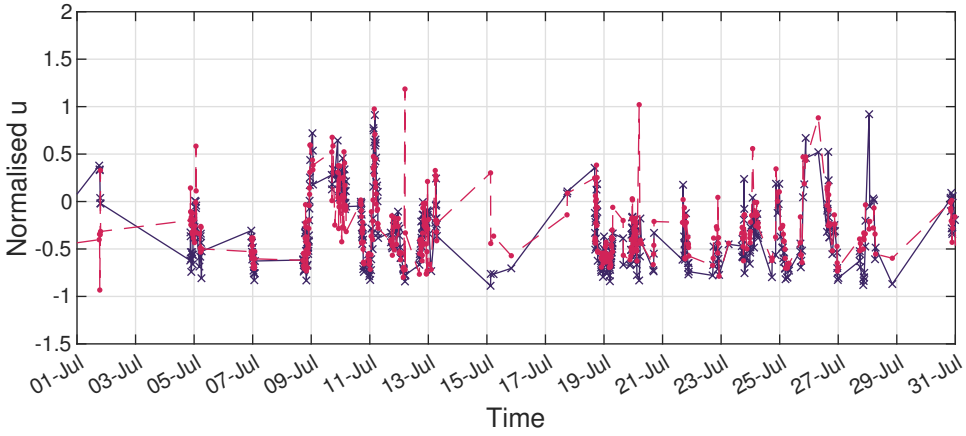
that the time series shown here provide a close-up view of the entire testing data set and only cover the wind variation over a month.



(a) With a relatively low level of NMSE (Turbine 23)



(b) With a medium level of NMSE (Turbine 39)



(c) With the highest NMSE (Turbine 32)

Figure 7.7: The GP predicted mean wind speed over the period of a month.

In the low NMSE case (Figure 7.7a), the mean prediction follows the target trend

closely, with occasional small-scale mispredictions that can be comfortably considered insignificant. Moving on to the second case (Figure 7.7b), the general trend of the target is said to be captured by the model predicted mean, but with mispredictions occurring more frequently. These mispredictions are more notable than those seen in the first case, such as the ones occurring around 15-07. There is no consistent under- or overpredictions that exceed the timescale of a few days. Some undesirable fluctuations are also seen in the prediction.

In the high error case (Figure 7.7c), worse fluctuations are observed. It might be deduced that the undesirable fluctuations are associated with the poor prediction at the switching boundaries of the GP-SPARX models. As well as the fluctuations, discrepancies between target and prediction become more notable in Figure 7.7c, with a typical example seen during the period from 13-07 to 16-07. The positive side is that the overall trend of the time series prediction shown in Figure 7.7c — corresponding to the least accurate prediction across the wind farm — follows roughly the target trend, which implies that the proposed method is effectively capturing the spatiotemporal variation.

7.3 Modelling Turbine Performance across a Wind Farm

As the predictability of wind variation via the switching GP-SPARX method has been demonstrated in Section 7.2, the same method is applied to model the variation of turbine performance across an operating wind farm, in this case, the power produced by individual turbines. An accurate prediction of the power across a wind farm can be useful for (a) performance monitoring and (b) power forecast. The former is in line with the purpose of the mapping method introduced in Chapter 4; a wind farm performance indicator can be developed based on the predicted normal behaviour. The latter provides a forecast in farm-wise power production that may contribute to turbine control decisions, for example, to optimise the overall power production.

Given the strong correlation between power and wind speed [54], the turbine power outputs across a wind farm are also expected to vary in a spatial autoregressive manner. A slight modification is made to the GP-SPARX formulation in Equation 7.1

for adaptation to the new target variable,

$$\mathbf{p}_i = g_i\left(\mathbf{p}_j, \mathbf{v}_0\right) \tag{7.2}$$

where $\mathbf{p}$ indicates turbine power output. It is to be clarified that the the GP-SPARX model used in Section 7.2 focuses on capturing the spatiotemporal variation in wind speed across a wind farm, whereas the model used in this section aims to capture also the wind-power correlation — in Equation 7.2, turbine power is predicted given free-stream wind speed as the only exogenous input. Hence, a successful model based on Equation 7.2 can potentially be used for wind farm power forecast given the wind forecast from weather station(s).

Following the discussion in Section 7.2.1, a candidate pool size of 50 is chosen, from which the GP-SPARX model that provides the relative best MPO prediction is picked out iteratively using a greedy algorithm. As is the case of wind prediction, the switching model used here contains 30 GP-SPARX components to allow comparisons between the results. The prediction errors change with the size of the switching model in a manner shown in Figure 7.8, where comparisons are made between the wind and the power prediction cases. Both cases demonstrate (almost) exponential decay in NMSE, with similar levels of error. It indicates that the spatiotemporal variation in power can also be effectively modelled as a set of "wake" patterns, and that, within the selected operating range, turbine performance is highly dependent on the environmental condition. Given that a static wind-power model has been the mainstream method used by wind farm operators, this thesis points out an alternative direction — instead of obtaining a power field forecast from a wind field forecast, it is also possible to obtain the power prediction across a wind farm from the wind forecast at a single point.

The map of testing errors for power prediction is illustrated in Figure 7.9, in which the spatial distribution of errors is similar to the case of wind prediction. It is once again demonstrated that, within an operating range, the power production across a wind farm is governed by a spatial autoregressive process. Comparing to the wind prediction case, higher errors are seen in similar regions, i.e. towards the frequently wake-shadowed turbines far away from the spatial reference; the turbine with the highest error in Figure 7.9 (i.e. turbine 17) is different from that in Figure 7.6 (i.e. turbine 32). As discussed in Section 7.2.3, the spatial trend in error is likely a result of (a) the spatially varying correlation between the exogenous input and the

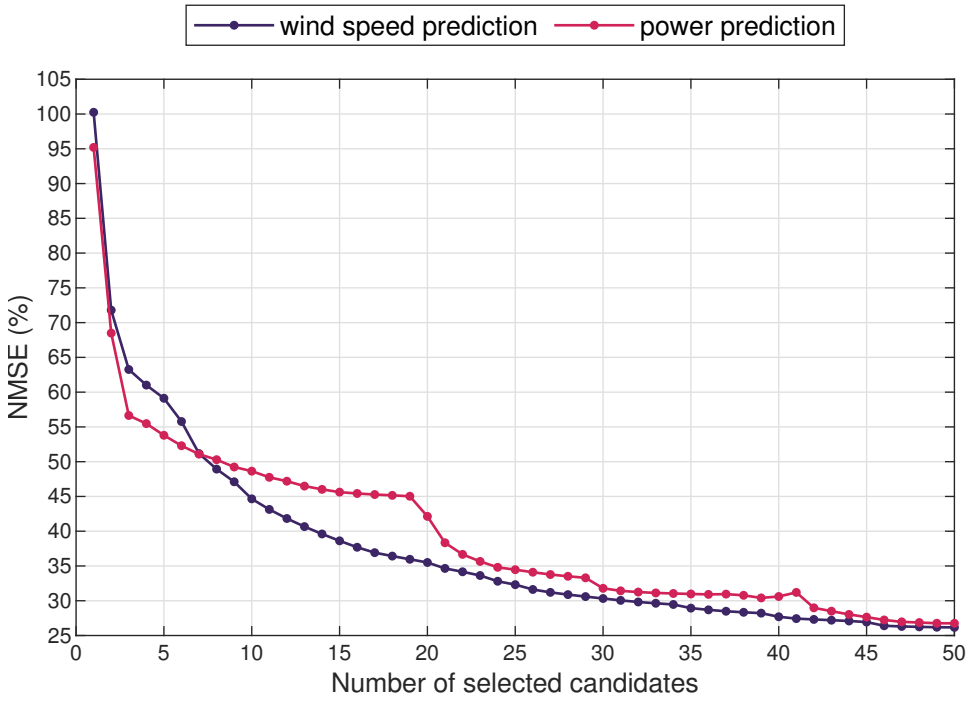


Figure 7.8: NMSE of the validation predictions for different output variables.

output and (b) the MPO prediction process. In Figure 7.9, the individual NMSE has an upper limit of 78.1%, with an overall NMSE of 44.05%; the error level is similar to that in Figure 7.6.

The times series of the predicted power is illustrated in Figure 7.10, which includes examples corresponding to low, medium, and high errors. In all three examples, the general trends of power variation are captured in the mean GP predictions. As is the case of wind prediction, higher values of NMSE are likely to result from (a) more frequent, larger discrepancies between target and prediction and (b) more fluctuated predictions close to the switching boundaries. It is reasonable to conclude that, under specific operating conditions, the power production across a wind farm can be modelled as a spatial autoregressive process using the proposed switching GP-SPARX method.

7.4 Conclusions

To further establish the GP-SPARX model, this chapter has demonstrated the capability of the switching GP-SPARX method to capture the spatiotemporal variation

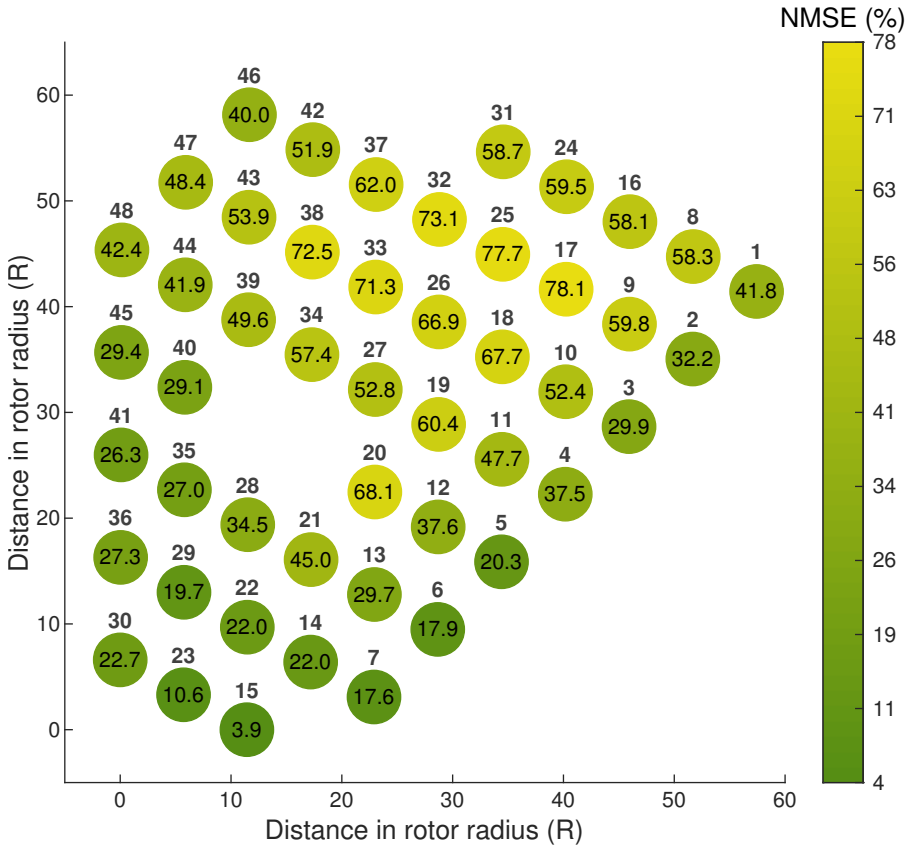
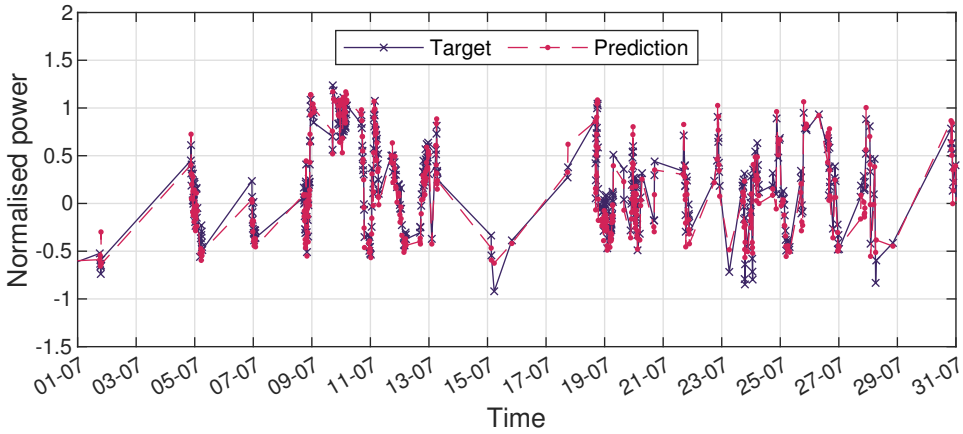
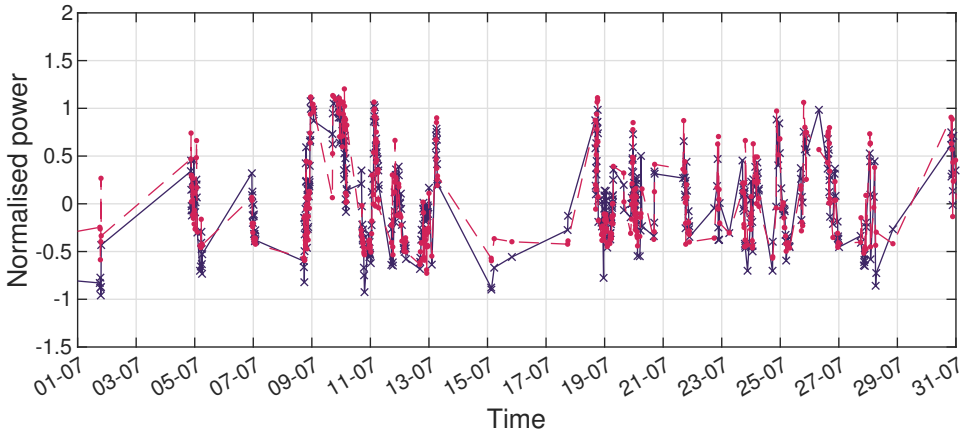


Figure 7.9: The NMSE map for the testing predictions obtained from the optimised switching model.

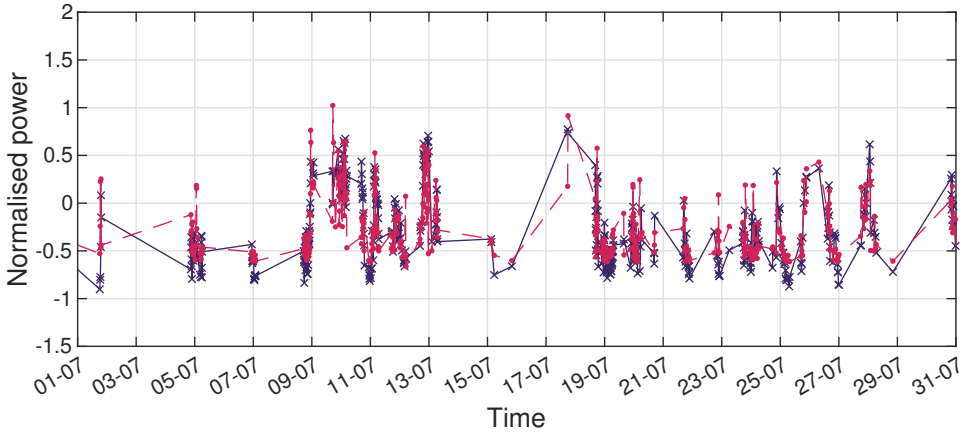
across an operating wind farm, Lillgrund. The chapter began with a discussion on the assumptions associated with the use of GP-SPARX, in light of the similarities and difference between simulated and real wind farm data (Section 7.1); it has been explained that the case studies in Chapter 6 and Chapter 7 are complementary and together provide a comprehensive assessment for the GP-SPARX model. The application of GP-SPARX to real wind farm data has been demonstrated in the rest of the chapter. Firstly, the switching GP-SPARX method was used to predict the time-varying wake patterns (i.e. wind variation) across the wind farm from a single spatial reference (Section 7.2), with the model selection process discussed in Section 7.2.1. Secondly, an equivalent model was applied to predict the power variation across the wind farm over the same period of time (Section 7.3). Results show that the switching GP-SPARX method can predict the spatiotemporal variation in wind and power (within a certain operating range) across a real wind farm, given the wind speed measurements from a single spatial reference. An overall summary of GP-SPARX across Chapter 5 to Chapter 7 is provided in the following.



(a) With a relatively low level of NMSE (Turbine 23)



(b) With a medium level of NMSE (Turbine 44)



(c) With the highest NMSE (Turbine 17)

Figure 7.10: The GP predicted mean power over the period of a month.

7.4.1 A Summary of GP-SPARX

As its name implies, the GP-SPARX describes the autoregression of a target variable across space. Like all models of the autoregressive type, the GP-SPARX does not require the information of every driving factor (i.e. explanatory variable) to describe a spatial process; this may imply less effort required into analysing the causal relationships in data. Inherent in GP-based NARX models is the problem of lag selection, which is usually done through exhaustive searches [72]. In the case of GP-SPARX, the lag selection problem is circumvented by a physics-inspired model structure, i.e. the use of one spatial lag that corresponds to the upstream neighbour(s), which significantly simplifies model structure and implementation. Another benefit of the autoregressive formulation lies in that the current version GP-SPARX model — spatially autoregressive with static temporal mapping — can be extended to incorporate temporal lags (to handle dynamic data) with relatively simple modifications.

In this work, the GP-SPARX model is primarily developed for wind farm modelling, where the physical understanding of the problem contributes to the refinement of model structure and implementation. However, this physics-inspired method might not fall into the category of what is conventionally understood as a grey-box (or physics-informed) approach [62, 75, 85, 86], since the model structure, in essence, is still a specific type of GP-NARX model (black-box). What is of importance here is the use of physical knowledge to form tailored solutions to specific engineering applications. In an era of machine learning "explosion", it has almost become a common practice to fit a generalised algorithm to a large data set in an attempt to get round the time-consuming, case-specific process of data analysis, which might be an efficient or even the only possible approach in certain circumstances. However, the author intends to remind one of the beauty of engineering problem solving. When combined with machine learning techniques, the physical understanding developed through numerous previous studies is likely to bring about parsimonious solutions which may save large amounts of computational power in the long run.

The switching GP-SPARX method is a form of hierarchical model used to address the two-level problem of wind farm modelling. The underlying philosophy is to divide an engineering problem into smaller, simpler-to-solve elements. Here, individual wake patterns are modelled by separate GP-SPARX models and a switching model is laid on top to capture the variation in wake patterns. In comparison to the single-GP

approach demonstrated by the mapping method in Chapter 4, the switching GP-SPARX method is considered to have splitted one GP into multiple GPs. Given that exact GP regression is notorious for its high cost in dealing with large data sets, a straightforward way to prevent this high cost is to use a smaller training data set. In the case of GP-SPARX, the high computational cost is circumvented by training multiple GPs, each with a relatively small data set, such that the total amount of training data allowed by the GP-SPARX approach is larger than that by a single GP. Hence, the multiple-model approach is able to learn and predict the data variation over a longer timescale, which makes the two approaches fit for different monitoring purposes.

Chapter 8

Conclusions and Future Directions

This thesis has attempted to develop new methods for anomaly detection in a PBSHM setting by accounting for the confounding influence of spatiotemporally changing environmental and operational conditions. As an alternative to the existing methods for PBSHM, the new spatiotemporal approach (a) recognises and captures the spatiotemporal correlations between structures across a population, (b) makes full use of the available EOV data, and (c) considers the structures collectively at the population scale. While other PBSHM methods seek to transfer knowledge between structures, the methods developed in this thesis focus on the correlations between structures, but, together, these methods serve the same purpose of sharing data/knowledge across a population of structures. An offshore wind farm is an example of a population of structures with spatiotemporally correlated EOV and, hence, is a main application of the spatiotemporal approach.

Chapter 2 reviewed the existing sensing techniques and modelling methods for the SHM of wind turbines and farms, from which three conclusions has been drawn. Firstly, more research is needed on methods of population-based turbine assessment. Secondly, the spatiotemporal interaction between turbines is one of the most important considerations during the development of population-based monitoring methods for offshore wind farms. Lastly, the existing models that account for spatiotemporal variations across a wind farm are mainly physics-based and not for monitoring purposes, whereas capturing spatiotemporal variations is not the focus of the existing

data-based methods for wind farm monitoring.

In light of the lack of research on a spatiotemporal approach to the population-based monitoring of wind farms, Chapter 3 reviewed the existing spatiotemporal models in other fields, which has been divided into three categories: (a) physics-based, (b) physics-informed, and (c) data-based. A current trend in SHM is to incorporate physics knowledge to enhance the accuracy of data-based monitoring methods, especially for extrapolation; the spatiotemporality of a process is a type of knowledge that may contribute to the formation of physics-informed monitoring methods. In the scope of this thesis, the developed methods are mainly at the data-based end of the spectrum, with the spatiotemporal correlation between turbines reviewed to shed light on how data-based models can be created.

In Chapter 4, the mapping method has been developed as a wind farm performance indicator that accounts for the spatiotemporal variation across a wind farm. The mapping method includes a modelling step, in which a data-based model is trained on the normal condition of the wind farm, and a detection step, where model predictions are compared with measurements to assess the performance condition of turbines. In the modelling step, the wind farm-wise spatiotemporal correlation is captured in terms of a spatiotemporal covariance function with a generic prior; this corresponds to a purely data-based model. Before exploring more complex (i.e. physics-inspired) methods for spatiotemporal modelling, an investigation into the ability of a data-based model in capturing the spatiotemporal wake effect is necessary. Case studies of an operating offshore wind farm has also been provided, which demonstrated the capability of the data-based mapping method as a wind farm performance indicator.

Chapter 5 introduced GP-SPARX as a physics-inspired model designed specifically for capturing the spatiotemporal variation across a wind farm. The GP-SPARX model assumes the farm-wise variation as a spatial autoregressive process; this autoregressive relationship is specified through wake paths, and the wake paths resulting from a specific wind direction constitute a wake pattern. That is, a GP-SPARX model aims to capture a wake pattern. In the case of modelling time-varying wake patterns, the switching GP-SPARX method is suggested, which switches between multiple GP-SPARX models to predict the time-varying wake patterns.

The main aim of Chapter 6 is to verify that the GP-SPARX introduced in Chapter 5 is suitable for modelling the spatiotemporal variation across a wind farm; to this end, a simulated data set has been considered such that the prediction results are a

representation of the modelling capability instead of an indication of the performance anomalies that might be present in operational data. A trained GP-SPARX model was able to successfully capture the spatiotemporal correlation depicted in a wake pattern. For the time-varying wake patterns stemming from the changing wind direction, the switching method performed better with the knowledge of wake patterns known. Without such knowledge, a switching model based on estimated wake patterns was created; the predictive accuracy of the model converged towards an acceptable level. Although incorporating the knowledge of wake patterns improved the accuracy of the switching GP-SPARX method, regions of (relatively) high errors were the same with or without such knowledge; for each turbine, high errors were found in wind directions where the turbine might be shadowed by wakes, especially around the switching boundaries of the model.

In Chapter 7, the GP-SPARX was applied to predict the spatiotemporal variation across an operating wind farm; in particular, a switching model was trained using a chosen set of training data to predict the time-varying wake patterns in wind speed and power, respectively. It has been demonstrated that the switching method is potentially capable of predicting across a farm based on the measurements from a single spatial reference. A main significance of the GP-SPARX developed in this work is the fact that it is a physics-inspired model; the physics knowledge provides guidance on model structure, including the selection of exogenous input and the lag length of autoregressive terms, which leads to a more parsimonious model. The GP-SPARX approach also demonstrates how a complex modelling problem, i.e. the modelling of the spatiotemporal variation across a wind farm, can be addressed by dividing it into simpler problems, i.e. the modelling of a wake pattern. Comparing with the mapping method developed in Chapter 4, the GP-SPARX approach is more capable of monitoring wind farm performance over long timescales.

8.1 Limitations and Future Work

8.1.1 SHM for Wind Farms Using SCADA Data

In the literature of SHM for wind turbines, SCADA-based methods have proven useful in providing early warnings of damage in turbines [42]. This work follows the footsteps of previous studies to develop a new approach to wind farm monitoring —

a spatiotemporal approach — that fits in the field of PBSHM. The SCADA system that is available in (nearly) every turbine aims to provide consistent collection and recording of a wide range of quantities, describing environmental and operational conditions as well as turbine performance; for each quantity, a statistical summary (including mean, maximum, minimum, and standard deviation) is available every ten minutes. In the scope of this thesis, only the mean values in SCADA data are analysed and the spatiotemporal correlation explored resides only in this mean variation.

Given the richness and broadness of SCADA data, there are a number of avenues for future research concerning SHM for wind farms. Firstly, the ever broadening variety of measured quantities provide ample opportunity to identify the cause of damage (i.e. identifying damage-sensitive features); methods based on neural networks may be useful here. Secondly, the statistical summaries associated with each quantity may be incorporated to enhance the monitoring method. One way to utilise the statistical summaries could be to propagate the values of standard deviation, as input uncertainty, through the model that predicts normal turbine behaviour. Methods that approximate the propagated uncertainty through GP-based models have been investigated previously in [87, 88]. Another possibility that may prove insightful is to model the maximum/minimum values as a means of monitoring and predicting potential extreme environmental or operational conditions. Thirdly, as the SCADA system continues to collect data, there will be a growing amount of SCADA data available; methods for constructing scalable models and/or dimensionality reduction may be worth investigating into.

8.1.2 Modelling Techniques for the Mapping Method

The mapping method, in essence, is an anomaly detector for the population of turbines in a wind farm, with separate steps for normal behaviour modelling and anomaly detection. In this section, potential improvements will be discussed mainly in terms of modelling techniques.

With regard to modelling techniques, two approaches have been developed in this work — GP regression (Chapter 4) and GP-SPARX (Chapter 5 to Chapter 7). The following text will focus on GP regression, and a discussion on the future directions of GP-SPARX will be found in Section 8.1.3. As a model for the normal behaviour

across a wind farm, GP regression has two practical limitations. Firstly, the exact GP is computationally expensive for large data sets, because of its complexity of $\mathcal{O}(N^3)$ in computation and $\mathcal{O}(N^2)$ in storage. One way to apply the exact GP regression without modifying the model for scalability is simply to control the size of data set used; this means that the GP model is to be trained on data that correspond to a shorter timescale, given a wind farm with a larger number of turbines. The second limitation of the GP-based model for the mapping method lies in that the wind speed measurements from several turbines are required as inputs to the model to allow interpolation across the geographical region covered by the wind farm.

It is encouraging that both limitations have been addressed by the GP-SPARX approach. The GP-SPARX model focusses on capturing the correlation between neighbouring turbines in a spatial autoregressive manner, instead of modelling the farm-wise spatiotemporal correlation in one go as is the case with GP regression. As a result, the GP-SPARX approach is able to monitor wind farm behaviour over longer timescales regardless of the size of the farm, and the prediction of wind farm behaviour can be made using a single spatial reference. A more detailed discussion is found in Section 7.4.1.

Future research on modelling techniques for the mapping method may be guided by two basic ideas — (a) to incorporate physcics into the model and (b) to allow online monitoring. Despite that physics-informed machine learning is a recent research trend, the incorporation of physics knowledge in data-based SHM approaches brings about a variety of benefits, one of which is the capability to extrapolate. This, for example, may allow the dynamic behaviour of a structure to be predicted ahead in time or at off-site locations. For the GP regression model in the mapping method, physics may be incorporated through the use of a spatiotemporal covariance function that describes the wake propagation process; physics-informed covariance functions have been developed for SHM problems in [62, 89].

Online machine learning methods, on the other hand, aim to address learning environments that evolve in time. In a SHM context, online methods may be motivated by the aim of having a continuous monitoring system, with the prediction models updated to accommodate changes in data that, for example, result from EOV. Such an online framework for SHM has been previously looked at in [90]. This idea requires further investigation.

It is noted that the wind farm prediction model in the mapping method may be used

for purposes other than performance indication/monitoring. Having an accurate prediction of the power production across a wind farm is the first step in power optimisation. The GP regression model, as well as the GP-SPARX model, may be adapted to provide information on how the individual turbines are to be controlled to ensure an overall maximised power production. Power prediction for wind farm optimisation has been attempted using physics-based wake model in [91]. Further investigation is required to develop this idea.

8.1.3 GP-SPARX

At the risk of repetition, as a physics-inspired modelling technique that is able to be used in the framework of the mapping method, the GP-SPARX benefits from its ability (a) to monitor a population of turbines over long timescales and (b) to predict across a wind farm using inputs from a single spatial reference.

The current version of GP-SPARX does not account for the propagation of model uncertainty in the MPO prediction process, as the prediction variance is assumed negligible across all positions in a wind farm; justifications for this assumption have been attempted in Section 5.3.1. An interpretation with regard to the physics of wind farms would be that the model uncertainty accumulated through the MPO prediction process may be a representation of the increasing turbulence level as the wind progresses downstream across a wind farm. Hence, the propagated model uncertainty may be an indication of the turbulence level around each turbine. To better predict this turbulence level, information on the standard deviation of the raw data streams, which is also available in the SCADA data, may be used in addition to the mean values, developing methods along the lines of [87, 88]. A GP-SPARX model that predicts turbulence levels as well as the mean wind would be a desirable progression, which shall be the focus of future work.

Currently, the GP-SPARX provides MPO predictions based on a spatial reference; without the data from the weather station, the wind speed measurements from a nearby turbine has been used as the reference. As a result, the reference input does not represent free-stream wind conditions. To address this, simulated data from, for example, climate models may be of use. Altenatively, measurements from a weather station that is not affected by turbine wakes may also prove useful.

To handle the time-varying wind directions, the current method utilises a “hard”

switch between the GP-SPARX components, resulting in relatively poor predictions along the switching boundaries. A promising approach to improve the predictions at the boundaries is to develop a "soft" switch, based on methods such as mixture of experts, so that the overall prediction can be derived as a weighted sum of the individual GP-SPARX predictions [92]; alternatively, a method based on treed GP may also be considered [93, 94]. Another approach is to treat the wind direction as a continuous variable, so that the discrete switch between the GP-SPARX components can be replaced by a continuous transition. Future work has been planned to investigate the deep learning option of GP-SPARX using methods such as [88].

Determine wake patterns from data

```
function [count_upsnbr,t_bdy_int,logi_upsnbr,theta_rcvr_s,
    theta_proj_s,d] = find_upsnbr(X_cart,Y_cart,phi_noiseless)
%
% Given:
% X_cart - the Cartesian coordinates along x-axis
% Y_cart - the Cartesian coordinates along y-axis
% phi_noiseless - the time series of each turbine angle
%
% Find:
% count_upsnbr - the number of upstream neighbours for each turbine
    at each timestamp
% t_bdy_int - timestamps when wake patterns change
% logi_upsnbr - logical array showing when a turbine is shadowed by
    wake(s)
% theta_rcvr_s - off-path angle of the wake shadowed turbine
% theta_proj_s - off-path angle of the wake producing turbine
% d - distance between turbines

%% predefine
theta_max = 20; % width of turbine wake

%% find logical indices of upstream (& downstream) turbines
t = [1:size(phi_noiseless,1)]';
t_len = size(phi_noiseless,1);
```

```
s_len = numel(X_cart);
Xs = X_cart(:);
Ys = Y_cart(:);

theta_proj_s = cell(1,s_len);
logi_proj_s = cell(1,s_len);
theta_rcvr_s = cell(1,s_len);
logi_upsnbr = cell(1,s_len);

X_mat1 = repmat(Xs',s_len,1);
Y_mat1 = repmat(Ys',s_len,1);
X_mat2 = repmat(Xs,1,s_len);
Y_mat2 = repmat(Ys,1,s_len);
[phis,d] = cart2pol(X_mat1-X_mat2,Y_mat1-Y_mat2);
phis_deg = phis/pi*180;
% column i: where is turbine i in relation to other turbines
% row i: where are all turbines relative to turbine i

for sidx = 1:s_len
% find logical indices of upstream turbines
phi_allwrts = phis_deg(sidx,:); % where are all turbines relative
    to turbine s
phi_allwrts(sidx) = NaN;

phi_rcvr_uniq = unique(phi_allwrts,'stable');
s_blocked = [];
if numel(phi_allwrts)>numel(phi_rcvr_uniq) % when there are tbs
    being blocked
for ii = 1:numel(phi_rcvr_uniq)
phi_ii = phi_rcvr_uniq(ii);
s_ii = phi_allwrts==phi_ii;
if sum(s_ii)>1
s_ii_far = d(:,sidx)>min(d(s_ii,sidx));
s_blocked = [s_blocked,find(s_ii&s_ii_far')]; % list of tbs being
    blocked
end
end
end

theta_rcvr_s{sidx} = repmat(phi_allwrts,t_len,1);
theta_rcvr_s_int = mod(phi_noiseless(:,sidx)-theta_rcvr_s{sidx
    },360); % step 1
theta_rcvr_s{sidx} = min(360-theta_rcvr_s_int,theta_rcvr_s_int); %
    step 2
```

```
% theta_rcvr_s shows whether turbine s aligns with wake path direction
% (regardless of whether the wake is shadowing turbine s or not)

theta_proj_s{sidx} = repmat(phi_allwrts,t_len,1); % This is correct !
theta_proj_s_int = mod(phi_noiseless-theta_proj_s{sidx},360); % step 1
theta_proj_s{sidx} = min(360-theta_proj_s_int,theta_proj_s_int); % step 2
logi_proj_s{sidx} = theta_proj_s{sidx}<=theta_max; % whether other tbs leave wakes that shadows tb s
% Steps 1&2 allow the correct angle difference to be calculated, regardless
% of the discontinuity in polar coordinate.

logi_upsnbr{sidx} = logi_proj_s{sidx};
logi_upsnbr{sidx}(:,s_blocked) = 0;

end

%% section time series into sections according to patterns
t_bdy = cell(s_len,1);
t_bdy_int = cell(s_len,1);
for sidx = 1:s_len
[t_endofcurrpd,~] = find(diff(logi_upsnbr{sidx}));
t_bdy{sidx} = unique(t_endofcurrpd);
t_bdy_int{sidx} = [[t(1);t_bdy{sidx}+1] [t_bdy{sidx};t(end)]];
end

%% find the maximum number of upstream neighbours at every time instance
% Note that count_upsnbr does not tell us which upstream neighbour has more
% upstream neighbours than others. Only the maximum lenth of turbine chain
% at each time instance is recorded.
count_upsnbr = cell(1,s_len);
for sidx = 1:s_len
count_upsnbr{sidx} = double(logi_upsnbr{sidx});

logi_tfrnt = ~any(logi_upsnbr{sidx},2); % when tb s does not have upstream neighbours
if sum(logi_tfrnt)<t(end) % if tb s is not at the front row all the
```

```
    time
t_haveups = find(~logi_tfrnt);

n_tsec = size(t_bdy_int{sidx},1);
for i_sec = 1:n_tsec
t_sec1 = t_bdy_int{sidx}(i_sec,1);
t_sec2 = t_bdy_int{sidx}(i_sec,2);
t_sec = t_haveups(t_haveups>=t_sec1&t_haveups<=t_sec2);
if ~isempty(t_sec)
logi_ups_s_sec = logi_upsnbr{sidx}(t_sec(1),:);
logi_ups_all_sec = cellfun(@(x) x(t_sec(1),:),logi_upsnbr,'
    UniformOutput',false);

clear tb_chain_s_sec
i_chain = 1;
tb_chain_s_sec{1,i_chain} = sidx;
logi_ups_i_sec = logi_ups_s_sec;
logi_closed_cycle = 0;
while sum(logi_ups_i_sec)~=0 && ~logi_closed_cycle
i_chain = i_chain+1;
tb_nxt = find(logi_ups_i_sec);
tb_chain_s_sec{1,i_chain} = tb_nxt;
logi_ups_i_sec = any(cell2mat(logi_ups_all_sec(tb_nxt)'),1);
logi_closed_cycle = any(cellfun(@(x) isequal(x,tb_chain_s_sec{1,
    i_chain}),tb_chain_s_sec(1:i_chain-1)));
fprintf('Turbine %d, section %d, position in turbine chain %d \n',
    sidx,i_sec,i_chain)
end
count_upsnbr{sidx}(t_sec,logi_ups_s_sec) = size(tb_chain_s_sec,2)
    -1;
end
end
end

end
```

Bibliography

[1] A. Rytter. *Vibrational Based Inspection of Civil Engineering Structures*. PhD thesis, Department of Building Technology and Structural Engineering, Aalborg University, 1993.

[2] C. Farrar and K. Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley and Sons Ltd, 2013.

[3] K. Worden, E. Cross, N. Dervilis, E. Papatheou, and I. Antoniadou. Structural health monitoring: from structures to systems-of-systems. *IFAC*, 48:1–17, 2015.

[4] K. Worden, L. Bull, P. Gardner, J. Gosliga, T. Rogers, E. Cross, E. Papatheou, W. Lin, and N. Dervilis. Structural health monitoring: from structures to systems-of-systems. *Frontiers in Built Environment*, 6, 2020.

[5] L. Bull, P. Gardner, J. Gosliga, A. Maguire, C. Campos, T. Rogers, M. Haywood-Alexander, N. Dervilis, E. Cross, and K. Worden. Towards population-based structural health monitoring, Part I: Homogeneous populations and forms. In Z. Mao, editor, *Model Validation and Uncertainty Quantification, Volume 3*, Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham, 2020.

[6] J. Gosliga, P. Gardner, L. Bull, N. Dervilis, and K. Worden. Towards population-based structural health monitoring, Part II: Heterogeneous populations and structures as graphs. In B. Dilworth and M. Mains, editors, *Topics in Modal Analysis & Testing, Volume 8*, Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham, 2021.

[7] J. Gosliga, P. Gardner, L. Bull, N. Dervilis, and K. Worden. Towards population-based structural health monitoring, Part III: Graphs, networks and communities. In B. Dilworth and M. Mains, editors, *Topics in Modal Analysis & Testing, Volume 8*, Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham, 2021.

[8] P. Gardner, L. Bull, J. Gosliga, N. Dervilis, and K. Worden. Towards population-based structural health monitoring, Part IV: Heterogeneous populations, transfer and mapping. In Z. Mao, editor, *Model Validation and Uncertainty Quantification, Volume 3*, Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham, 2020.

[9] C. Wickramarachchi, D. Brennan, W. Lin, A. Maguire, D. Harvey, E. Cross, and K. Worden. Towards population-based structural health monitoring, Part V: Networks and Databases. In R. Madarshahian and F. Hemez, editors, *Data Science in Engineering, Volume 9*, Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham, 2022.

[10] K. Worden. Towards population-based structural health monitoring, Part VI: Structures as geometry. In S. Pakzad, editor, *Dynamics of Civil Structures, Volume 2*, Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham, 2021.

[11] W. Lin, K. Worden, A. Maguire, and E. Cross. Towards population-based structural health monitoring, Part VII: EoV fields: environmental mapping. In B. Dilworth and M. Mains, editors, *Topics in Modal Analysis & Testing, Volume 8*, Conference Proceedings of the Society for Experimental Mechanics Series. Springer, Cham, 2021.

[12] E. Papatheou, N. Dervilis, A. Maguire, I. Antoniadou, and K. Worden. A performance monitoring for the novel Lillgrund offshore wind farm. *IEEE Transactions on Industrial Electronics*, 62:6636–6644, 2015.

[13] E. Papatheou, N. Dervilis, A. Maguire, C. Campos, I. Antoniadou, and K. Worden. Performance monitoring of a wind turbine using extreme function theory. *Renewable Energy*, 113:1490–1502, 2017.

[14] L. Bull, P. Gardner, J. Gosliga, T. Rogers, N. Dervilis, E. Cross, E. Papatheou, A. Maguire, C. Campos, and K. Worden. Foundations of population-based SHM,

Part I: Homogeneous populations and forms. *Mechanical Systems and Signal Processing*, 148, 2021.

[15] P. Gardner, L. Bull, J. Gosliga, N. Dervilis, and K. Worden. Foundations of population-based SHM, Part III: Heterogeneous populations – Mapping and transfer. *Mechanical Systems and Signal Processing*, 149, 2021.

[16] J. Gosliga, P. Gardner, L. Bull, N. Dervilis, and K. Worden. Foundations of population-based SHM, Part II: Heterogeneous populations – Graphs, networks, and communities. *Mechanical Systems and Signal Processing*, 148, 2021.

[17] G. Tsialiamanis, C. Mylonas, E. Chatzi, N. Dervilis, D. Wagg, and K. Worden. Foundations of population-based SHM, Part IV: The geometry of spaces of structures and their feature spaces. *Mechanical Systems and Signal Processing*, 157, 2021.

[18] P. Gardner, L. Bull, N. Dervilis, and K. Worden. Domain-adapted gaussian mixture models for population-based structural health monitoring. *Journal of Civil Structural Health Monitoring*, 12:1343–1353, 2022.

[19] J. Poole, P. Gardner, N. Dervilis, L. Bull, and K. Worden. On statistic alignment for domain adaptation in structural health monitoring. *Structural Health Monitoring*, 2022.

[20] J. Poole, P. Gardner, N. Dervilis, L. Bull, and K. Worden. On the application of partial domain adaptation for PBSHM. In P. Rizzo and A. Milazzo, editors, *Lecture Notes in Civil Engineering*, volume 270 of *European Workshop on Structural Health Monitoring. EWSHM 2022*. Springer, Cham, 2022.

[21] J. Lee and F. Zhao. Global wind report 2022. Technical report, Global Wind Energy Council, 2022.

[22] K. Dykes, P. Veers, E. Lantz, H. Holttinen, O. Carlson, A. Tuohy, A. Sempreviva, A. Clifton, J. Rodrigo, D. Berry, D. Laird, S. Carron, P. Moriarty, M. Marquis, C. Meneveau, J. Peinke, J. Paquette, N. Johnson, L. Pao, P. Fleming, C. Bottasso, V. Lehtomaki, A. Robertson, M. Muskulus, J. Manwell, J. Tande, L. Sethuraman, O. Roberts, and J. Fields. Results of IEA wind TCP workshop on a grand vision for wind energy technology. Technical report, IEA Wind TCP, 2019.

[23] J. O'Brien, T. Young, D. O'Mahoney, and P. Griffin. Horizontal axis wind turbine research: A review of commercial CFD, FE codes and experimental practices. *Progress in Aerospace Sciences*, 92:1–24, 2017.

[24] C. Archer, A. Vasel-Be-Hagh, C. Yan, S. Wu, Y. Pan, J. Brodie, and A. Maguire. Review and evaluation of wake loss models for wind energy applications. *Applied Energy*, 226:1187–1207, 2018.

[25] J.-A. Dahlberg. Assessment of the Lillgrund windfarm: Power performance and wake effects. Technical report, Vattenfall Vindkraft AB, 2009.

[26] F. González-Longatt, P. Wall, and V. Terzija. Wake effect in wind farm performance: Steady-state and dynamic behavior. *Renewable Energy*, 39:329–338, 2012.

[27] K. Lee, Z. Huque, R. Kommalapati, and S. Han. Fluid-structure interaction analysis of NREL phase VI wind turbine: Aerodynamic force evaluation and structural analysis using FSI analysis. *Renewable Energy*, 113:512–531, 2017.

[28] H. Sohn. Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365:539–560, 2007.

[29] K. Worden, C. Farrar, G. Manson, and G. Park. The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463:1639–1664, 2007.

[30] E. Cross. *On Structural Health Monitoring in Changing Environmental and Operational Conditions*. PhD thesis, Department of Mechanical Engineering, University of Sheffield, 2012.

[31] W. Hu, S. Thöns, R. Rohrmann, S. Said, and W. Rücker. Vibration-based structural health monitoring of a wind turbine system part II: Environmental/-operational effects on dynamic properties. *Engineering Structures*, 89:273–290, 2015.

[32] N. Cressie and C. Wikle. *Statistics for Spatio-Temporal Data*. John Wiley and Sons Ltd, 2011.

[33] G. Christakos and D. Hristopulos. *Spatiotemporal Environmental Health Modelling: A Tractatus Stochasticus*. Kluwer Academic Publishers, 1998.

[34] C. Ciang, J. Lee, and H. Bang. Structural health monitoring for a wind turbine system: A review of damage detection methods. *Measurement Science and Technology*, 19(12), 2008.

[35] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. Song. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable Energy Reviews*, 13:1–39, 2009.

[36] A. Motil, A. Bergman, and M. Tur. State of the art of brillouin fiber-optic distributed sensing. *Optics & Laser Technology*, 78:81–103, 2016.

[37] M. Jones. *On Novel Machine Learning Approaches for Acoustic Emission Source Localisation: A Probabilistic Perspective*. PhD thesis, Department of Mechanical Engineering, University of Sheffield, 2023.

[38] M. Haywood-Alexander. *Development of Novel Tools for Application of Ultrasonic Guided Waves in Fibre-Composites*. PhD thesis, Department of Mechanical Engineering, University of Sheffield, 2022.

[39] W. Staszewski, B. Lee, L. Mallet, and F. Scarpa. Structural health monitoring usingscanning laser vibrometry: I. Lamb wavesensing. *Smart Materials and Structures*, 13:251–260, 2004.

[40] A. Manohar, J. Tippmann, and F. L. di Scalea. Localization of defects in wind turbine blades and defect depth estimation using infrared thermography. In *Proceedings Volume 8345, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. Society of Photo-Optical Instrumentation Engineers, 2012.

[41] S. Hwang, Y. An, and H. Sohn. Continuous line laser thermography for damage imaging of rotating wind turbine blades. In *Procedia Engineering*, volume 188, pages 225–232, 2017.

[42] J. Tautz-Weinert and S. Watson. Using SCADA data for wind turbine condition monitoring – a review. *IET Renewable Power Generation*, 11:382–394, 2016.

[43] Y. Feng, Y. Qiu, C. Crabtree, H. Long, and P. Tavner. Monitoring wind turbine gearboxes. *Wind Energy*, 16:728–740, 2013.

[44] Y. Li, S. Liu, and L. Shu. Wind turbine fault diagnosis based on gaussian process classifiers applied to operational data. *Renewable Energy*, 134:357–366, 2019.

[45] D. Astolfi, F. Castellani, and L. Terzi. Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm. *Diagnostyka*, 15: 71–78, 2014.

[46] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, pages 437–472, 1996.

[47] N. Rayner, D. Parker, E. Horton, C. Folland, L. Alexander, and D. Rowell. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108:4407, 2003.

[48] R. Hijmans, S. Cameron, J. Parra, P. Jones, and A. Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25:1965–1978, 2005.

[49] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. M. noz Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. D. Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146:1999–2049, 2020.

[50] E. K. Akpinar and S. Akpinar. An assessment on seasonal analysis of wind energy characteristics and wind turbine characteristics. *Energy Conversion and Management*, 46:1848–1867, 2005.

[51] Y. Zhu, F. Li, and W. Bao. Fatigue crack detection under the vibration condition based on ultrasonic guided waves. *Structural Health Monitoring*, 20, 2019.

[52] J. Dowds, P. Hines, T. Ryan, W. Buchanan, E. Kirby, J. Apt, and P. Jaramillo. A review of large-scale wind integration studies. *Renewable and Sustainable Energy Reviews*, 49:768–794, 2015.

[53] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41:15:1–15:57, 2009.

[54] J. Manwell, J. McGowan, and A. Rogers. *Wind Energy Explained: Theory, Design and Application*. John Wiley and Sons Ltd, 2010.

[55] A. Kusiak, Z. Zhang, and A. Verma. Prediction, operations, and condition monitoring in wind energy. *Energy*, 60:1–12, 2013.

[56] T. Rogers, P. Gardner, N. Dervilis, K. Worden, A. Maguire, E. Papatheou, and E. Cross. Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression. *Renewable Energy*, 148:1124–1136, 2020.

[57] V. Heine. Models for two-dimensional stationary stochastic processes. *Biometrika*, 42:170–178, 1955.

[58] M. Stein. Space–time covariance functions. *Journal of the American Statistical Association*, 100(469):310–321, 2005.

[59] S. Shirota and A. Gelfand. Space and circular time log gaussian cox processes with application to crime event data. *The Annals of Applied Statistics*, 11(2): 481–503, 2017.

[60] A. Ezzat, M. Jun, and Y. Ding. Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting. *IEEE Transactions on Sustainable Energy*, 9(3):1437–1447, 2018.

[61] P. White and E. Porcu. Nonseparable covariance models on circles cross time: A study of mexico city ozone. *Environmetrics*, 30, 2019.

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

[63] B. Sansó, A. Schmidt, and A. Nobre. Bayesian spatio-temporal models based on discrete convolutions. *The Canadian Journal of Statistics*, 36(2):239–258, 2008.

[64] F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society B*, 72(4):423–498, 2011.

[65] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge MA, 2006.

[66] A. Cliff and K. Ord. Spatial autocorrelation: A review of existing and new measures with applications. *Economic Geography*, 46:269–292, 1970.

[67] J. LeSage and R. Pace. *Introduction to spatial econometrics*. Taylor & Francis Group, LLC, 2009.

[68] D. Griffith. *Spatial autocorrelation and spatial filtering*. Springer, 2003.

[69] L. Lee. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925, 2004.

[70] R. Frigola-Alcalde. *Bayesian Time Series Learning with Gaussian Processes*. PhD thesis, Department of Engineering, St Edmund's College, University of Cambridge, 2015.

[71] K. Worden, W. Becker, T. Rogers, and E. Cross. On the confidence bounds of Gaussian process NARX models and their higher-order frequency response functions. *Mechanical Systems and Signal Processing*, 104:188–223, 2018.

[72] T. Rogers. *Towards Bayesian System Identification: With Application to SHM of Offshore Structures*. PhD thesis, Department of Mechanical Engineering, University of Sheffield, 2018.

[73] J. Jeppsson, P. Larsen, and Å. Larsson. Technical description Lillgrund wind power plant. Technical report, Vattenfall Vindkraft AB, 2008.

[74] J. Ferziger and M. Perić. *Computational Methods for Fluid Dynamics*. Springer-Verlag, third edition, 2002.

[75] E. Cross, S. Gibson, M. Jones, and D. Pitchforth. *Structural Integrity*, chapter Physics-Informed Machine Learning for Structural Health Monitoring, pages 347–367. Springer International Publishing, 2022.

[76] M. McKay, R. Beckman, and W. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 1979.

[77] K. Worden, G. Manson, G. Hilson, and S. Pierce. Genetic optimisation of a neural damage locator. *Journal of Sound and Vibration*, 309:529–544, 2008.

[78] N. Jensen. A note on wind generator interaction. Technical report, Risø National Laboratory, 1983.

[79] S. Frandsen, R. Barthelmie, S. Pryor, O. Rathmann, S. Larsen, and J. Højstrup. Analytical modelling of wind speed deficit in large offshore wind farms. *Wind Energy*, 9:39–53, 2006.

[80] G. Larsen. A simple wake calculation procedure. Technical report, Risø National Laboratory, 1988.

[81] R. Shakoor, M. Hassan, A. Raheem, and Y. Wu. Wake effect modeling: A review of wind farm layout optimization using jensen's model. *Renewable and Sustainable Energy Reviews*, 58:1048–1059, 2016.

[82] S. Andersen, J. Sørensen, S. Ivanell, and R. Mikkelsen. Comparison of engineering wake models with cfd simulations. In *Journal of Physics: Conference Series*, volume 524, page 012161, Copenhagen, Denmark, 2014.

[83] H. Liu, Y.-S. Ong, X. Shen, and J. Cai. When gaussian process meets big data: A review of scalable GPs, 2019.

[84] S. Billings. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-temporal Domains*. John Wiley & Sons, Ltd, 2013.

[85] G. Karniadakis, I. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021.

[86] M. Jones, T. Rogers, and E. Cross. Constraining gaussian processes for physics-informed acoustic emission mapping. *Mechanical Systems and Signal Processing*, 188, 2023.

[87] A. Girard. *Approximate Methods for Propagation of Uncertainty with Gaussian Process Models*. PhD thesis, Department of Computing Science, University of Glasgow, 2004.

[88] A. Damianou. *Deep Gaussian Processes and Variational Propagation of Uncertainty*. PhD thesis, Department of Neuroscience, University of Sheffield, 2015.

[89] E. Cross, S. Gibson, M. Jones, D. Pitchforth, S. Zhang, and T. Rogers. Physics-derived covariance functions for machine learning in structural dynamics. In

A. Cury, D. Ribeiro, F. Ubertini, and M. Todd, editors, *Structural Health Monitoring Based on Data Science Techniques*, volume 21 of *Structural Integrity*. Springer, Cham, 2022.

[90] L. Bull, T. Rogers, C. Wickramarachchi, E. Cross, K. Worden, and N. Dervilis. Probabilistic active learning: An online framework for structuralhealth monitoring. *Mechanical Systems and Signal Processing*, 134, 2019.

[91] M. Howland, S. Lele, and J. Dabiri. Wind farm power optimization through wake steering. *PNAS*, 116(29), 2019.

[92] Z.-H. Zhou. *Ensemble Methods: Foundations and Algorithms*. Taylor & Francis Group, LLC, 2012.

[93] R. Gramacy. *Bayesian Treed Gaussian Process Models*. PhD thesis, Department of Computer Science, University of California Santa Cruz, 2005.

[94] T. Zhang. *On Treed Gaussian Processes for Modelling Structural Dynamic Systems*. PhD thesis, Department of Mechanical Engineering, University of Sheffield, 2018.