Non-Stationarity in Correlation Structures of Wind Turbine Data

Zum Erwerb des akademischen Grades
Doktor der Naturwissenschaften
(Dr. rer. nat.)
der Fakultät für Physik der Universität Duisburg-Essen vorgelegte

Dissertation

von

Henrik Maria Bette
aus Oberhausen

Erstgutachter:	Prof. Dr. Thomas Guhr
Zweitgutachter:	Prof. Dr. Joachim Peinke
Tag der Disputation	24.05.2024

List of Publications

Parts of this thesis are included in the following publications:

[1] H. M. Bette, E. Jungblut, and T. Guhr. *Nonstationarity in correlation matrices for wind turbine SCADA-data.* Wind Energy 26.8 (2023), 826-849.

[2] H. M. Bette, C. Philipp, M. Wächter, J. Freund, J. Peinke, and T. Guhr. *Dynamics of wind turbine operational states.* (2023). arXiv:23.10.06098 [physics.flu-dyn].
submitted to Wind Energy

[3] C. Philipp, H. M. Bette, M. Wächter, J. A. Freund, T. Guhr, and J. Peinke. *Extension of the Langevin power curve analysis by separation per operational state.* (2023). arXiv: 2305.15512 [physics.data-an].

[4] H. M. Bette, M. Schreckenberg, and T. Guhr. *Sensitivity of principal components to system changes in the presence of non-stationarity.* Journal of Statistical Mechanics: Theory and Experiment 2023.10 (2023), 103402.

[5] E. Jungblut, H. M. Bette, and T. Guhr. *Spatial Structures of Wind Farms: Correlation Analysis of the Generated Electrical Power.* (2022). arXiv: 2210.12960 [stat.AP].
submitted to Physica A

The following publication is not part of this thesis:

[6] R. Knecktys, H. M. Bette, R. Kiesel, and T. Guhr. *Risk Theory and Pricing of "Pay-for-Performance" Business Models.* (2022). arXiv: 2212.09585 [q-fin.MF].

The abstract of this thesis contains parts from the abstracts of publications [1–5]. Chapter 3 contains modified parts of publication [1]. The Chs. 4 to 8 are based on the publications [1] to [5], respectively.

My contributions to the publications [1–5] were as follows:

[1] The objectives of the research were defined by T. Guhr and myself. I developed the necessary tools and carried out the data analysis with support from E. Jungblut. The text was written by me with contributions from E. Jungblut and T. Guhr.

[2] The objectives of the research were defined by all authors together. C. Philipp and myself developed the necessary tools and carried out the data analysis. The text was written by myself with contributions from the other authors.

[3] The objectives of the research were defined by all authors together. C. Philipp and myself developed the necessary tools and carried out the data analysis. The text was written by C. Philipp with contributions from myself and the other authors.

[4] The objectives of the research were defined by T. Guhr and myself. I developed the necessary tools and carried out the simulation study and data analysis. The text was written by myself with contributions from M. Schreckenberg and T. Guhr.

[5] The objectives of the research were defined by T. Guhr, E. Jungblut and myself. E. Jungblut developed the necessary tools and carried out the data analysis with my support. The text was written by E. Jungblut with contributions from myself and T. Guhr.

Abstract

We investigate non-stationarity in the mutual correlations of wind turbine data and to some extent its effect on existing analysis methods. Our data stems from Supervisory Control and Data Acquisition (SCADA) systems, with which all modern utility-scale wind turbines are equipped.

First, we evaluate Pearson correlation matrices for a variety of observables at a single turbine with a moving time window. We show that a clustering algorithm applied to the correlation matrices reveals distinct states. We develop the method on a single turbine and then show that it easily transfers to multiple turbines. Thereby, we find that the state is primarily determined by wind speed. This is in accordance with known turbine control systems. Our analysis shows that for high frequency data the control mechanisms of a turbine lead to automatically detectable non-stationarity in the correlation matrix. The presented methodology allows an automated distinction of the operational states solely based on SCADA data.

Moreover, we combine the clustering analysis with a construction of a stochastic process to study the dynamics of those states in more detail. Calculating the distances between correlation matrices we obtain a time series that describes the behavior of the complex system in a collective way. Assuming this time series to be governed by a Langevin equation, we estimate the deterministic (drift) and stochastic (diffusion) components of the dynamics to understand the underlying non-stationarity. After adapting our method to specific features of our data, we are able to study the dynamics of operational states and their transitions as well as to resolve hysteresis effects.

Next, we study the influence of the measured non-stationarity on the established method of Langevin power curves. The estimation of drift and diffusion for the power output conditioned on wind speed is improved by accounting for the identified operational states. The operational states effectively separate between distinct dynamics. Thereby, we show that for each operational state only one fix point in the power output exists for each wind speed.

Non-stationarity also affects change detection. We study the sensitivity of principal components to changes in arbitrary correlated systems that display non-stationarity. We perform a numerical study to analyze changes that occur in mean values, standard deviations or correlations of the system's observables. Running Monte Carlo simulations for different system dimensions and numbers of normal states, we clearly show that knowledge about the non-stationarity of the system greatly improves change detection sensitivity for all principal components. We illustrate our results with an example using real traffic flow data.

Lastly, we study correlations of a whole wind farm instead of single turbines. The operational data of two offshore wind farms are analyzed. For the correlations of the active power between turbines over an entire wind farm, we find a dominant collective behavior. We manage to subtract the collective behavior and find a significant dependence of the correlation structure on the spatial structure of the wind farms. Our method provides a tool for aggregated assessment of turbine interactions in a wind farm.

Zusammenfassung

Wir untersuchen die Nichtstationarität in den wechselseitigen Korrelationen von Windturbinendaten und in Teilen ihre Auswirkungen auf bestehende Analysemethoden. Die verwendeten Daten stammen aus Überwachungs- und Datenerfassungssystemen (Supervisory Control and Data Accquisition, kurz SCADA), mit denen alle großen, modernen Windenergieanlagen im ausgestattet sind.

Zunächst werten wir Pearson-Korrelationsmatrizen für eine Vielzahl von Beobachtungsgrößen an einer einzelnen Turbine auf einem gleitenden Zeitfenster aus. Wir zeigen, dass ein Clustering-Algorithmus unterschiedliche Zustände erkennen lässt. Die entwickelte Methode ist problemlos auf mehrere Turbinen übertragbar. Wir zeigen, dass der Zustand primär durch die Windgeschwindigkeit bestimmt wird. Dies passt zu bekannten Turbinenkontrollsystemen. Die vorgestellte Methodik erlaubt eine automatisierte Unterscheidung der Betriebszustände allein auf Basis von SCADA-Daten.

Wir erweitern die Clustering-Analyse mit der Konstruktion eines stochastischen Prozesses, um die Dynamik der Zustände genauer zu untersuchen. Der Abstand der akutellen Korrelationsmatrix zu einem Clusterzentrum erweist sich als Zeitreihe, die das Verhalten des komplexen Systems auf kollektive Weise beschreibt. Unter der Annahme, dass diese Zeitreihe durch eine Langevin-Gleichung bestimmt wird, schätzen wir die deterministischen (Drift) und stochastischen (Diffusion) Komponenten der Dynamik. Nach Anpassung unserer Methode an die spezifischen Merkmale unserer Daten sind wir in der Lage, die Dynamik der Betriebszustände und ihrer Übergänge sowie Hystereseeffekte zu untersuchen.

Weiterhin analysieren wir den Einfluss der identifizierten Nichtstationarität auf die etablierte Methode der Langevin-Leistungskurven. Die Schätzung von Drift und Diffusion für die von der Windgeschwindigkeit abhängige Leistungsabgabe wird durch die Berücksichtigung der Betriebszustände verbessert. Diese trennen effektiv zwischen verschiedenen Dynamiken. Wir zeigen, dass es für jeden Betriebszustand nur einen Fixpunkt in der Leistungsabgabe pro Windgeschwindigkeit gibt.

Nichtstationarität wirkt sich auch auf die Erkennung von Systemveränderungen aus. Wir untersuchen die Empfindlichkeit der Hauptkomponenten gegenüber Veränderungen in beliebigen korrelierten Systemen, die Nichtstationarität aufweisen. Mithilfe einer numerischen Studie analysieren wir Änderungen in Mittelwerten, Standardabweichungen oder Korrelationen der Beobachtungsgrößen des Systems. Monte-Carlo-Simulationen zeigen deutlich, dass Wissen über die Nichtstationarität eines Systems die Erkennung von Änderungen mit Hauptkomponenten erheblich verbessert. Wir veranschaulichen unsere Ergebnisse anhand eines Beispiels mit realen Verkehrsflussdaten.

Zuletzt untersuchen wir die Korrelationen innerhalb von zwei vollständigen Windparks. Für die Korrelationen der Wirkleistung zwischen den Turbinen in einem Windpark finden wir ein dominantes kollektives Verhalten. Es gelingt uns, das kollektive Verhalten zu subtrahieren. Danach identifizieren wir eine signifikante Abhängigkeit der Korrelationsstruktur von der räumlichen Struktur der Windparks. Unsere Methode bietet ein Instrument zur aggregierten Bewertung der Wechselwirkungen zwischen den Turbinen in einem Windpark.

Acknowledgements

While my name is printed on the title page of this thesis, it would not have been possible without constant support, help and sometimes also distraction by many others.

First among those to be mentioned is certainly my supervisor, Thomas Guhr. Without his initiation of the project this thesis would never have been started. During my work he provided countless suggestions and improvements. I could always turn to him when looking for advice and help. Thanks for everything!

For scientific discussions and a great work environment, I must also thank the other members of my working group: Toni Heckens, Juan Henao Londoño, Shanshan Wang, Nico Hahn, Nils Gluth, Felix Meier, Edgar Jungblut, Thorben Karlshofer, Lars Uebbing, Katharina Prinz and Efstratios Manolakis. I think that lively discussions are at least as important as silent sessions in front of equations or computer code in science. Also, some of you are great at providing healthy breaks during long days.

As this thesis was written as part of a large research project, I had the advantage of having a second great work environment! I want to thank everybody, who had any part in the WiSAbigdata project for fruitful discussions. Special thanks for regular meetings and answers to all my wind related questions go to Christian Wiedemann, Timo Lichtenstein, Karoline Pelka, David Bastine, Victor von Maltzahn, Matthias Wächter, Jan Freund and Joachim Peinke. And thanks again to Christian, to whom I could always vent about any frustrations and problems during this work.

Some of the greatest coffee breaks – even though often online – were provided by my old study colleagues: Tobias Braun, Tim Vranken and Lukas Martinetz.

For organizational help with more things than I can remember, I want to thank Annemarie Tappert.

Now, of course, there were also those people that had nothing to do with the physics contents of this thesis, but who had to listen to me yapping on about them far too often. Thank you for your patience, support and reminders that other things might also be fun! I cannot name you all. But I must at least thank Max, who surely now has forgotten more about complex systems than he ever wanted to hear in the first place.

Thanks to my family, especially my parents and my sister, without whom I would have never had the chance or means to write this thesis. You put me on this path and were always ready to help when needed!

And lastly, thank you, Mona! Whenever I thought that I might never finish this, you did not allow those doubts to grow.

Contents

CHAPTER 1

Introduction

Wind energy, being a widely available resource, was a popular driver for economic development throughout history [7, 8]. With the invention of steam power and the use of fossil fuels, wind – at first – became less popular. However, fossil fuels were never as readily available. Furthermore, their supply chain is prone to crisis [8]. We see peaks in the development and use of wind energy (among other renewable sources) in times of fossil related uncertainty [8]. Today we live in the era of climate change. Fighting this crisis necessitates a change in our energy production on a global scale [9]. At the same time the amount of energy consumed by humanity is ever increasing [10]. Additionally, oil and gas retain their supply chain problems. Recently, this was made all too clear by the Russian invasion of Ukraine [11]. Many countries now acknowledge the problems that come with their dependence on fossil supplies from other – and often less democratic – countries. Unsurprisingly, renewable energy sources are more popular than ever [10]. In this context, the main sources of green energy are solar, water and wind. They all have their own advantages and problems. Therefore, all of them are part of a successful energy transformation.

Focusing on wind, we see an accordant rise in the capacity of wind energy production globally and locally [12]. This, in turn, drives innovation and development ranging from physical and technical issues to sociological and economic ones [13]. Researchers aim for improved understanding of atmospheric flows and the interactions of wind turbines within a wind farm, optimal design (including materials) of wind turbines, optimized control strategies for wind turbines and a strategy for the integration of wind energy into the power grid [14]. One outcome of these efforts is a variety of types and sizes of wind turbines. These range from small devices, which are primarily designed for private use, over mid-size turbines powering remote locations or small businesses to large turbines generating multiple Gigawatts, which serve as public power stations [7, 8, 15, 16]. While local solutions are sometimes reasonable, in general the efficiency of large turbines is much higher. This leads to an ever increasing size of turbines [16, 17]. Apart from size there are differences in the employed business models. Some private cooperatives run only a few or even just one large turbine. Other big energy companies run whole farms of large wind turbines. These are generally divided again into onshore and offshore turbines. The first term refers to any wind turbine being built on land, the second to those built within the sea. While onshore is, rather obviously, easier in terms of building, accessibility and maintenance, it brings along a whole range of problems. From a technical viewpoint there are many obstacles that reduce the wind speed on land ranging from mountains to trees. Politically, setting up new wind turbines is desired in general, but locally hindrances appear as turbines produce noise and moving shadows. Furthermore, some people feel that they disfigure the landscape. Many of these problems are solved with the use of offshore turbines built in the sea. The wind speeds are higher and steadier, obstacles are rare and no people live close-by. Of course, especially construction and maintenance are more complicated and costly offshore. Together with the harsh environmental conditions, this makes operation and maintenance one of the main cost drivers for offshore wind energy [18]. Accordingly, research efforts are undertaken both in the scientific and commercial community to reduce these costs. One way is to make the machines themselves more robust and reliable on a material level [19, 20]. Another approach has become possible due to increasing amounts of data from the turbine's own Supervisory Control and Data Acquisition (SCADA) system: optimization of control and operating strategies for single turbines as well as for whole farms and the predictive analysis of possible failures [18, 21, 22]. While the first increases efficiency of running turbines, the latter allows for better planning of service missions and helps to reduce costly downtimes.

The possible improvements to clean energy production provide enough societal motivation to tackle the optimization challenges in wind energy. Additionally, the system itself is quite interesting from a physics point of view. Traditional disciplines, like aerodynamics and fluid dynamics, are obviously connected to wind energy generation. Early on, scientists such as Gottfried Wilhelm Leibniz, Daniel Bernoulli, Leonhard Euler and Charles Augustin de Coloumb studied windmills and their rotors [8]. Moreover, modern wind turbine systems also display many interesting qualities that make a study worthwhile from a statistical physics viewpoint. Wind turbines operate in complex environments where external conditions change continuously. They are subjected to various stochastic processes such as turbulent wind flow on short time scales [16]. Simultaneously, weather changes and seasonal effects create non-stationarity on mid-term and long-term scales. In these conditions a modern wind turbine is a complex machine in itself. Many mechanical and electrical parts are governed by a control system, which tries to ensure optimal efficiency depending on the current conditions [23]. This further increases the complexity of the total system. In fact, multiple operational states are created in which the system behaves differently. Such a behavior might be compared to different phases in a physics system, where transitions are induced by changes in the external conditions. Looking past the individual turbine, complexity is further increased as the operation of turbines in a wind farm is not independent from each other. Wind turbines interact with each other through wake effects, usually described in research by complex fluid dynamics [24]. Incorporating randomness, multi-agent interaction, non-linearity and non-stationarity, it is well justified to consider wind turbines as an example of a complex system in the physics-sense.

The interdisciplinary field of complex systems touches upon various scientific domains, engineering fields, and the social sciences. While it originates from statistical physics and non-linear dynamics, where phenomena such as magnetization in spin systems and phase transitions are studied [25–28], the developed analytical and computational tools were soon adopted to other fields – usually those that exhibit randomness, non-linearity, non-stationarity and other complicated phenomena. Those tools have proven useful to analyze systems containing many interacting individual components or agents. The focus lies on uncovering collective dynamics, often resulting in phenomena that cannot be easily predicted by examining the components in isolation [29]. In biology, the applications range from understanding the flocking behavior of birds and schooling patterns of fish to the complex organization of cells and evolutionary processes [30–32]. Geophysicists also employ complex systems theory to understand natural phenomena like earthquakes and weather patterns, which result from the interplay among subsystems like tectonic plates or atmospheric layers [33–36]. Thus, complex system science also plays an important role in the study of Earth's climate system [37–40]. Important work in this field was honored with the Nobel prize "for groundbreaking contributions to our understanding of complex physical systems" in 2021 further underlining the importance of the discipline. Syukuro Manabe and Klaus Hasselmann were awarded half of it "for the physical modelling of Earth's climate, quantifying variability and reliably predicting global warming". The other half went to Giorgio Parisi "for the discovery of the interplay of disorder and fluctuations in physical systems from atomic to planetary scales" [41]. Furthermore, complex system science have proven their applicability also in initially less physics-related fields. Econophysics leverages its concepts to better understand the intricate dependencies in economic systems such as the stock market [42–45]. Finance tools like risk assessment and portfolio management now rely on results from complex systems theory [46, 47]. Sociology applies complex system models to study social networks, opinion dynamics, and even collective decision-making [48–50]. In traffic, a system influenced by many agents and their individual behavior, the complex systems approach helps to understand the resulting traffic flow phenomena and model whole urban environments [51–54]. The list goes on to epidemiology [55, 56], linguistics[57], neuroscience [58, 59] and many more. Not least among the common denominators of these systems is non-stationarity. Economic relations change, traffic jams from and dissolve again, and even the naming of biological evolution and climate change contradicts stationarity.

Similar to these systems, wind turbines and farms are also subject to non-stationarity. The conditions, in which they operate, change constantly on short and long time scales. Hence, as a collective result from external factors such as wind speed and internal processes turbine behavior changes as well. Changing the view from individual to multiple turbines in a wind farm, the collective behavior is also non-stationary. The availability of high-frequency SCADA data from wind turbines now facilitates the study of wind turbine systems with complex systems methods. Thereby, we gain a deeper understanding of the interactions and behaviors of the different observables. In this thesis we especially focus on the analysis of non-stationarity in wind turbine systems as well as its consequences for other analyses, such as failure detection. We study cross-correlations between different available data channels. The analysis of correlation matrices has proven to be a useful tool in other complex systems like financial markets or traffic [42, 51]. It has helped to understand stability and non-stationarity in these systems. We therefore explore how it can help to classify and understand wind turbine behavior without additional external input. This includes, on the one hand, correlations between different measurements for single turbines and, on the other hand, correlations between different turbines in one wind farm.

1.1 Outline of this Thesis

We start by establishing some basic knowledge about wind energy and especially wind turbines in Ch. 2. We present a brief history of wind energy, describe the sophisticated machines they have become today and discuss their importance. In Ch. 3 we present the theory needed throughout this thesis. This includes correlation matrices, whose analysis has been helpful in many other complex and interconnected systems to characterize the mutual dependencies and identify non-stationarity. We also look at Principal Component Analysis, a data analysis technique that is closely connected to the correlation matrices via their eigenvectors. It helps us to understand the meaning of the eigenvectors and their connection to finding and predicting failures in correlated systems.

In this thesis, we focus on analyzing non-stationarity in high frequency wind turbine data through the mutual correlations of observables. We first study the correlation matrices of different observables measured at individual wind turbines. We characterize their non-stationarity on time scales of several minutes and explore its influence on other data analysis methods in the field. Furthermore, we show how the analysis of correlations between all turbines in a farm provides a statistical assessment of the total interactions between multiple turbines.

The analysis of non-stationarity in the correlation matrices of single turbines is carried out in Ch. 4. It lays the ground for any subsequent analysis thereof. Applying a clustering algorithm to a time series of correlation matrices allows the distinction of multiple distinct clusters. These are shown to be linked to the control system of the turbine. We characterize their main dependency on the current wind speed. The clusters represent different operational states of the turbine. In subsequent chapters, the terms clusters and states are often used somewhat interchangeably. During our analysis this is no problem as we identify clusters with states, but generally speaking the term cluster rather refers to the mathematical group while the term state characterizes the condition of the analyzed system corresponding to the grouping. The method is shown to be transferable onto multiple wind turbines. It effectively allows an automatic distinction of different operational states, which can be used as an input for other analysis methods. Applying stochastic process analysis in Ch. 5 facilitates an understanding of the state dynamics. For this, we construct a measure based on correlation matrix distances, which describes the behavior of the system in an collective way. Langevin drift maps and resulting potential landscapes for this measure reveal the state transition dynamics. An adaptation of the drift estimation process allows us to resolve hysteresis in the dependency of operational states on wind speed.

After analyzing the non-stationarity directly, we study how it affects other analyses. First, we once again combine the non-stationarity analysis with Langevin analysis in Ch. 6. Here, we study how Langevin power curve analysis, i.e. stochastic analysis of the produced power output versus wind speed, might be improved by accounting for the different operational states. Some regions of the Langevin power curve are shown to exhibit different properties per individual state. The traditional analysis that does not take the non-stationarity into account, only reveals the average of these separate behaviors. Second, we study in Ch. 7 how non-stationarity influences change (or novelty / anomaly / failure) detection in arbitrary correlated systems. We shift the focus from wind energy to correlated systems in general with a theoretical and systematic analysis. Based on earlier results for stationary systems, we carry out a simulation analysis of the effect of non-stationarity on change detection with Principal Component Analysis. The results clearly show that in a system with multiple possible states – as we find it for wind turbines in Ch. 4 – the detectability of small changes is hindered if no knowledge about the non-stationarity exists. The chapter ends with a traffic system example of how the results transfer into real data analysis. This concludes the exploration of non-stationarity in single wind turbine correlation matrices and its effect on existing methods.

In Ch. 8 we return to wind energy systems, but broaden our view to whole wind farms. The correlations of active power outputs measured at all turbines in two wind farms are studied. After subtracting the collective behavior, distinct correlation structures emerge. We show that they depend on wind direction. The analysis provides a statistical tool to quickly assess the total interaction between multiple turbines, which arise from complicated processes such as the wake effect.

To conclude this thesis, we summarize and discuss our findings in Ch. 9.

CHAPTER 2

Wind Energy

This chapter gives a brief introduction into the field of wind energy. We do not attempt to cover all aspects of this broad topic. Since the statistical analyses performed in this thesis are not primarily specific to wind turbines, it is not necessary to become familiar with all the details of how wind turbines and farms operate. However, to understand the results, their interpretation and the motivation behind the presented analyses, it is important to have a basic grasp of the topic. Hence, the focus in this chapter lies on familiarizing the reader with the general concept of wind energy and the used technology as well as highlighting its importance. Thereby, we aim to pave the way for understanding of the environment in which the present study is carried out.

2.1 The Wind Resource

Wind energy is an abundant yet highly variable resource. It is subject to both geographical and temporal fluctuations. The amount of available energy in the wind is proportional to the cube of the wind speed, making the predictability and understanding of wind patterns crucial for efficient energy harvesting. We provide a brief outline of the different aspects based on the more extensive description by Burton et al. [16].

Geographically, wind energy is influenced by a variety of factors, primarily driven by the sun's energy. The sun heats the Earth's surface unevenly, leading to the rise of warm air in some areas and the descent of cooler air in others. Coriolis forces, due to Earth's rotation, further shape these airflows, resulting in a large-scale global circulation pattern. This produces different climatic regions with varying wind conditions. Medium-scale variations exist due to the physical geography of a region, including the ratio of land to sea, the presence of mountains, and other landscape features. On a smaller scale, topographical elements such as hills, valleys, and even man-made structures like buildings or trees can have a substantial impact on wind speed.

Temporal wind variability occurs on multiple scales. Over long timescales of years or decades, wind patterns can shift in ways that are not fully understood, complicating the assessment of potential sites for wind farms. At subannual timescales seasonal changes are typically predictable, as are weather-induced variations over the span of days. However, precise prediction for extended periods remains challenging. Diurnal variations, occurring within a single day, can be strong. These depend on the location, but are often predictable enough to allow for efficient integration of wind energy systems into the electricity grid.

On the shortest timescales, wind is turbulent. Turbulence occurs over minutes, seconds, or even shorter intervals. It is largely caused by friction between the air and Earth's surface as well as thermal effects and has a significant impact on the design and performance of individual wind turbines. Theoretically, the complex processes behind turbulence could be described by a set of differential equations that take into account factors like temperature, pressure, density, humidity, and air motion. However, the process is highly chaotic, i.e. slight variations in initial conditions might lead to vastly different outcomes. Hence, in practice, statistical descriptions are often employed instead.

While this topic, and especially turbulence, is of high interest from a statistical physics point of view, they are not the prime subject of the present study. Burton et al. [16] and Gasch & Twele [7] are recommendable for further reading on this topic from a wind energy viewpoint. For the sake of this study the main take away is that wind turbines operate in a highly variable and stochastic, non-stationary environment.

2.2 A Brief History of Wind Power

The use of wind energy by humanity dates back to the first sails billowing in the wind. However, the generation of mechanical power from wind also has a history that starts with ancient civilizations, such as the Persians [7, 8, 60]. The codex of Babylonian King Hammurabi, who reigned from 1792 to 1750 BC, mentions the plan to apply wind power in irrigation projects. However, it is unsure if this was implemented [61]. Similarly, some researchers claim to have found remains of 3000 years old windmills in Egypt near Alexandria [62]. Later on, in the same location Heron of Alexandria put down a detailed description of a wind powered organ in the first century. While this is primarily a toy, the design already features a horizontal axis windmill [63].

The first practically used windmills, of which we are sure, were Vertical Axis Wind Turbines (VAWTs) or rather Vertical Axis windmills. They consisted of a vertical driveshaft with six to twelve rectangular sails [64]. They made their appearance in Persia on the modern day territories of Iran, Pakistan and Afghanistan between the 7th and 9th century [8, 65]. Their primary functions were to grind grain and pump water. Later, an arguably separate development took place in Middle Age Europe. Here, windmills with a horizontal driveshaft and cloth-covered wooden blades make their appearance. They were among the first Horizontal Axis Wind Turbines (HAWTs). Historical documents point to early sunk-post mills in the province of Brabant (1119) and Normandy (1180). While the dates on these documents are subject of uncertainty, we have definite documentation of a windmill dated 1185 in Weedley, Yorkshire [8]. The horizontal driveshaft made yawing necessary. For sunk-post mills this was realized by mounting the whole wooden building on a post sunk into the earth around which the building was rotated.

In contrast to water power, wind could be harvested anywhere instead of only at fast-flowing streams and was not hindered by freezing temperatures. This lead to a fast adaptation across Europe [66]. As in the Middle East, the mills were primarily used for milling grain and pumping water. A famous large-scale use of such pumps are the polders of the Netherlands, where wind power was used to claim land from the sea [8]. This necessitated the transmission of power to fixed machines such as scoop wheels for pumping water. With the moving buildings of post mills this was impractical. The hollow post mill was invented in the early 15th century: a stationary pyramid shape building on top of which the windmill was fixed via a hollow post. Through this post ran a vertical shaft powering a machine in the base building. In the Mediterranean regions the tower windmill became popular. A fixed stone tower with a wind wheel consisting typically of triangular sails. Rough yawing was sometimes achieved by manually moving the whole rotor to a different position of the tower. The need for ever more powerful windmills sparked the development from hollow post windmill to the typical Dutch windmill: A fixed, stone mill house with a rotating cap. The rotors typically consisted of four wooden blades with cloth sails. Many of these are still seen around the countryside in Europe today.

Unsurprisingly with such prominent machines, they became subject to systematic study as soon as physical-mathematical thinking established itself in the 17th and 18th century [8]. Gottfried Wilhelm Leibniz involved himself in the topic and wrote about the "Wind Arts". In his works he suggested new designs as well as improved constructions for windmills [67, 68]. After formulating the basic laws of fluid dynamics, Daniel Bernoulli soon applied them to improve the design of windmill sails [8]. The twist of the sails was calculated correctly for the first time by Leonhard Euler [8, 69]. The scientific interest in windmills was, however, not only theoretical. Two Scotsmen, Meikle and Lee, invented automatic yawing with a fantail for the Dutch mill in 1750. In 1792 Meikle again improved the Dutch windmill: Till then the sailcloth on the blades had to be reefed by hand during storms. Meikle replaced the sails with hinged shutters connected by a rod. These could be opened and closed easily. The wooden – and later sheet metal – shutters were sometimes even connected with springs to the surrounding frame, effectively automating the process. A crude speed and power regulation of the windmill became possible for the first time [8]. Later systematic aerodynamic experiments in 1821 by Charles Augustin de Coloumb showed that the new blade design was, however, less efficient than traditional sailcloth blades [70]. With all these efforts, the Dutch windmill was refined, so that it became the dominant windmill type in the middle of the 19th century, when wind power, in general, was an important economic factor with around 200,000 mills in Europe [71]. In 1920 Albert Betz, director of the AERODYNAMISCHE VERSUCHSANSTALT GÖTTINGEN (Aerodynamical Research Institute Göttingen), which is today's DEUTSCHES ZENTRUM FÜR LUFT- UND RAUMFAHRTTECHNIK (DLR, German Aerospace Center), investigated the rotor blades further. He established the physics behind wind-energy conservation and found Betz's law: No machine can capture more than 59.3% of the kinetic energy of the wind. Although named after Betz, it was derived by two more scientists, Lanchester and Joukowsky, of that era independently [72]. This research, together with advancements in modern airfoil design from aircraft engineering, lead Major Kurt Bilau to undertake the last large improvement on the Dutch windmill. He partnered with Betz and developed the "Ventikanten" sail [8]. When looking at a Dutch windmill equipped with these aluminum airfoil blades, it is hard to think of it as a traditional windmill. They resemble a fusion of modern and traditional technology.

A different development took place on the American continent in the early 19th century, where rapidly moving settlers and railroad constructions were in need of lighter windmills to pump water as they moved westward. Additionally, settlers complained about the first available wind wheels with sailcloth blades, which needed constant supervision to avoid damage during strong winds. Around 1850 Daniel Halladay invented a solution: the Westernmill [73]. Inspired by safety valves in steam engines, which were opened by flywheel governors, he designed a self-regulating wind wheel. The wooden blades were loosely suspended on a ring and connected by a second movable ring collar. This, in turn, was triggered by flyweights. Furthermore, the wheel was divided into six areas, that could be pitched out of the wind separately. Covering the whole surface of the wheel with thin blades, very different aerodynamic characteristics compared to the classical European mills were achieved. The wheel starts turning at low speeds and turns relatively slow while generating high torque. Ideal conditions for operating the water pumps. The only problem remained in the complexity of the machine. Reverend Leonhard R. Wheeler soon developed a simpler solution [8]. Instead of multiple areas the wind vane responsible for yawing was suspended with springs and a second fixed vane was installed parallel to the wheel surface. In strong winds this would cause the whole wheel to turn out of the wind. These mills soon became popular with the production total adding up to 6 million by 1930. Wind energy was a mass-produced article for the first time [8].

After the dawn of electricity, it did not take long until wind turbines were adapted to generate electrical power. While big power plants were soon installed in cities, the electrification of rural areas was a slow process. It is likely that innovative farmers, for example in America, were among the first to equip their existing windmills with small generators [8]. The first documented turbine to produce electricity was installed at the Vienna Electrical Exhibition in 1883 by Josef Friedländer [74]. A few years later in 1887 James Blyth, an electrical engineering professor in Glasgow, powered electrical installations at his holiday cottage with a small vertical axis turbine in the garden [75]. A bigger project was undertaken one year later in Cleveland, Ohio, by Charles F. Brush. His turbine had a 17 m rotor, produced up to 12 kW and was used to charge batteries [76]. A key figure in the transition to modern wind energy technology was Danish scientist Poul la Cour. First, researching and perfecting traditional windmills, he then turned to producing electricity with wind turbines. While his early turbines are designed much like a traditional windmill for simple use in rural areas, he focused much of his research on aerodynamics. He carried out wind tunnel experiments, likely the first ones, and founded the ASSOCIATION OF DANISH WIND POWER ENGINEERS [8, 77]. Even though La Cour's design was manufactured industrially by the LYKKEGARD company and used to power rural homes, the interest in wind energy was waning during the early 20th century. The growth and adoption of fossil fuels, coupled with the advent of steam engines, pushed wind energy to the margins of energy production. Only when fossil resources became problematic and expensive during World War 1 and 2 did interest arise again. This led, for example, to the construction of a first really large turbine in 1941 in the USA: The Smith-Putnam wind turbine with a rotor diameter of 53 m and a rated output of 1250 kW [78]. Other early examples include the 100 kW Balaclava wind turbine in the USSR in 1931 with a rotor diameter of 30 m, the 200 kW 24 m diameter Gedser machine built in 1956 in Denmark, a 1.1 MW test turbine in France with a diameter of 35 m and the 1942 MAN-Kleinhenz project with 10 MW using a 130 m diameter rotor [8, 16]. Fascinating because of its different technology is the Andrea Enfield 100 kW turbine constructed in the UK during the 1950s. It's 24 m diameter rotor consisted of hollow blades with openings at the tip. Centrifugal forces created an air stream through the turbine, which powered an air turbine sitting in the tower [8]. Some influential designs right to this day were developed by Ulrich Hütter, who also had a part in founding the STUDIENGESELLSCHAFT WINDKRAFT E.V. (Research Association Wind Energy) in 1949. In 1958 he started to develop the W-34, which had a rotor diameter of 34 m and a rated power of 100 kW. Here, the aerodynamically refined airfoil rotor blades were made of advanced fibre-glass material and connected to the rotor shaft via a teetering hub, which allowed for compensation of asymmetrical aerodynamic loads [79].

It took the 1970s oil crisis to reignite broad interest in wind power as an alternative and sustainable energy source. Oil prices were up and Western countries were made painfully aware of their dependency on imports. Various states provided extensive funds to researchers and industries for large scale prototype turbines. There was no consensus yet as to which type of turbine and which number of rotor blades would be most efficient. This led to very different prototypes all over. In the USA, NASA, the NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, was tasked with this problem. In cooperation with industrial companies two bladed prototypes from the 100 kW Mod-0 with a 38 m rotor diameter in 1975 to the 2.5 MW Mod-5B with a 97.5 m rotor diameter in 1987 were built [8]. The USA also tried vertical axis turbines. A 34 m diameter Darrieus type was built in the Sandia Vertical Axis Test Facility. A 4 MW turbine of the same type was constructed in Canada under the name Éole [16]. Operation was, however, not promising and the project was decommissioned soon after. In Denmark, original government-funded research was focused on overhauling the 200 kW three-bladed Gedser turbine in cooperative efforts with NASA. Other three-bladed turbines were erected such as the Tvind turbine with 2000 kW power output and 52 m rotor diameter and the Nibe A and B turbines with 630 kW each and rotor diameters of 40 m. In Germany, subsidised work on wind energy dates back to 1974. The DLR and the FWE, FORSCHUNGSINSTITUT FÜR WINDENERGIE (Research Institute for Wind Energy), both worked on a study assessing the possibilities of wind energy. The study finds that the previously constructed W-34 turbine with two rotor blades of Ulrich Hütter provided appropiate technology for modern turbines. Even though the study suggested a smaller test, politics called for an upscale model of the W-34 with a rotor diameter of 113 m and 3 MW power output. The project was titled GROWIAN (Große Windkraft-Anlage).

Many technical problems were encountered and subsequently bad press followed. However, the final success in wind energy utilization could not be stopped anymore [8]. Further experimental turbines in Europe followed in the late eighties, but were of smaller dimensions. Also, the focus had shifted from national programs to coordinated programs of the EU Commission. In the WEGA-I program a large number of operating wind turbines, among others the so-called GROWIAN 2 with 1.2 MW on Helgoland, were evaluated [80]. Strategic studies were performed concerning the relationship of turbine size and economic feasibility as well as development in the MW-range. The subsequent WEGA-II program already focused on commercial prototypes [81]. All in all, the first generations of large prototypes did not meet the political expectations in terms of economic energy generation. However, they laid the foundations for modern wind turbines. A scientific and industrial basis was formed for present-day wind energy technology [8].

Commercial use at that time was restricted to smaller turbines. Denmark was able to promote small wind turbines for private use. Earlier developments were picked up again and, by 1990, 2500 turbines with an approximated total of 200 MW had been installed [8]. Some companies such as VESTAS, which were formerly active in different fields, took to building small wind turbines after the original Danish concepts with three rotor blades and grid-connected induction generators. The first units were relatively small with only about 60 kW power output and 16 m rotor diameter. Apart from single private turbines, here also for the first time community installations consisting of multiple turbines owned by groups of people were built [8]. The first large wind farms were erected in the state of California, USA around 1980. This was sparked by additional state subsidies for regenerative energies, which offered large tax benefits to investors. While some of the early farms were constructed with American built turbines, the more refined Danish turbines were soon in demand here as well [8].

2.3 Modern Wind Turbines

Today, wind energy has transcended its status as merely an alternative source of power. It has become a mainstream and indispensable element in the global energy landscape. The isolated prototypes and small-scale systems have largely been replaced by big, grid-connected wind farms for commercial electricity production. The installed worldwide capacity of wind energy is steadily rising [82]. Advancements in materials science enable us to build larger and more efficient wind turbines, significantly increasing the megawatts generated per installation. Turbines are no longer restricted to being built on land (onshore), but large wind farms are being built offshore in the water. Here, more steady and higher wind speeds can be harvested.

Simply spoken, the goal of modern turbine design is to increase the power output P. Its dependence is well known:

$$P = \frac{1}{2} c_P(\lambda) \rho A u^3, \tag{2.1}$$

where c_P is the power coefficient, ρ is the density of air, A is the area swept by the rotor and u is the wind speed. The power coefficient is dependent on the tip

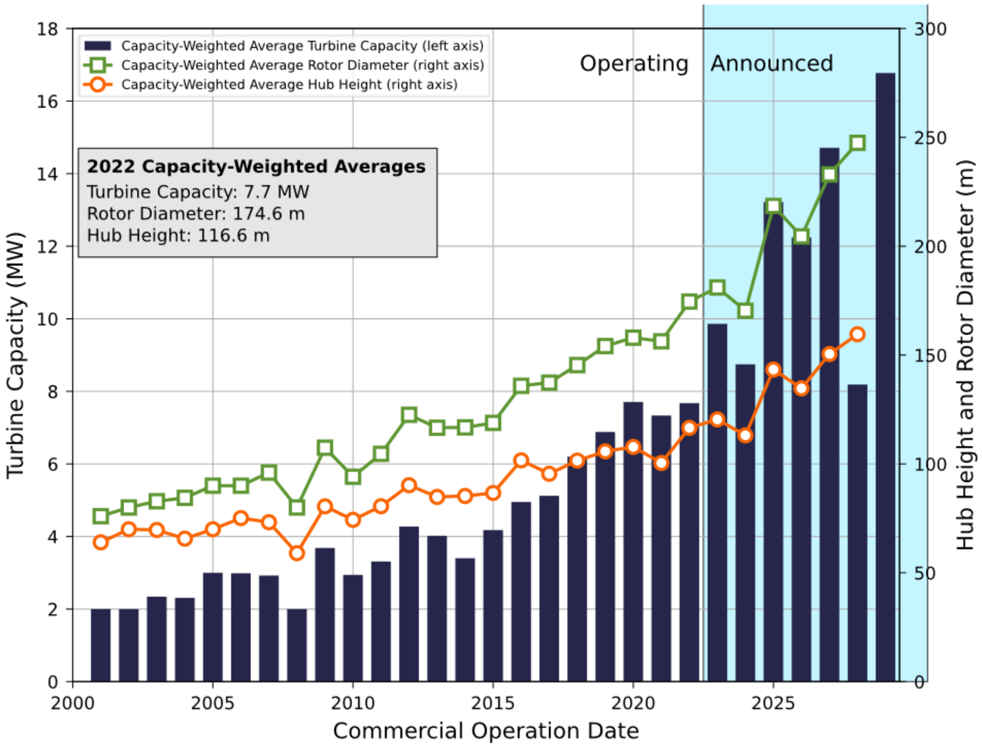


Figure 2.1: Global averages of offshore turbine rotor diameters, hub-heights and capacity. The mean is capacity-weighted, i.e. each installation contributes to this mean weighted according to its capacity. Figure taken from [17]

speed ratio λ (rotor tip speed divided by wind speed) in such a way that always one maximum exists [16].

The increase of the power coefficient is one way to make turbines more efficient. This is achieved by improved aerodynamics and other detailed changes to the turbine. Operation at variable speed ensures that the maximum power coefficient is maintained over a wide range of wind speeds. However, the Betz limit gives a theoretical maximum value of 0.593 of the power coefficient [72, 83]. The basis for this are conservation of mass and energy. Basically, one cannot extract all energy from the wind, because this would mean a standstill behind the turbine blocking new inflow.

Another lever to increase power output is the enlargement of the rotor swept area A. Doubling the rotor diameter leads to quadrupled power output. Hence, in the last years turbines have been getting larger and larger. A trend, which is expected to continue on. In Fig. 2.1, one clearly sees the steady increase in rotor diameter and subsequently also in capacity and hub-height. Some prototypes of the large expected turbines are already built [17].

The last possibility to increase power output according to Eq. (2.1) is to utilize higher wind speeds. Here, a doubling of wind speed even leads to an eight-fold increase in power output. Not surprisingly, considerable efforts are underway to ensure optimal locations of wind turbines [16]. One obvious development is the increasing amount of offshore turbines, where less obstacles are present and wind speed is generally higher. Onshore, the hub-height is not only increased to accommodate larger rotors, but also to take advantage of stronger winds in high altitudes.

The following sections provide a brief introduction to modern wind turbine technology. It is in large parts based on the more detailed overview of Ritschel & Beyer [23]. Even more extensive descriptions are found, for example, in Hau [8] and Gasch & Twele [7].

2.3.1 Turbine Types and Geometry

Many of the different concepts for turbines still exist today. VAWTs with different designs are still being built and used. They have advantages in certain applications. For example, there is no necessity for wind direction alignment and electrical components can easily be housed on ground-level [8]. However, their power coefficients are usually lower than those of HAWTs. Furthermore, the structural parts needed to construct a rotor comparable in size to modern HAWTs are large and heavy. Hence, almost everybody immediately thinks of a typical three-bladed HAWT when it comes to modern wind turbines. Some turbines with different numbers of rotor blades are used, but three blades are the lowest number, which realizes the same inertia along any axis perpendicular to the rotational axis. This makes the rotor dynamically more balanced than designs with less blades. Designs with less blades also need higher tip speeds for the same efficiency making them louder and environmentally challenging. Hence, while other forms are present, the HAWT with three blades is clearly dominant in today's large wind turbines.

For a HAWT two geometries are typically feasible: the rotor can either be downwind or upwind from the tower. While the downwind position has some advantages, the strong tower shadow is responsible for all commercial turbines to be built with upwind geometry. However, this results in the blades of the rotor bending towards the tower due to the wind. The simplest solution to this problem is sufficient horizontal distance between the rotor and the tower. Such a rotor overhang is structurally challenging and expensive. The prominent solution is therefore to tilt the rotor axis slightly (approximately 5 degrees) upward from the horizontal axis. Further distance from the tower is achieved by curving the blades away from the tower, effectively tilting them about 3 degrees out of the rotor plane.

2.3.2 Structure and Components

A wind turbine consists of many different components working together. Figure 2.2 shows the most frequent ones and their location in a typical HAWT. While some components are more crucial for the operation of the turbine and others might be considered rather auxiliary, all are part of different systems that are common in almost all modern wind turbines. Of course, the overall system is the wind turbine itself.

Looking at it from the outside, one easily distinguishes tower, nacelle and rotor. The tower rests on some form of foundation, which can either be a classical one on land or some more complicated construct for offshore turbines in the water. It provides the height necessary to hold the rotor, which in turn consists of the blades connected to the hub. This connection is not static, but rather allows for the turning of the rotor blades around their axis via a pitch system. The hub itself is connected at the main bearing to the nacelle. Here, the rest of the drive train system is hosted: main shaft, most often a gearbox, the generator and a brake. Grid integration of the created electrical energy makes further electrical systems, such as converter and transformer, necessary. Many of these components are temperature controlled by one or more cooling systems. As with the blade's pitch system, the mounting of the whole nacelle on the tower is not static either. A yaw system, consisting of bearing, drives and brakes, makes sure that the turbine can follow the wind direction.

All these systems are managed by a control system. Here, information is gathered by sensors and turned into instructions for the other systems by controller hardware and system. A stand-alone safety system is also in place to ensure stopping of the turbine in case of danger.

2.3.3 Rotor

Rotor blades are key to aerodynamic efficiency and mechanical robustness in wind energy systems. They are designed to optimize the power coefficient, but must at the same time regulate power and generate minimal noise. Aerodynamically, the blades are akin to aircraft airfoils but adapted for the specific requirements of wind turbines. These adaptations include a variable angle of attack and chord length across the blade's radius. This is necessary as the relative wind speed changes along the radius due to the rotational speed.

Of course, theoretical designs must make the transition into reality. Here, structural integrity is paramount. An interplay between modern materials and sophisticated engineering ensures that the blade can handle both static and dynamic loads, including the vibrations that might arise during operation. The outer shell is made of sandwich materials. Nowadays the predominant base material is Fiber-Reinforced Plastic (FRP). For large parts the fibers used are glass. Carbon fibers, though stronger and more suitable for highly stressed components, are generally more expensive and thus used sparingly. Internally most blades feature shear webs and spar caps above them, which create a sort of backbone that imparts stability to the blade. Here, stronger materials such as carbon fibers are often used.

The trend in rotor blade design leans towards slimmer profiles. This shift to lower solidity (blade area divided by rotor swept area) has not compromised the power performance of the blades. It necessitates higher rotational speeds to maintain efficiency, but this is somewhat desired as the same generator speed is then obtained with reduced gearbox ratios [16].

Lastly, attaching the blades to the hub is a non-trivial challenge. The blade-to-hub connection must be able to transfer high mechanical loads securely. This is usually achieved through a series of threaded rods. Various types of connections, from laminated bushings to T-bolts, are employed. Given that composite materials like FRP are considered less reliable than metals like steel, each new blade design undergoes rigorous testing to ensure its stability under extreme conditions.

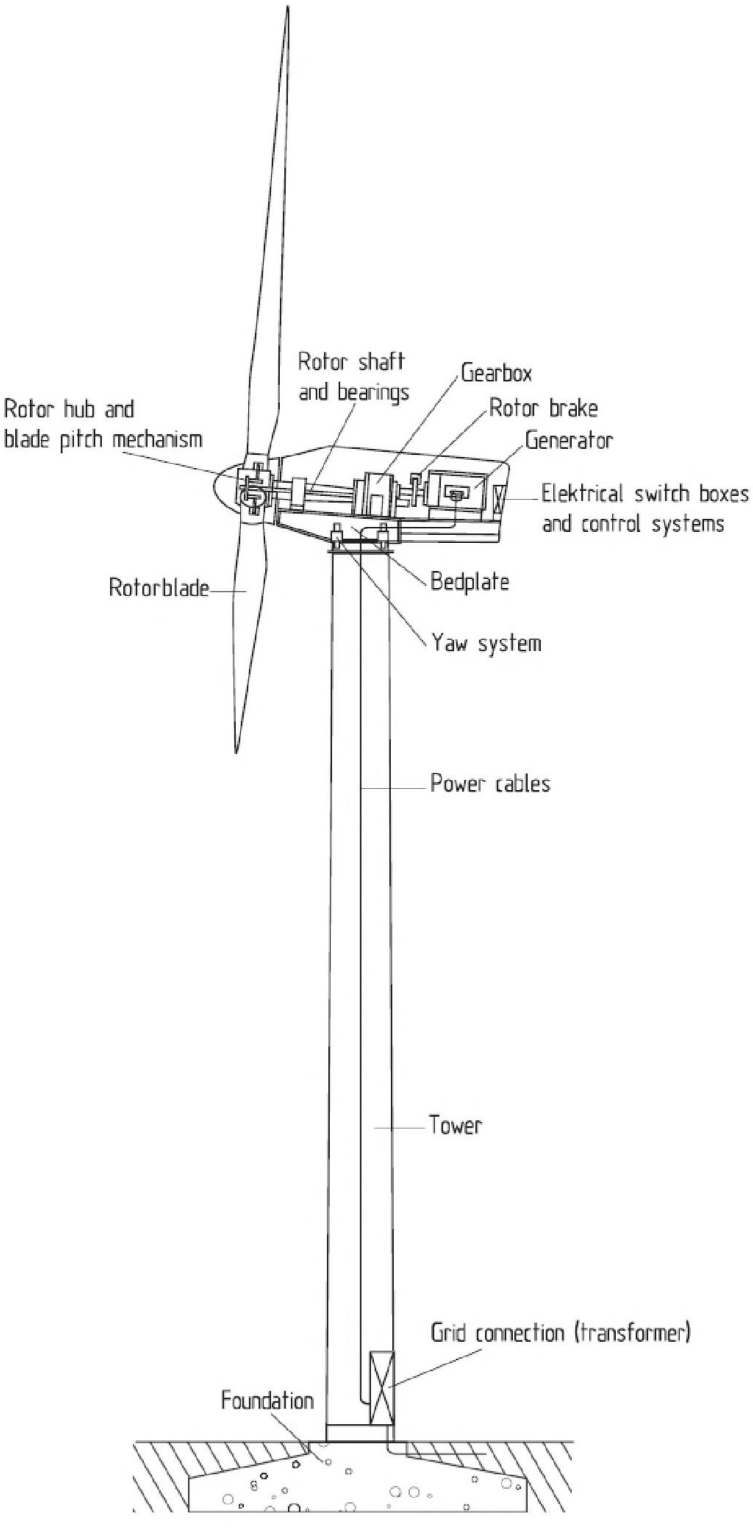


Figure 2.2: Common components in a Horizontal Axis Wind Turbine. Figure taken from [84]

2.3.4 Tower

Wind turbines are typically built using one of three main types of towers: tubular steel towers, lattice towers, and concrete towers. In general, a large diameter of the tower is desirable as it provides the highest stability. This is especially true at the bottom of the tower where bending moments are largest. In consequence, many towers are – at least in parts – conical. The heavy weights and large dimensions make the tower one of the most expensive components in a wind turbine.

Tubular steel towers are the most widely used and are constructed from steel plates that are bent and welded together. For onshore turbines however, they are restricted in size as their parts must be transported on roads. This effectively limits the onshore hub-height on a tubular steel tower to 140 m. No such restriction exists for offshore construction. Another steel-based option are lattice towers. They consist of main beams connected by a lattice of thinner steel beams. These towers are easier to transport and offer flexibility in terms of size, but their assembly is time-consuming. Moreover, many find them aesthetically unpleasing, limiting their popularity. Concrete towers offer a third alternative. Their on-site construction lessens the transport restrictions. However, while concrete comes with the ability to withstand large pressure loads, it can only endure very limited tensile stress. This can be overcome by pre-stressing the tower, so that the concrete always remains under pressure no matter the current bending moment. A popular option for new large scale installations on land is a hybrid version with a concrete base and a tubular steel top.

2.3.5 Drive Train and Electrical Generation

Converting the mechanical energy harnessed from the wind to electricity is key for modern wind turbines. The process of this conversion occurs between the generator's rotor and stator. The turning rotor induces voltage in the stator coils mediated via the magnet field. Two primary types of generators are commonly used in wind turbines: synchronous and induction (also called asynchronous) machines. Both produce 3-phase Alternating Current (AC) but differ in their rotor circuitry. The terms "synchronous" or "asynchronous" refer to whether the rotor speed matches that of the rotating magnetic field in the stator. In synchronous generators, these speeds are the same, while in asynchronous generators, they differ slightly allowing more adaptation of the rotational speed.

As the frequency of the produced AC is dependent on the rotational speed, this speed must be fixed by the frequency of the electrical grid, if the turbine is directly connected to it. While this was the case in older turbines, modern machines overcome this constraint by the use of converters. These converters enable variable-speed operation, allowing the wind turbine to operate at speeds that maximize power output (cf. Sec. 2.3.6). The idea is to use two converters in sequence: the rectifier and the inverter. In a fully-fed setup the total AC current from the generator is converted into Direct Current (DC) by the rectifier. The inverter then turns it back into AC but at the grid's frequency. This allows seamless integration of the generated electricity into the grid.

Another popular variant is known as the Doubly-Fed Induction Generator (DFIG).

It also allows for variable speed but requires only a partial-rating (usually about 30%) converter, often making it a more economical choice. The stator of such a generator is directly connected to the electrical grid, while the rotor current's frequency is adjusted independently through the converter. This allows the wind turbine to operate within a certain speed range, albeit narrower than that of fully-fed systems. The analyzed turbines in, for example, Ch. 4 are of the DFIG-type.

The final necessary part is a transformer. It increases the voltage from the generator level (around 700 V) to higher voltages suitable for long-distance transmission (around 30 kV), minimizing energy loss during transportation without the need for massive cable cross-sections. Converters and transformers are sometimes found at the base of the wind turbine tower for easier maintenance. However, this necessitates large cable cross-sections in the tower. The trend therefore leans toward integration of these components into the nacelle.

2.3.6 Control Systems

Modern wind turbines are highly automated machines, which can handle various operating conditions on their own. They automatically align with the wind directions, regulate rotor speed and adjust the pitch angle of the blades. The detailed control mechanisms are complicated, differ between turbine types and are often kept secret by the manufacturer. However, the general concepts are the same among all modern utility-scale turbines.

Below rated power – the maximum power output, for which the turbine is designed – the aim is to extract as much power from the wind as possible. Here, variable speed control is a crucial feature that enables optimal power extraction from varying wind speeds in modern turbines. The fundamental concept relies on the dependency of the power coefficient c_P on the tip speed ratio λ. For optimal performance, the wind turbine aims to operate at a specific design tip speed ratio λ_D, where the power coefficient is at its maximum. To achieve this, the turbine uses torque control, which involves measuring the rotational speed of the rotor and setting the generator's torque accordingly. If the rotor's rotational speed leads to a tip speed ratio smaller (higher) than λ_D, the aerodynamic torque will be higher (lower) than the generator's counteracting torque, causing the rotor to speed up (slow down). This ensures self-regulating optimized power generation. This process is known as maximum power point tracking. It allows the turbine to follow a pre-defined optimal power curve, even under variable wind conditions. The system is flexible enough to allow for minor deviations from this optimal curve due to technical or physical limitations, such as system inertia or wind turbulence.

Above rated wind speed, i.e. the wind speed at which the rated power can be produced, it is important to limit the power production to avoid damages. The common method used in almost all current utility-scale turbines is pitch control. The blades are rotated along their axis to change the angle of attack of the wind. Thereby, the local aerodynamic forces, namely lift and drag coefficients, and consequently the effective power coefficient c_p of the rotor are changed. Usually the dynamic control of the pitch angle is based on the rotor speed. The deviation between measured rotor speed and the desired rated rotor speed are used to calculate the necessary changes in pitch angle. This allows the system to keep the produced power almost constant. Another – nowadays less used – concept to limit the power output is through stall control. Here, the blade's design favors stall effects, i.e. the onset of turbulence at the rotor blades, at high wind speeds. While this effectively reduces the torque, the thrust is not decreased. Hence, this control approach is unsuited for large turbines as their structures would become too expensive.

As mentioned above, the individual control of a turbine usually employs more complex variants of these basic ideas. Additionally, for an individual turbine design slight changes to this general control concept are always possible. For example, in Ch. 4 we show that the analyzed turbine has an intermediate state, where power still increases with higher wind speeds, but the rotation is kept constant for optimal performance of the generator. Furthermore, the objective of a controller might be changed in certain situations. For example, restrictions in the electricity grid sometimes make curtailment, i.e. the reduction of the maximum power produced by a turbine, necessary.

2.3.7 Wind Farms

Today, many turbines do not operate on their own, but in so called farms, which consist of many turbines. This allows operation of multiple turbines with only one connection into the main electrical grid. Furthermore, maintenance measures can be pooled, which reduces costs.

However, many aspects must be considered when designing a wind farm, among them also the complicated task of wind farm layout [85]. Each turbine influences the wind stream downwind of itself creating wakes with lesser wind speeds. If another turbine is hit by such a wake, its power production is reduced. Therefore, the setup of turbines inside a farm must be carefully considered, even using very detailed fluid dynamics simulations to find the best layout considering dominant wind directions [24, 86]. Furthermore, farm control strategies are being researched. Here, the goal is not to maximize power output of individual turbines, but often to maximize power output from the total farm [22]. This can be done by slightly rotating the turbines in such a way that their wakes steer clear of other turbines, but this of course means that they are not perfectly aligned with the wind direction. Other objectives of farm control include minimizing mechanical loads and complying with grid requirements [87].

2.3.8 New Concepts

New concepts and technologies are still being developed to improve efficiency, reduce costs, and address specific challenges in wind power generation. Airborne wind energy involves using tethered flying devices to capture wind energy at higher altitudes. Offshore floating concepts explore the use of floating platforms to harness wind energy in deep waters. Smart rotors incorporate sensors and actuators to optimize the performance of wind turbine blades. Unconventional power transmission systems aim to improve the efficiency of transmitting power from wind turbines to the grid. Multi-rotor turbines utilize multiple smaller rotors instead of a single

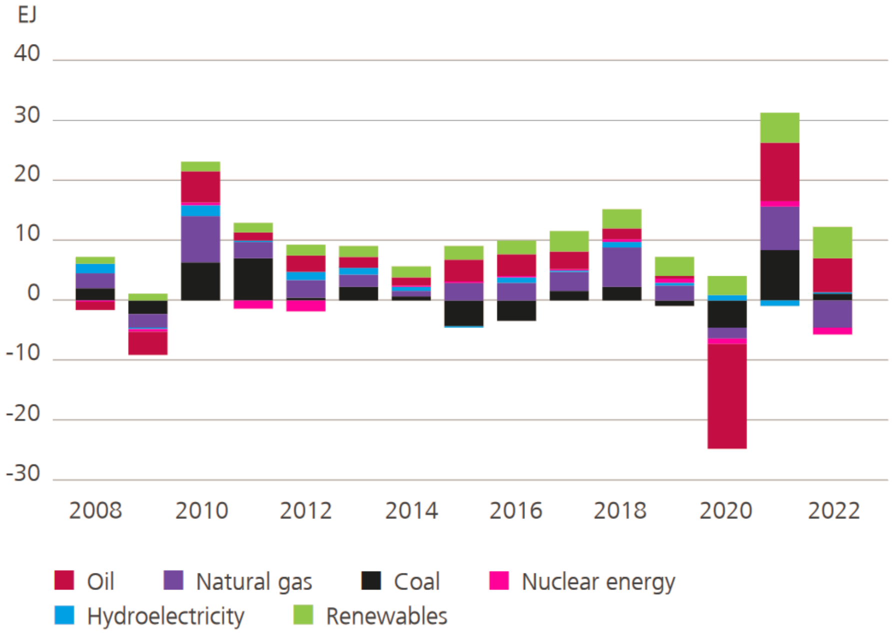


Figure 2.3: The worldwide change in primary energy consumption by fuel. Figure taken from [10]

large rotor to increase efficiency and reduce loads. Alternative support structures explore new designs and materials for wind turbine towers and foundations. Innovative blade manufacturing techniques, such as automated manufacturing and additive 3D printing, offer cost reduction and improved performance. Wind-induced energy harvesting devices capture energy from aeroelastic phenomena, such as vibrations and flutter, to generate electricity. These technologies are at various stages of development with interesting future outlooks. Most still require further research and funding to realize their full potential. For a more detailed review on the prospects of these new technologies we refer to Watson et al. [88].

2.4 Significance of Renewable Energy

In recent years the reality of climate change has become ever more apparent with an increasing amount of natural disasters and weather extremes. Science has proven humanity's influence in this [9, 38]. Thus, the focus on renewable sources of energy has surged, becoming a global priority amid an ever-growing global demand for energy. Already 174 countries have set renewable energy targets, with 37 aiming for 100% renewable energy in their electricity [12].

While global primary energy consumption saw a modest 1% increase in 2022, with fossil fuels accounting for 82% of primary energy consumption [10], the data in Fig. 2.3 clearly shows that the increase in energy consumption is steady (2020 marking the exception caused by the COVID-19 pandemic) with Renewables gaining more shares. Green energy is making gradual but significant inroads, reaching a 14.2% share in the primary energy sector in 2022 [10].

Focusing on electrical power generation, the system is undergoing similar changes. Global electricity generation rose by 2.3% in 2022, with wind and solar achieving a record high with a 12% share in global electricity generation [10]. This upswing in renewable energy is further supported by the will to become independent from fossil imports in many countries and the increase in their prices. Natural gas prices reached record levels in Europe and the Asia Pacific regions in 2022 [10]. Renewable energy technologies, particularly solar and wind, are scaling up at an impressive pace. In 2021 84% of total capacity additions to worldwide electricity production were from Renewables, emphasizing their increasing importance [89]. The following year brought newly installed renewable energy capacity of 348 GW, another 13% more than in 2021. About 70% and 22% of this new capacity came from solar and wind power, respectively [12]. Already in 2020 eleven countries met over 20% of their electricity demands through wind energy and photovoltaics [90]. Globally, the share of of Renewables in the energy mix reached 12% in 2022 while some countries exceed this by far such as Denmark with 61%, Uruguay with 36% or Germany with 32% [12]. Two years earlier during the COVID-19 pandemic Germany produced more electricity from renewable sources (44%) than from all fossil energy sources (40%) for the first time [90]. Here, wind power is the leading electricity producer at 27% of the electricity mix, 78% of which was offshore wind [90]. Globally, the total wind energy capacity in 2022 stood at 899 GW, with Europe hosting 252 GW and Germany taking the lead in Europe with 66 GW. The recorded growth of electricity generation by wind power in 2022 was 13.5% [10].

Despite these positive trends, the challenge to achieve a sustainable energy future is daunting. The International Energy Agency's Net-Zero plan by 2050 demands increases in the annually installed capacity [91]. To achieve net-zero emissions by 2050, newly installed renewable capacity needs to be 2.5 times larger by 2030 than it was in 2022. Specifically for wind power, annual additions must increase by 3.7 times by then.

In conclusion, given the urgent need to reduce greenhouse gas emissions in the face of climate change, wind energy stands out as one important player among other renewable energy sources in shaping a sustainable future. Wind turbines have already become a significant source of renewable energy[92], but further improvement and research of wind energy technology is paramount for meeting global energy demands in a sustainable manner [13, 14].

2.5 Operational Challenges

While the Levelized Cost Of Energy (LCOE) of eletricity from wind energy is competitive [93], operation and maintenance of wind turbines is challenging. Maldonado-Correa et al. [18] list various authors claiming that Operation and Maintenance (O&M) costs account for 20-35% of the total expenditure for offshore wind farms. The corresponding numbers are lower but still significant (approx. 10-15 %) for onshore turbines [21, 94]. Thus, the advantage of steady and high wind speeds at offshore locations comes at the cost of harsh environmental conditions and increased difficulties for O&M [95]. For example, offshore turbines cannot be reached every day for maintenance measures due to unsuitable weather conditions.

Not surprisingly, among the many topics to be studied in the wind turbine field [13, 14], O&M is on focus for researchers and industry alike. They undertake increased efforts to effectively optimize O&M procedures for wind turbines and thereby reduce this cost factor [96–101]. Improved understanding of wind turbine behavior is key to achieving this goal. Data driven methods are developed to control problems such as, for example, yaw misalignment or under-performance [102, 103]. In general, increasing the reliability of turbines is important [104]. However, failures or problems will always occur. Hence, another prominent topic is the prediction of failures in wind turbines with sufficient lead time to react and carry out preemptive maintenance instead of correctional maintenance. This reduces not only the money lost in turbine downtime, but also enables cheaper maintenance. The idea is to optimize assets by replacing components exactly when needed [105]. The wind energy branch follows a general trend in most industries, aiming at moving from scheduled maintenance towards condition-based maintenance to reduce costs and efforts [106].

2.6 WiSAbigdata Project

Clearly, improving operation and maintenance of wind turbines through the use of available operational data is critical in optimizing the performance of wind energy systems. The WiSAbigdata (Wind farm virtual Site Assistant for O&M decision support — advanced methods for big data analysis) project, in which this study was carried out, tackles this problem. Researchers and industry have partnered up to develop an assistance system based on enhanced data analysis methods. The goal is to leverage the data potential with existing and new methods to provide useful information for the daily operations of wind farms. One focus of the project lies on predictive analysis. This means the capability to predict turbine behavior – and especially turbine failures and faults – before they happen. If such a system could be reliably employed, it would strongly benefit the planning of maintenance activities.

Overall, the project aims at a better and thorough understanding of the (statistical) information contained in high frequency wind turbine data. Various approaches to big data analysis are tested by an interdisciplinary team and directly evaluated with industry.

2.7 Data

A plethora of data are gathered in modern wind turbines. A Supervisory Control and Data Acquisition (SCADA) system is installed in all major wind turbines since commission. It is comprised of many sensors gathering various information from wind speed to electrical measurements as well as component temperatures. The exact set of observables contained in the SCADA data are turbine dependent, but Tab. 2.1 lists common data channels. Usually, the set of observables is primarily defined by what is necessary for turbine control. However, singular turbines, e.g. for research purposes, can be equipped with many further sensors [107]. Additionally, many operators choose to install Condition Monitoring Systems (CMS) in their turbines. These are typically multiple vibration sensors placed on different parts of the turbine. Their signals are used to monitor the turbine and, if possible, predict failures. However, such additional sensors have their own inherent costs, which can be problematic if the goal is to reduce the O&M costs. They will not be considered in this study, but we recommend Stetco et. al [108] for a review on approaches using such CMS data.

The different data serve multiple purposes. In operation they are necessary for the controller of the wind turbine to ensure optimal operating conditions (cf. Sec. 2.3.6). On a larger scale, aggregated data from multiple turbines can be used to make informed decisions about energy storage and grid distribution [109]. Furthermore, remote monitoring and assessment of the turbine becomes possible. Recorded historical data are used to analyze and optimize the operation of wind turbines.

While many observables are measured with high frequency, some even multiple times in one second, the recorded historical SCADA data is usually averaged over 10 minute intervals. Some further statistical measures, such as standard deviation in the 10 minute interval, are often recorded as well. Many developed methods try to employ them for different types of analysis. The reader is referred to Maldonado-Correa et al.[18] and Tautz-Weinert et al.[110] for reviews. Common methodologies include neural networks, physical models and statistical analysis [99, 111–118]. These authors also raise two important points: First, it is often complicated or impossible to reliably label events in the data due to scarcity of available log and maintenance data. Second, Ulmer et al. [98], who apply convolutional neural networks for failure detection, mention that the 10 minute averaging process naturally leads to a loss of information. This effect is specifically studied by Beretta et al. [119]. Some researchers have tried to avoid these problems by using simulated high frequency data [116, 120].

For this study we have, as part of the WiSAbigdata project, access to high frequency SCADA data from three industry partners. It was immediately clear that, while all are SCADA data, they are very different. The recorded observables, the recording frequency, the accuracy of measurements and the data format are inhomogeneous. Sometimes the recording frequency is even different within one dataset. For example, temperatures, which are considered to be rather slowly changing observables, are measured a lot less frequently than other observables. The inhomogeneity between different data sets becomes even more pronounced for the alarm logs and maintenance data, which were also provided by the industry partners. These consist of automated alarms as well as service reports written by personnel. Standardization of operational and maintenance data would greatly improve usability in science and allow easier testing of developed methods. We employed a semi-automatic process to transform the service data into a combination of Zustands-Ereignis-Ursachen-Schlüssel (Condition-Event-Cause-Key – ZEUS) [121] and Reference Designation System for Power Plants (RDS-PP) [122]. The ZEUS-Code allows us to understand how the state of the turbine (or a single component) is influenced and what maintenance measure was carried out. The RDS-PP-Code specifies which component was afflicted.

While this event data is present, its usage is problematic. First, the translation into the ZEUS- and RDS-PP-Code is not always sufficiently clear. Some service reports do not contain the necessary information for this sort of classification. Sec-

Table 2.1: Common SCADA data observables as stated by Tautz & Weinert [110]

Environmental	Electrical characteristics	Part temperatures	Control variables
wind speed	active power	gearbox bearing	pitch angle
wind direction	power factor	gearbox lubricant oil	yaw angle
ambient temperature	reactive power	generator winding	rotor shaft speed
nacelle temperature	generator voltages	main bearing	generator speed
	generator phase current	rotor shaft	fan speed / status
	voltage frequency	generator shaft	cooling pump status
		generator slip ring	number of yaw movements
		inverter phase	set pitch angle / deviation
		converter cooling water	number of starts / stops
		transformer phase	operational status code
		hub controller	
		top controller	
		converter controller	
		grid busbar	

ond, the present data are of limited use in establishing when exactly a turbine is faulty or not. One easily understood example for this problem is the following: The replacement of a subsystem indicates that this component was faulty, but it is entirely unclear since when the component showed faulty behavior. Hence, even with operator provided event data, the ground truth needed for the development and evaluation of analysis methods – especially failure detection and prediction methods – is hard to establish. The service documentation kept by the industry rather cover economic points than those needed from a research viewpoint. Such problems are also identified by Tavner [94].

A detailed description of the explicitly used data in any analysis is given in the corresponding chapters.

CHAPTER 3

General Theory

While specific theory and calculations are introduced in the chapters when needed, some general concepts are useful for the understanding of all chapters and their connections in this thesis. We introduce these in this chapter.

First, we will discuss the way our data is structured. Then, the Pearson correlation, as a measure for the linear dependency of two different observables, is introduced. It will play a central role in this thesis, especially in the form of correlation matrices, which measure multiple dependencies between observables. Furthermore, we introduce the eigenvalue analysis of these matrices and the connection to Principal Component Analysis (PCA).

3.1 Time Series Data

The data used for analysis in this thesis are in the form of time series. Such data consists of a sequence of numerical data points indexed in successive order. Time series data can arise from a wide range of scientific disciplines such as physics, economics, biology, and engineering, among others. The main objective of their analysis is to uncover the underlying structure of the data, which could consist of trend components, inter-dependencies, seasonality, and noise. The mathematical framework for analyzing time series often involves stochastic processes, statistical tests, correlations and various forms of regression models. Understanding the characteristics of a time series is critical for both descriptive and predictive analysis of many systems. [123]

For this thesis, an observable (or signal) k, for example wind speed or active power output, is measured at multiple discrete times $t = 1, \dots, T_{\text{end}}$. We assume that the times are given as unit free steps. This is possible as the interval between measurements is constant. With multiple measured signals $k = 1, \dots, K$ each time series is denoted as a sequence of data points

$$X_k(t),\ k = 1, \dots, K\ ,\ t = 1, \dots, T_{\text{end}}. \tag{3.1}$$

To analyze this data in an aggregated form, we arrange multiple time series into rectangular $K \times T_{\text{end}}$ data matrices

$$X = \begin{bmatrix} X_1(1) & \dots & X_1(T_{\text{end}}) \\ \vdots & & \vdots \\ X_k(1) & \ddots & X_k(T_{\text{end}}) \\ \vdots & & \vdots \\ X_K(1) & \dots & X_K(T_{\text{end}}) \end{bmatrix}, \tag{3.2}$$

where each row is the time series of observable k.

Quite often the different time series consist of values that range on inhomogeneous scales. This makes an objective comparison between different signals k difficult. To circumvent this problem it is useful to normalize each time series to a mean value of zero and a standard deviation of one. This allows us to treat all observables on equal footing. The mean value of a time series is calculated as

$$\mu_k = \frac{1}{T_{\text{end}}} \sum_{t=1}^{T_{\text{end}}} X_k(t) \tag{3.3}$$

and the standard deviation as

$$\sigma_k = \sqrt{\frac{1}{T_{\text{end}}} \sum_{t=1}^{T_{\text{end}}} (X_k(t) - \mu_k)^2} \ . \tag{3.4}$$

Then, the normalized time series $M_k(t)$ are given by

$$M_k(t) = \frac{X_k(t) - \mu_k}{\sigma_k}, \ k = 1, \dots, K. \tag{3.5}$$

Analogous to Eq. (3.2) we define the normalized data matrix

$$M = \begin{bmatrix} M_1(1) & \dots & M_1(T_{\text{end}}) \\ \vdots & & \vdots \\ M_k(1) & \ddots & M_k(T_{\text{end}}) \\ \vdots & & \vdots \\ M_K(1) & \dots & M_K(T_{\text{end}}) \end{bmatrix} . \tag{3.6}$$

While for Chs. 4 to 6 the different k denote different measurements at one turbine, in Ch. 8 they differentiate between turbines in one wind farm. In the general analysis of an arbitrary system in Ch. 7 no explicit meaning is given to k except for the presented traffic data example, where k denotes different street sections.

3.2 Pearson Correlation Matrices

For the analysis of multiple time series, i.e. $K > 1$, the mutual dependencies represent essential information. One linear measure of such dependencies is the

Pearson correlation coefficient. It provides information on the relationship between different observables and plays a central role in this thesis. Using the normalized time series for two observables k and l it is defined as

$$C_{kl} = \frac{1}{T_{\text{end}}} \sum_{t=1}^{T_{\text{end}}} M_k(t) M_l(t). \tag{3.7}$$

The values of C_{kl} lie in the range between -1 and $+1$. Positive values indicate that the two observables tend to move in the same direction, whereas negative values indicate that they tend to move in opposite directions. If $C_{kl} = 1$, the two time series of observables k and l are completely correlated, i.e. they are identical under normalization. On the other hand, $C_{kl} = -1$ indicates complete anti-correlation. Two time series are uncorrelated (in terms of the Pearson correlation), if the coefficient is equal to zero.

An aggregated view onto multiple time series is provided by the Pearson correlation matrix C. The elements of the matrix are the Pearson correlation coefficients C_{kl}. It can be calculated using the normalized $K \times T_{\text{end}}$ data matrices defined in Eq. (3.6). The time average of Eq. (3.7) is realized as a matrix product

$$C = \frac{1}{T_{\text{end}}} M M^{\dagger}, \tag{3.8}$$

where $M^{\dagger}$ denotes the transpose of M. This $K \times K$ matrix C is positive-semidefinite and real symmetric. It contains the correlation coefficients between all K signals and the diagonal values are one by definition.

Another closely related measure is the covariance matrix Σ. Instead of the correlation coefficients it contains the variances of each time series on the diagonal and the covariances

$$\Sigma_{kl} = \frac{1}{T_{\text{end}}} \sum_{t=1}^{T_{\text{end}}} X_k(t) X_l(t) - \mu_k \mu_l \tag{3.9}$$

in the off-diagonal elements. It is related to the correlation matrix by

$$\Sigma = \sigma C \sigma, \tag{3.10}$$

where $\sigma = \text{diag}(\sigma_1, \ldots, \sigma_K)$ is the diagonal matrix of the standard deviations.

Figures 3.1 and 3.2 contain examples of correlation matrices. In Fig. 3.1 we see matrices calculated for one wind turbine over time periods of 30 minutes. Figure 3.2 shows correlations between all wind turbines in a wind farm for the measured active power output over time periods of 12 hours.

Sometimes the use of the Pearson correlation matrix is problematic. Its entries are averaged over times, which might include trends or other variations. This non-stationarity could lead to estimation errors. In general, many complex systems, for example financial markets [42] and traffic systems [51], exhibit non-stationarity in the correlation matrix. With the varying outside conditions and different control strategies it is plausible to expect the correlations of wind turbine SCADA data to be non-stationary as well. This fact is also already visible in the example correlation matrices in Figs. 3.1 and 3.2. One way to deal with the average over such non-stationarities is to calculate the correlation matrices in intervals instead of for the

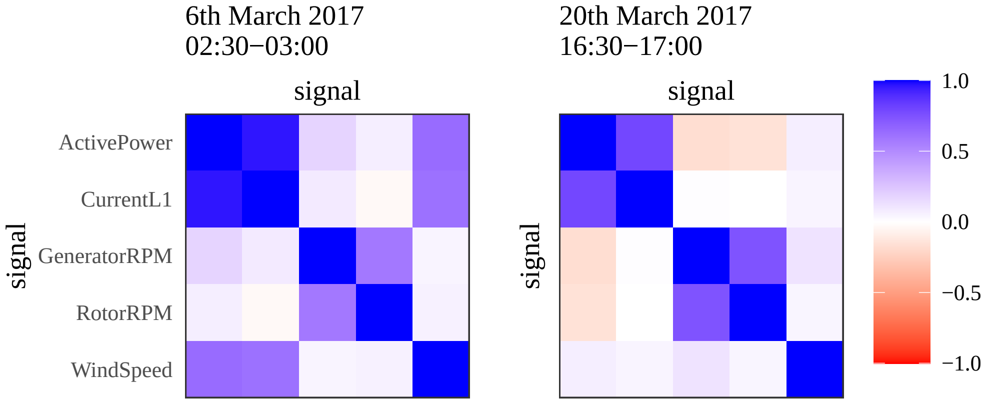


Figure 3.1: Two example correlation matrices for one turbine with multiple measured signals. The signals are the same on the x- and y-axis: generated active power (ActivePower), generated current (CurrentL1), rotation per minute of the rotor (RotorRPM), rotation per minute of the high speed shaft at the generator and wind speed (WindSpeed). The value of the matrix entries, i.e. the Pearson correlation coefficients, is indicated by the color. Each matrix is calculated over the indicated 30 minute interval.

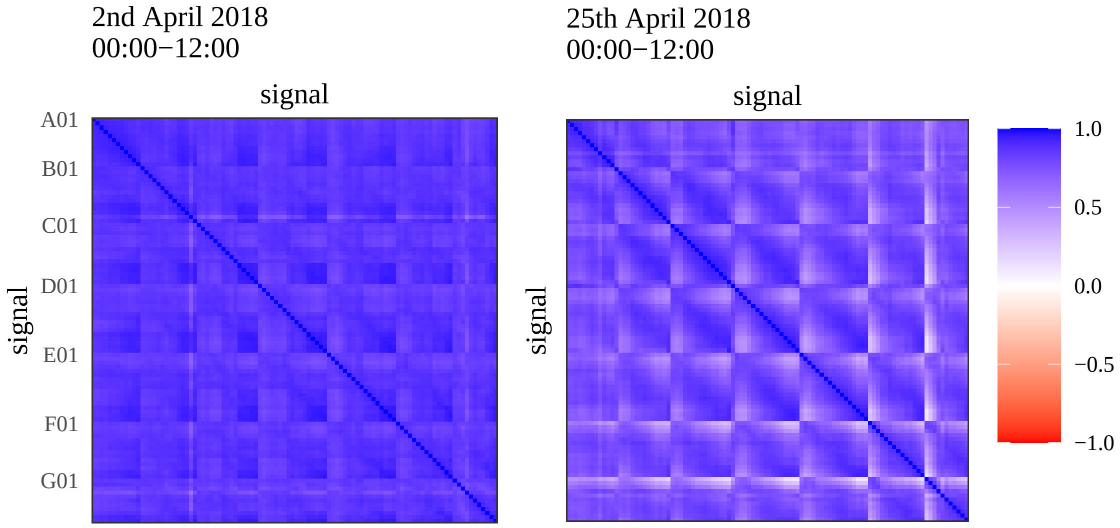


Figure 3.2: Two example correlation matrices for the active power output of all turbines in the Thanet wind farm. The x- and y-axis are the same and display the different turbines. We list only the first (01) turbine of each row (indicated by a letters A-G) for readability. The value of the matrix entries, i.e. the Pearson correlation coefficients, is indicated by the color. Each matrix is calculated over the indicated half day interval.

whole time series. This also allows us to study how the correlation matrix changes in time. We refer to these intervals as epochs and label them with the new time variable τ equal to the starting time of the epoch. Their length is denoted by T. The total time series is then divided into T_{end}/T non-overlapping intervals with a possibly occurring rest at the end being omitted. For each of these intervals, the normalization is done separately with mean value $\mu(\tau)$ and $\sigma(\tau)$ calculated during the epoch. This allows us to calculate a time series of correlation matrices

$$C(\tau),\ \tau = 1, T+1, 2T+1, \ldots, (T_{\text{end}}/T - 1)T + 1. \tag{3.11}$$

The correlation coefficients $C_{kl}(\tau)$ in these matrices then signify the correlation strength between the observables k and l only in the epoch τ, i.e. for time $t \geq \tau$ and $t < \tau + T$. We analyze the non-stationarity of wind turbine and wind farm correlation matrices in chapters 4, 5 and 8.

When choosing the length T of the epochs, we have to compromise. To best distinguish non-stationarities in the correlation structure of a system, the epoch length should be short. As wind changes constantly, even on the time scale of seconds and minutes, this holds true for wind turbine data. However, the finite length of the time series during an epoch introduces measurement noise to the correlation coefficients. Considering K different time series, the correlation matrix contains $K(K-1)/2$ entries. They are calculated from K time series of length T. With smaller T the calculated correlations become noisy. If T is smaller than K the correlation matrix will be singular. In conclusion, there is a trade-off between choosing T large enough to keep noise at low levels and small enough to resolve non-stationarities. Such compromises are common when dealing with correlation matrix time series [43, 124].

Another important fact to keep in mind, when analyzing Pearson correlation matrices, is that they only contain information on linear dependencies. In fact, it is a good correlation measure only for elliptical distributions such as the multivariate normal distribution. A common measure, that also includes non-linear dependencies is Spearman's ranked correlation. However, while non-linear dependencies definitely exist in wind turbine data, for example the power is dependent on the third power of wind speed according to Eq. (2.1), we find that the Pearson correlation measure reveals much about the existing dependencies and yields good results during our analyses. Nonetheless, some calculations from Ch. 4 are repeated with Spearman's ranked correlation in App. B, but no significant differences are found.

3.3 Eigenvalue Analysis

In the context of correlation matrices, it is often useful to analyze their eigenvalues $\lambda_j,\ j = 1, \ldots, K$ and eigenvectors $\boldsymbol{v}_j,\ j = 1, \ldots, K$. It is important to note that the numbering is not in the same order, i.e. $j \neq k$. Only the number of possible values K is equal for both. The eigenvalues and eigenvectors are defined via

$$C\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j, \tag{3.12}$$

i.e. an eigenvector of C does not change its direction under multiplication with C, only its length is modified by the corresponding eigenvalue.

As the correlation matrix is real symmetric and positive-semidefinite, the eigenvalues must be non-negative real numbers. Therefore, we can sort them and assume $\lambda_1 \geq \cdots \geq \lambda_K$. The sum of eigenvalues is

$$\sum_{j=1}^{K} \lambda_j = K. \tag{3.13}$$

This is because the trace of a matrix (sum of its diagonal elements) remains invariant under a change of basis, and the eigenvalues represent the variances of the observables in the new basis defined by the eigenvectors. Employing the eigenvectors as a basis is possible as they must be orthogonal as long as the eigenvalues are non-degenerate, because C is real symmetric. Furthermore, they can be normalized to length one, because eigenvectors are only defined up to a multiplicative factor. Hence, we can assume that the set of $\boldsymbol{v}_j,\ j = 1, \ldots, K$ is orthonormal:

$$\boldsymbol{v}_i^T \boldsymbol{v}_j = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j. \end{cases} \tag{3.14}$$

This poses an obvious and interesting question: Does the basis defined by the eigenvectors have a special meaning?

In a classical physics interpretation, the eigenvalues of the correlation matrix can be thought of as revealing the 'modes' of the system. In the context of vibrational modes of a molecule, for example, the eigenvectors indicate the direction and manner in which atoms vibrate, and the eigenvalues provide the associated energy or strength of that vibration. From this example, we already learn one important fact about the eigenvector basis: The size of the eigenvalue indicates the importance or strength of the behavior identified by the corresponding eigenvector for a given system.

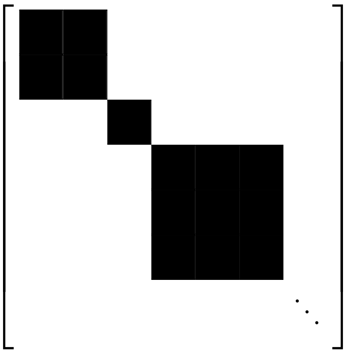
Now, we look at what the eigenvectors and eigenvalues would look like in a random system, i.e. one where all mutual dependencies are random and not determined by systematic coupling of any sort. In such a system, we do not expect any dominant behaviors of the total system. We find such examples in Random Matrix Theory (RMT). RMT is the study of large matrices, i.e. large K, with random variables as entries. After such matrices were introduced by by John Wishart in multivariate statistics in 1928 [125], RMT became popular in nuclear physics when Eugene Wigner used it to explain the energy levels of complex nuclei in the 1950s [126]. Nowadays, RMT is also applied in many other scientific disciplines, for example quantum physics [127, 128] and statistics [129]. In finance, as an example of a complex system, RMT has become an important tool [44, 130, 131]. In the context of our question about the meaning of the eigenvector basis, we can learn from RMT what the eigenvalue density would look like, if our system was random. Then, for large K and T_{end} the density of eigenvalues of a correlation matrix $C = \frac{1}{T_{\text{end}}} AA^{\dagger}$ calculated from random $K \times T_{\text{end}}$ data matrices A with Gaussian-distributed random entries is given by the Marchenko-Pastur distribution

$$\rho(\lambda) = \frac{\beta}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{3.15}$$

where $\beta = T_{\text{end}}/K$, σ^2 is the variance of the random entries in A and $\lambda_{\pm} = \sigma^2(\beta^{-\frac{1}{2}} \pm 1)^2$ [132]. Hence, for a random system, the eigenvalues would all lie in a bulk of small values. In real world complex systems, for example finance, we often find that such a bulk exists, but that there are some eigenvalues which are significantly larger. Sticking with our finance example, we usually observe one very large eigenvalue, which is connected to an eigenvector with homogeneous entries, that fluctuate only little. This describes a behavior, where all measured observables – usually price changes in finance – tend to move in the same direction. In finance this is a collective market mode [133, 134]. For wind farms we find in Ch. 8 that such a farm mode also exists, when comparing measured active power values from all turbines in a wind farm. Other large eigenvalues, i.e. larger than λ_+, which are smaller than the market eigenvalue, usually signify branches in financial stocks. For general systems, this translates to groups of observables that behave similarly.

The influence of group correlated behavior on the eigenvalues and eigenvectors can be analyzed by employing factor models. Guhr & Kälber [135] employed a 1-factor model to construct time series, which are purely group correlated, i.e. their correlation matrix has non-zero values only on diagonal blocks:

$C =$



, with white indicating zero and black non-zero. (3.16)

Correlations exist within groups of observables, but no correlations exist outside of these groups in the limit of large T_{end}. Here, they find that for each group in the observables one large eigenvalue exists with a value proportional to the size of the group. The matching eigenvector has homogeneous values for all entries corresponding to the group and zero otherwise. Further eigenvectors with small eigenvalues then represent relative motion inside of these groups.

In summary, a correlated system of large but finite T often shows two types of eigenvalues and eigenvectors. Large eigenvalues that, together with the corresponding eigenvectors, represent common behavioral patterns of the observables and small eigenvalues that represent (random) fluctuations and less common behavioral patterns. If one's interest lies with the detection of anomalies or failures in a system, it might be that the small eigenvalues become the important ones. The separation of common and uncommon behavior is also exploited in Principal Component Analysis (PCA), which provides us with another viewpoint on the meaning of correlation matrix eigenvalues and eigenvectors in the following section.

3.4 Principal Component Analysis

Principal Component Analysis (PCA) is a widely used tool in the context of statistics, machine learning and data science. Its primary objective is dimensionality reduction and feature extraction. It was first developed by Karl Pearson in 1901 [136] – a first clue to its close connection to the Pearson correlation matrix. In the 1930s Harold Hotelling developed it independently and named it [137, 138]. Today, it is a common technique in many disciplines [139, 140].

The basic idea is to transform a set of possibly correlated observables into a new set of observables called the principal components. These new observables are chosen in such a way that the first one is the direction of maximum variance in the K-dimensional data set. The second one is chosen the same way with the additional condition that it must be orthogonal to the first one. This continues on, so that each new component lies in the direction of maximum variance under the condition of orthogonality with all previously chosen components. We can imagine PCA as the fitting of a K-dimensional ellipsoid to the data set comprised of our K observables. The axes of the ellipsoid then form the principal components. In mathematical terms this is equivalent to the calculation of the eigenvectors of the covariance matrix, if the data is centralized to mean value zero in each observable before the fitting of the ellipsoid. If the data are also normalized to one before the fitting, the principal components are equivalent to the eigenvectors of the correlation matrix. Both approaches are common and the choice depends on the analyzed observables. As the different measurements at wind turbines range on very different scales, we use the second, correlation matrix related approach, unless stated otherwise.

The first principal component is the same as the eigenvector to the largest eigenvalue. As mentioned in Sec. 3.3, the eigenvalues represent the variance of the data along the directions specified by the eigenvectors. Therefore, due to $\lambda_1 \geq \cdots \geq \lambda_K$ the sorting of principal components and eigenvectors of the correlation matrix is equivalent.

When performing PCA, the original data is projected into the basis of the principal components. Each normalized data point

$$\boldsymbol{M}(t) = \begin{bmatrix} M_1(t) \\ \vdots \\ M_k(t) \\ \vdots \\ M_K(t) \end{bmatrix}, \tag{3.17}$$

which is a single column of the normalized data matrix in Eq. (3.6), is projected onto the eigenvectors $\boldsymbol{v}_j$:

$$M'_j(t) = \boldsymbol{v}_j^T \boldsymbol{M}(t), \quad j = 1, \ldots, K. \tag{3.18}$$

If one then keeps all K projections, no information is lost and we have simply performed a coordinate transformation into the basis of the eigenvectors. However, the interpretation in PCA is that axes with large variances represent the normal, dominant behavior of the system. Hence, for dimensionality reduction not all K

projections are kept, but only the first few. The number of kept dimensions is usually based on the total explained variance, i.e. one chooses the number of principal components that together explain, for example, 95% of the total variance in the data. Here, we have a more direct interpretation of the eigenvectors of the correlation matrix, which is nonetheless very similar to the one we found in the previous section. Eigenvectors with large eigenvalues represent normal, dominant behavior in a correlated system. In consequence, eigenvectors with small eigenvalues, referred to as minor components, represent abnormal, non-occurring behavior.

We have already discussed that the correlation matrix is often non-stationary in complex systems in Sec. 3.2. Hence, its eigenvectors must also change in time. Therefore, in systems with non-stationarity one has to be careful when applying PCA. This serves as a good example of how non-stationarity influences a common method, which usually assumes stationarity. After characterizing non-stationarity in wind turbines, we will study how knowledge about it improves stochastic process analysis and PCA to some extent in Ch. 6 and Ch. 7, respectively.

3.4.1 Variants

Many variants of PCA exist. They were developed to tackle different specific problems. Their detailed understanding is not necessary for this thesis, but some better known ones include static and dynamic robust PCA [141, 142], kernel PCA [143], dynamic kernel PCA [144], moving window PCA [145], recursive PCA [146] and incremental PCA [147]. The latter three tackle the important problem of non-stationarity. We refer the reader to the review by Ketelaere et al. [148] for an introduction to PCA on time-dependent data. A performance comparison for system monitoring was undertaken by Rato et al. [149].

3.4.2 Change Detection with Principal Components

An important use case for PCA is novelty, change or failure detection, which is also one of the goals of the WiSAbigdata project, see Sec. 2.6. PCA might be applied in this context for its original purpose: dimensionality reduction in a data set before the application of an arbitrary other method. It might also be applied more directly. The original, non-changed data is projected into the system of principal components to calibrate what values are normal for these projections. New data is then projected and compared to the normal space using, for example, Hotelling T^2-statistics or the sum of squared prediction error together with a threshold [116, 150]. It has been applied to change detection in many different systems, including wind turbines [117], wastewater treatment plants [151], healthcare institutions [152] and traffic [153]. An important question is which of the principal components to look at. For the use-case of dimensionality reduction one generally employs a projection onto the major components [140]. However, the residual subspace, i.e. the minor components, is often most useful for outlier and change detection [154–156]. This underlines our earlier conclusion that abnormal behavior is represented in the minor components. Shyu et al. [157] proposed a method combining the components relating to the largest and the smallest eigenvalues for novelty detection. Tveten [158] studied the sensitivity of different principal components in a more structured and complete way confirming the high sensitivity of minor components to changes. In Ch. 7 we analyze how the presence of non-stationarity in a system influences this sensitivity.

CHAPTER 4

Operational States in Wind Turbine Correlation Matrices

A big challenge for wind turbine analysis and monitoring is presented by the varying external (e.g. wind speed, temperature) and internal (e.g. turbine control, curtailment) operation conditions. We studied this in paper [1], of which this chapter is a modified version. Non-stationarity is important also in many applications aside from wind turbines [147, 159–163]. We show in Ch. 7 that non-stationarity in correlated systems is important for the detection of anomalies. For wind turbines, it has been shown to have an effect on failure detection [164, 165]. Furthermore, different states in frequency data measured by a CMS system have been identified due to operational regimes [166]. Different behavior of the SCADA data for such regimes is to be expected also due to the turbine control mechanisms [96]. In general, complex systems containing many different observables, mechanisms and external influences show non-stationarity in their cross-correlations. The stability of correlations in financial stock market data was analyzed, for example, by Buccheri et al. [167]. Münnix et al. [42] shows that the correlation matrix of this data inhabits different states over time by using cluster analysis. The stability of these states was further analyzed by Rinn et al. [168] and Stepanov et al. [169]. Similar studies were also done for traffic data [51]. Some general correlation analysis for wind turbine data was carried out by Braun et al. [170].

In this chapter, we aim to quantify the non-stationarity in correlation matrices for high frequency SCADA data from real wind turbines during normal operations. To this end, we apply cluster analysis to the correlation matrices of different SCADA signals calculated over 30 minute time intervals. Distinct states with significantly different structures in the correlations matrices are found. We show that the prime cause for this is the turbine's own control mechanism. This allows us to develop a criterion based on wind speed to separate the cluster states. Such an automated distinction would, in principle, enable the usage of multiple normal states in applications via pre-processing. It is an important step towards accounting for the non-stationarity due to the operational states in an analysis such as failure detection with principal components.

This chapter is structured as follows: In Sec. 4.1 we will introduce the dataset we work with before moving on to the theoretical background of clustering correlation matrices in Sec. 4.2. Then we will present our clustering results for a single turbine in Sec. 4.3. We find proof of non-stationarity and identify the turbine control mechanism as the prime influence. Afterwards, we model the boundary wind speeds between the states in Sec. 4.4. Finally, we will show that the established method works for multiple turbines without further problems and can even be improved by the increase in available data in Sec. 4.5. We present our conclusions in Sec. 4.6.

4.1 Data

Our dataset includes 100 Vestas V90-3.0 MW wind turbines from the Thanet offshore wind farm south-east of Great Britain. It contains observables that are measured at approximately 5 second time intervals. To obtain aligned, synchronized data, we average the time series on 10 second time intervals resulting in a data frequency $\nu = 1/10\,\mathrm{s}$. This ensures continuity in the data even when the actual measurement frequency fluctuates around 5 s. This does not hinder the calculation of correlations for our purposes. In fact, it is rather similar to any measurement process: Every sensor will in reality average over a short time span to obtain a value. Taking the mean over 10 s seems therefore more natural than, for example, using only the last value from each interval. If at some time the deviation from the 5 s interval becomes stronger or if there are actually missing data points, data points will be missing in our averaged data as well. This might occur if there was a problem with, for example, sensors or communications. Another reason is simply a decreased measurement frequency when the turbine is switched off. This is underlined by the majority of missing values occurring in the very low wind speed regime beneath the turn-on wind speed. As the turbines are not running during these times, this wind range is not of interest for us. These missing values do not pose a problem for our analysis. For these reasons, we decided to transfer missing values from the original dataset into the 10 s data instead of replacing them by other means.

We are interested in identifying changes in the correlation structure, which emerge while the turbine is operating normally, in contrast to changes caused by failures. Therefore, in our main analysis we look at the following basic observables:

- generated active power (ActivePower)
- generated current (CurrentL1)
- rotation per minute of the rotor (RotorRPM)
- rotation per minute of the high speed shaft at the generator (GeneratorRPM)
- wind speed (WindSpeed)

As there are no deviations between the three current phases in our data, we simply choose one of them. The presented observables provide a good picture of the main turbine systems. Wind speed makes the rotor move. Its rotation is transmitted via gears to the rotation of the high speed shaft at the generator. This, in turn, generates electrical current and power. Two pairs of strongly correlated observables exist in this set. Deviations between generated active power and current could only occur, if large amounts of reactive power are generated. The low and high speed rotation of rotor and generator are directly coupled as well. These expected results are confirmed during our analysis. We include these pairs in our analysis as examples for group structures in the correlations. Knowledge of such structures is indispensable for monitoring a complex system: Are the groups stable? Do they break up? Do correlations across groups exist? In the presence of anomalies, such correlations, which are deemed normal and obvious, might be the structures which break up. Grouping is, in general, an important aspect in the study of complex systems [43, 51, 171–173].

While measurements of temperatures are very common and useful for failure analysis [18, 110, 111], they are rather slowly changing observables. Hence, in our data they are not measured at high frequency and without decimals, which makes the calculation of short-term correlations impossible. Furthermore, it seems reasonable to assume that their behavior in normal states is strongly coupled to mechanical observables, e.g. higher rotation speeds will lead to increased bearing temperatures. Of course, they would also be influenced by, for example, seasonality or cooling mechanisms. Thus, while excluded from the study at hand, further analysis of temperature correlations is nevertheless desirable for future work.

Two additional important control observables are the pitch angle of the rotor blades (BladePitchAngle) and the ratio between the blade tip speed and the current wind speed (TipSpeedRatio). Our data does not contain three separate measurements for the blades, but only one. We assume this to be a mean of the three individual blades. The pitch angle is excluded in our main analysis due to many missing values that hinder the calculation of correlation matrices. The tip speed ratio is not directly present in the data, but results from easy linear calculation

$$\text{TipSpeedRatio} = \frac{2\pi\text{RotorRPM} \cdot \text{RotorRadius}}{\text{WindSpeed}}. \tag{4.1}$$

It is disregarded in the main analysis, because we study linear correlations and it is also linearly derived from two already present observables. As both omissions are prominent observables when studying wind turbines, we include additional results with consideration of the pitch angle of the blades and the tip speed ratio for the basic cluster analysis in Sec. 4.3. To do this we had to fill the missing values for the pitch angle. A possible explanation for the missing values is a data acquisition system that only writes values whenever a new measurement is different from an old measurement, i.e. if the value changed. We could not establish whether this is actually the case in our data, but it seems reasonable. Thus, for the additional results including the pitch angle observable we treated the data as if this assumption were true. This means we filled any missing values in the 10 s data with the last measured value before. Of course, thereby we also fill in any values that might actually be missing instead of being left out for data storage reasons. Furthermore, with the pitch angle being a rather slowly changing observable compared to the other observables, the filling sometimes leads to stable values over long time periods. This, of course, hinders once again the calculation of short-term correlations. Therefore, one must be cautious when considering these additional results and we did not consider pitch angle or tip speed ratio directly for the analysis following the basic clustering in the present work.

We used approximately three weeks of data from 5 March 2017 to 24 March 2017. The data from such a time span are still easy enough to handle while providing enough data points to obtain reasonable clustering results. In view of possible practical applications, three weeks is a short enough time span to make it easily usable. There would be no need to collect huge amounts of operational data beforehand. However, it is of course necessary that the data used for identifying different operational states covers all possible states. In practice it turns out later that this means, we need a wide range of wind speeds in our data. The actual time span was chosen, because for at least one wind turbine there are no manual or automatic alarms or services during this period (cf. Sec. 4.3). Two turbines have no recorded data for this time span, effectively reducing our data set to 98 turbines.

Due to confidentiality agreements we will never show absolute values of any observable. In fact, only wind speed is shown directly and is then presented in units of the nominal wind speed $\tilde{u}_{\mathrm{nom}}$ at which the turbine starts to produce its nominal power output according to the manufacturer. The tilde is introduced to mark it as the rated value provided by the manufacturer as we will later on try to infer this value also from the data.

As we are looking at K different observables per turbine and also L different turbines, we introduce an additional identifier l to the notation of Sec. 3.1 to indicate the turbine. Each of the measured signals k for each turbine l yields a time series of data points $X_k^{(l)}(t)$, $k = 1, \ldots, K$, $l = 1, \ldots, L$, $t = 1, \ldots, T_{\mathrm{end}}$. In the case of our dataset we have $K = 5$, $L = 98$ and - assuming complete data - $T_{\mathrm{end}} = 20 \cdot 24 \cdot 60 \cdot 6 = 172800$. The data is arranged into L rectangular $K \times T_{\mathrm{end}}$ data matrices $X^{(l)}$ according to Eq. (3.2).

4.2 Theoretical Background

Our analysis to distinguish different states is based on identifying differences in the correlation matrix of observables listed in Sec. 4.1. In Sec. 4.2.1 we define the way in which we are calculating the correlation matrices.

To identify non-stationarity in the time series of these matrices we will use a distance measure and a clustering algorithm. These are introduced in Sec. 4.2.2.

4.2.1 Correlation Matrix Time Series

To identify changes over time in the correlation structure, the correlation matrices are calculated with a moving time window of 30 minute as explained in Sec. 3.2. The time intervals do not overlap. We effectively create a time series of matrices. To this end, the signal time series $X_k^{(l)}(t)$ are divided into disjoint intervals of 30 min, i.e. the lengths of the intervals is $T = 180$ in our dimensionless time variable. We refer to these intervals as epochs. Hence, we have $T_{end}/T = 960$ epochs. To avoid notational confusion, we use the new time variable τ for labeling the epochs. We reserve the notation $X_k^{(l)}(t)$ for the original time series and write $X^{(l)}(\tau)$ for the $K \times T$ data matrix containing the different time series for turbine l from τ to $\tau + T - 1$. The length of 30 minutes represents a compromise as already discussed in Sec. 3.2. Longer time spans would provide more data points per correlation matrix and would thereby decrease noise. However, we want to distinguish different states in time. Considering external conditions, e.g. wind, changing on short time scales of several minutes to hours, we have to choose relatively short epochs to ensure resolution of the non-stationarity.

During each of the epochs, we calculate the correlation matrix $C^{(l)}(\tau)$ for each turbine l according to the procedure described in Sec. 3.2. This matrix then contains information on the linear dependency of the different observables at turbine l during the epoch τ.

While the dependency of observables in a wind turbine is not always linear, which is already seen in the well-studied power curve, the linear Pearson correlation yields important and good results for the structure of the mutual dependencies. We have repeated our analysis with Spearman's rank correlation, which also measures non-linear dependencies, but did not find substantial differences. Results for the case with five observables are shown for comparison in App. B.

4.2.2 Clustering

We will now introduce the clustering, which allows us to sort the correlation matrices into groups (clusters) and check, whether different typical states do exist. If we can identify these, we will refer to them as OPERATIONAL STATES. An integer will be assigned to each of them and the algorithm will label each matrix in the time series with one such integer. Instead of a time series of correlation matrices, we then have a new integer time series $s(\tau)$ with the range $s \in \{1, \ldots, S\}$ when S is the number of clusters created.

The first outcome will then be as follows: If any decent clustering solution can be found, it is proof that typical states of the correlation structure exist. Then, analyzing the resulting integer time series $s(\tau)$ can much easier reveal dependencies of the state on time or other factors.

Any method separating objects into groups needs a distance measure defined between those objects. For the correlation matrices we choose the euclidean distance [43]. The reader can imagine that all matrix entries are written into a vector, effectively arranging the columns of the matrix underneath each other, so that the standard euclidean distance between vectors can be applied. The distance between the correlation matrices for the epochs starting at τ and τ' of turbine l is then

$$d^{(l)}(\tau, \tau') = \sqrt{\sum_{i,j} (C_{ij}^{(l)}(\tau) - C_{ij}^{(l)}(\tau'))^2} = ||C^{(l)}(\tau) - C^{(l)}(\tau')|| \ . \tag{4.2}$$

We choose the bisecting k-means algorithm to perform our clustering [174, 175]. This is the algorithm that was also used to determine states in the financial markets by [42] and [43]. It can be described as a hybrid of standard k-means clustering [176, 177] and hierarchical clustering [178]. While the former directly divides the whole set of objects into k groups (cf. App. A), the latter is performed step-wise. In each step either two groups are merged (agglomerative) or one group is divided into two (divisive). Bisecting k-means is a divisive clustering algorithm, meaning that at the start all objects belong to one big cluster and during each step one of the existing clusters is split into two new clusters. This bisection is performed by running a standard k-means on all objects within the group to be split with $k = 2$. Which of the $\tilde{S}$ currently existing clusters is split during a step, is decided based on the average internal distance of all objects in a cluster z_s , $s = 1, \ldots, \tilde{S}$

$$d_s^{(l)} = \frac{1}{|z_s^{(l)}|} \sum_{\tau \in z_s^{(l)}} ||C^{(l)}(\tau) - \langle C^{(l)} \rangle_s|| \quad . \tag{4.3}$$

Here, $|z_s^{(l)}|$ denotes the number of objects in cluster $z_s^{(l)}$ and $\langle C^{(l)} \rangle_s$ is the centroid of cluster $z_s^{(l)}$ defined by the element-wise mean:

$$\langle C_{ij}^{(l)} \rangle_s = \frac{1}{|z_s^{(l)}|} \sum_{\tau \in z_s^{(l)}} C_{ij}^{(l)} \quad . \tag{4.4}$$

Each step the cluster with the largest average internal distance is bisected. The algorithm is terminated when a set number S of clusters is reached.

This is slightly different from the approach used in former works by [42], [168], [169] and [43], where a threshold is introduced and all clusters are bisected until no single existing cluster has an average internal distance larger than the threshold. However, the threshold is then set based on the number of clusters wanted. It can easily be understood that the resulting clustering will be the same if one either uses our approach to produce S clusters or chooses the threshold in such a way that S clusters are produced.

Applying this algorithm, we split the set $Z^{(l)} = \{C^{(l)}(1), C^{(l)}(1+T), \ldots, C^{(l)}(T_{\text{end}} - T)\}$ of all correlation matrices into S subsets $\{z_1^{(l)}, \ldots, z_s^{(l)}, \ldots, z_S^{(l)}\}$. The centroid of each cluster according to Eq. (4.4) is interpreted as the mean correlation matrix of a cluster representing its typical correlation structure. Thereby, we only need to look at S matrices and a series of T_{end}/T integers $s(\tau)$ instead of as many matrices.

Later on we will see that the emerging typical correlation matrices correspond to different control settings of the turbines. This explains in a simplified way, why the hierarchical k-means works better than a normal k-means in our case. Approximately, we can describe the controller of a wind turbine as a mechanism fixing certain signals to a fixed value. This means the correlation of that signal with other non-fixed signals should vanish. The divisive clustering will first extract a group where signal A might be fixed. Then this group might be further divided into subgroups where signal B is either fixed or not. And in theory this could go on. Such a problem is very well suited for divisive clustering.

In order to check, if our clustering is sensible, we will do two things. Firstly, we will just look at the cluster centroids and see, if we can interpret them and if they are substantially different from each other. Secondly, we will calculate silhouette coefficients [179]

$$\zeta^{(l)}(\tau) = \begin{cases} \frac{b(\tau) - a(\tau)}{\max(b(\tau), a(\tau))} & , |z_{s(\tau)}| > 1 \\ 0 & , |z_{s(\tau)}| = 1 \end{cases} \tag{4.5}$$

with the average distance to all other matrices in the same cluster

$$a(\tau) = \frac{1}{|z_{s(\tau)}| - 1} \sum_{\tau' \in z_{s(\tau)}, \tau' \neq \tau} d^{(l)}(\tau, \tau') \tag{4.6}$$

and the smallest average distance to a single other cluster

$$b(\tau) = \min_{r \neq s(\tau)} \left(\frac{1}{|z_r|} \sum_{\tau' \in z_r} d^{(l)}(\tau, \tau') \right) \ . \tag{4.7}$$

This coefficient will take values between -1 and $+1$ with larger positive values representing matrices that are well clustered and negative values showing matrices that are closer to another cluster than to their own. To get an indicator for the overall clustering we will use the average silhouette coefficient

$$\bar{\zeta}^{(l)} = \frac{T}{T_{\text{end}}} \sum_{\tau} \zeta^{(l)}(\tau) \ . \tag{4.8}$$

4.3 State Identification via Clustering

In the following, we analyze the correlation matrix time series of one turbine, which will henceforth be referred to as turbine 1 (WT1). We are singling out this wind turbine, because the time frame of the total dataset was selected in such a way that for WT1 no problems were listed in the automatic alarm logs and manual service reports. The idea is that this will make analysis and definition of normal states easier as there was no (reported) unusual behavior.

The correlation matrices are calculated for non-overlapping epochs of 30 minutes each. This results in 960 matrices per turbine. However, due to several reasons there might be missing data in the time series. In such a case, any time stamp missing one or more of the measured observables is excluded from the calculation of the correlation matrix. Hereby, no estimation of values, which could influence the actual correlation coefficient, is necessary. More data could be used by calculating the correlation coefficients pair-wise, i.e. for any two observables just remove the time stamps where one of them is missing data. However, this does not result in a well defined positive semi-definite correlation matrix. For further analysis only those matrices, for which at least half of the expected data points (90 out of 180) exist, are considered. Furthermore, epochs in which the standard deviation $\sigma_k(\tau) = 0$ for any signal k have to be disregarded as they cannot be normalized.

We will first provide extensive results for the five considered observables in Sec. 4.3.1 and afterwards repeat some analysis including the pitch angle and tip speed ratio in Sec. 4.3.2.

4.3.1 Main Results for Five Observables

Without pitch angle no epoch includes a time series k, for which the standard deviation becomes zero. The disregarding of epochs with too many missing values is not a problem when looking at the five main observables as missing measurements

Table 4.1: Minimum, first quartile, median, mean, third quartile and maximum of silhouette coefficients for the clustering solutions with 2-5 clusters for correlation matrices of WT1.

clusters	min	1st Qu.	median	mean	3rd Qu.	max
2	-0.046	0.473	0.571	0.540	0.664	0.734
3	-0.157	0.379	0.537	0.491	0.640	0.718
4	-0.262	0.343	0.508	0.465	0.631	0.716
5	-0.352	0.314	0.479	0.439	0.626	0.707

usually stem from turbines being operational but switched off during times of very low wind speeds u smaller than turn-on wind speed u_{on}. For WT1 the average wind speed for 746 epochs with enough data is 10.01 ms^{-1}, while the average for 214 epochs where no correlation matrix could be calculated is only 4.34 ms^{-1}. It is obvious that these times where a wind turbine is not operating are unsuited for an identification of operational states. Of course, there might also be other reasons causing the missing data, e.g. a problem with the measurement of a signal. However, as for WT1 there are no alarms or services logged, we would not know what happened in those cases anyway. Any estimation of missing values would therefore need considerable guessing. As our results show that using only the epochs with enough data points is sufficient to reach a good differentiation of operational states, we are confident that just excluding missing values instead of estimating them is a good approach for our purposes.

When applying the hierarchical k-means algorithm described in the previous section to the set of matrices, the first step is to decide how many clusters provide a good solution. To this end, we calculate the silhouette coefficients for solutions with 2-5 clusters. The silhouette plots can be found in Fig. 4.1. The fifth cluster is almost imperceptible as it consists only of 3 matrices. Some descriptive statistics for these silhouette coefficients are shown in Tab. 4.1. The MEAN corresponds to the average silhouette coefficient from Eq. (4.8). All statistics provided decrease with increasing cluster number, implying that a few different states are sufficient to describe the dynamics of the analyzed correlation matrices. In the plots we can see some negative coefficients implying elements that would fit better into a different cluster. Such imperfections are to be expected when using heuristics like clustering algorithms. It is however clear, that all solutions provide a good grouping with largely positive silhouette coefficients. This is a strong indication that non-stationarities influence the correlation matrix. The influence is strong enough to be detected via simple clustering. Here, we observe that the assumption of a stationary correlation structure, for e.g. principal component analysis, is not justified.

As mentioned before, we also look at the cluster centroids to see if the matrices show indeed different behavior and if this distinction facilitates clear identification. Figure 4.2 shows the matrices calculated via Eq. (4.4) in a dendogram for the hierarchical clustering. The solutions for two and three clusters show distinctly different structures in the matrices, whereas the fourth cluster stems from cluster three, but is structurally very similar to cluster one, only differing in the strength of the mean

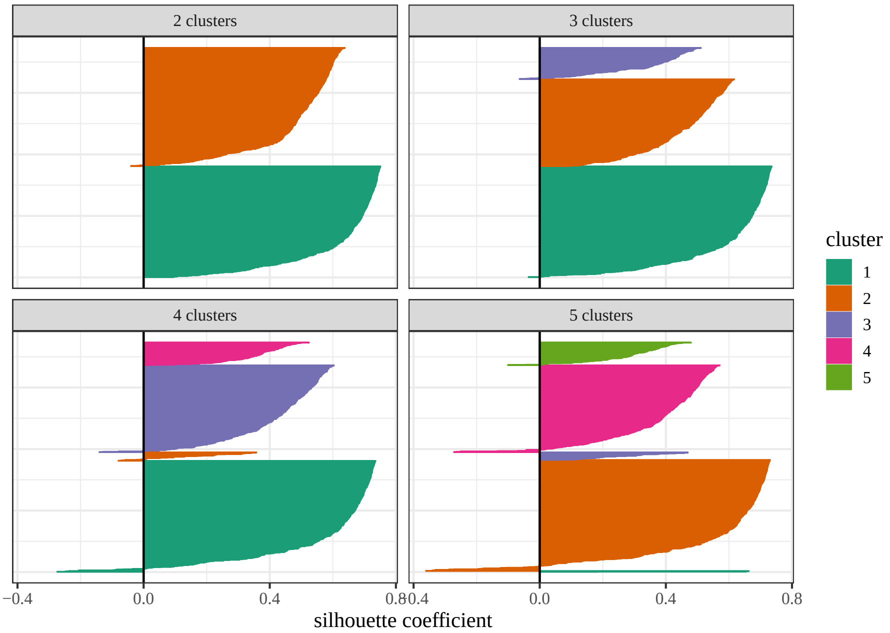


Figure 4.1: Silhouette plots for clustering solutions with 2-5 clusters. Each clustered element (matrix) is represented by a horizontal line the length of which is the silhouette coefficient for that element. Different clusters are color coded.

correlation. The introduction of a fifth cluster only produces a very small cluster with only three elements. The algorithm does not identify new groups, but rather starts to classify outliers. While the average silhouette coefficient favors two clusters, we continue our analysis with three clusters as we have seen that up to this point structural differences in the matrices occur and we will later see that these can be interpreted very reasonably. Here, we also point out that structural differences in the matrix have a stronger influence on the structure of the eigenvectors, i.e. principal components, than differences in average correlation strength. They are therefore more important to distinguish when using methods like principal components and Mahalanobis distance.

The classification of the matrices for three clusters is shown as an integer time series in Fig. 4.3. All three states appear to have a certain stability. Consecutive epochs often belong to the same cluster. However, there is no obvious behavior in dependency of the time. State 3 appears far less often than states 1 and 2. There is no emergence of new or disappearance of old states over time as is sometimes seen in other complex systems [42]. To get a better idea what each state might represent, we look at the matrices for the cluster centroids calculated according to Eq. (4.4) once more. They are seen in the third row of Fig. 4.2. Generally, as the differences between the matrices are quite clear we can conclude - in accordance with the silhouette coefficient - that the clustering does indeed separate the matrix time series into meaningful groups. The correlation matrices are non-stationary and automatically separable with a clustering algorithm.

In every cluster the strongest correlations are clearly visible between the observable pairs ActivePower-CurrentL1 and RotorRPM-GeneratorRPM. This was to be

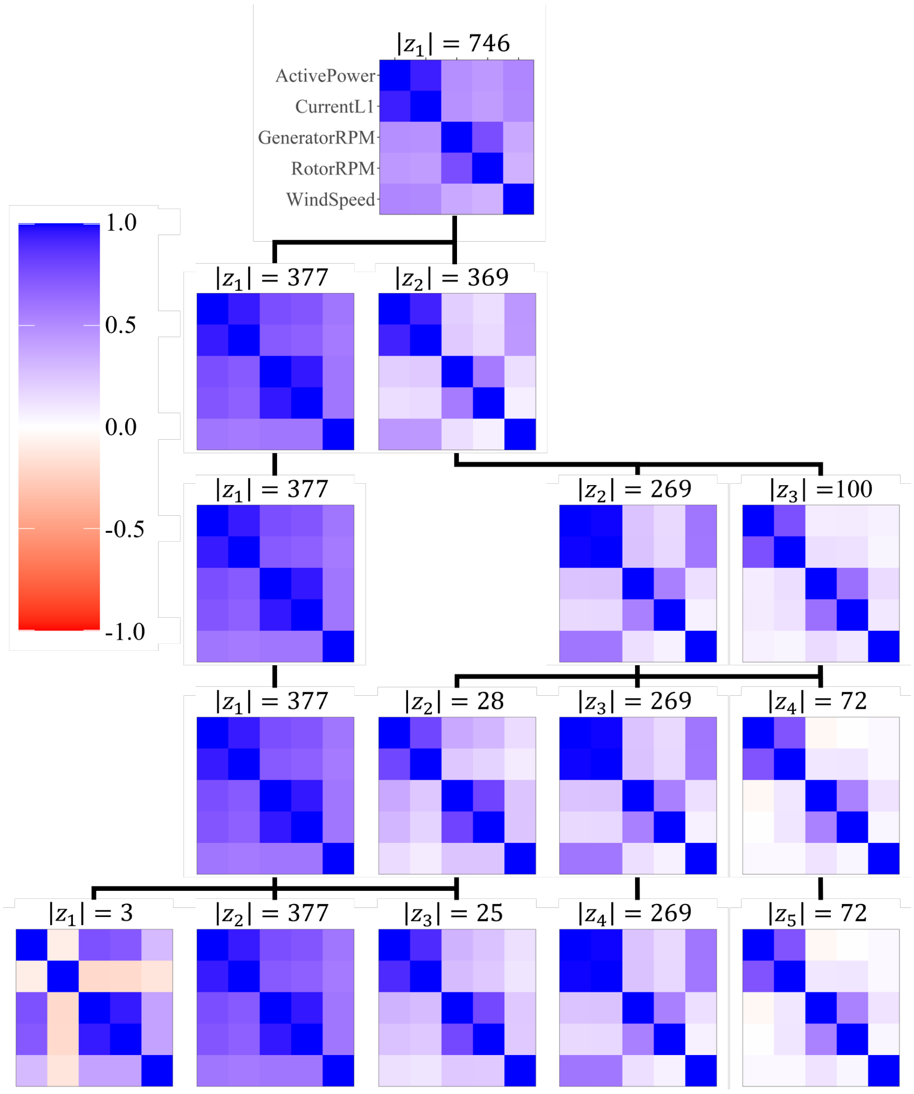


Figure 4.2: Cluster centroids as calculated in Eq. (4.4) for WT1 for different numbers of clusters. The color indicates the value of the correlation coefficient. Black lines connect child and parent clusters of the hierarchical algorithm and the number of cluster elements is given as $|z_s|$. Each cluster solution is ordered from low wind speeds (left) to high wind speeds (right) according to the average wind speed in a cluster.

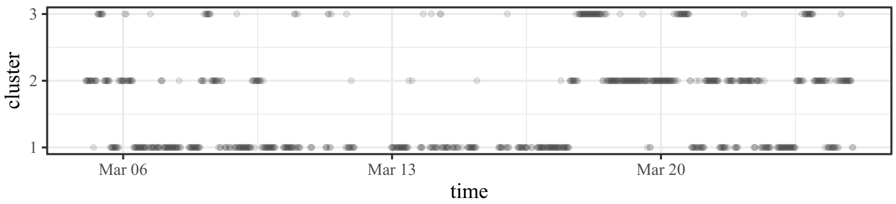


Figure 4.3: Cluster identifier s over time for WT1. Each dot represents a 30 minute epoch.

expected as these pairs are very directly linked. Apart from this we can see that for cluster 1 both of these pairs and the WindSpeed are all correlated with each other. Put differently, higher wind speed leads to faster rotation and thus to higher power. In cluster 2 this changes and the observables RotorRPM and GeneratorRPM, while still being closely correlated with each other, decouple from the other observables. Cross-correlations between these two and any other measurement vanish. The remaining cross-correlation between WindSpeed and the two observables ActivePower and CurrentL1 also vanish in cluster 3. Clearly, the three different states of the correlation structure identified via clustering are meaningful: They show distinctly different correlation behavior between different observables.

To interpret the meaning of the clustering solution, it is helpful to look at turbine control systems. The basic functionality of such a system is, for example, described by [180]. The specific functionality varies for individual turbine types, so it is likely that not all turbine types will show the same operational states. The turbine control system of the VESTAS V90 turbines analyzed here is one with variable pitch (VESTAS OPTITIP™) and variable speed (VESTAS OPTISPEED™). We can connect the three clusters to different operational states of this turbine type, which are separated by boundary wind speeds u_{on}, u_1, u_2 and u_{nom}. For very low wind speeds just above the turn-on wind speed $u_{\text{on}} \leq u < u_1$ the generator rotation is kept constant at the lowest possible value defined by the maximum slip in the generator. This results in a correlation structure as seen in cluster 2. Already for slightly higher wind speeds $u_1 \leq u < u_2$ the system controls the rpm proportional to the wind speed to operate at maximum aerodynamic efficiency of the rotor. This corresponds to cluster 1. With even more wind, but still not enough to reach nominal power output $u_2 \leq u < u_{\text{nom}}$ the turbine operates at fixed nominal rpm by controlling the torque. The rotational observables decouple again as for very low wind speeds and the correlation structure corresponds to cluster 2 again. Of course, fluctuations in wind speed will cause the rotation to fluctuate around the nominal value leading to some noise in the correlation structure. If wind speed is high enough to allow full power output $u_{\text{nom}} \leq u$ the nominal power output is reached and therefore kept constant alongside the rpm. This results in correlations as seen in cluster 3. All boundary wind speeds $u_{\text{on}}, u_1, u_2, u_{\text{nom}}$ are turbine dependent and usually not public knowledge.

We can see that our reasoning for the turbine at hand is correct by plotting the cluster state over the mean wind speed of the epoch instead of the time stamp in Fig. 4.4. For now disregarding the interval $u_{\text{on}} \leq u < u_1$ due to lack of enough data, cluster 1 represents low wind speeds, cluster 2 intermediate wind speeds and cluster

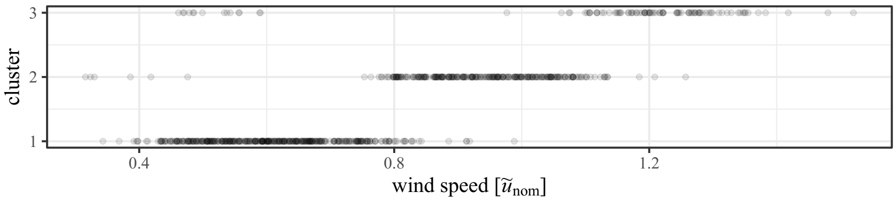


Figure 4.4: Cluster identifier s over wind speed for WT1. Each dot represents a 30 minute epoch.

3 high wind speeds, where nominal power output can be reached. Some exceptions are to be expected due to either wrong sorting during the clustering or wind speeds changing during the 30 minute epoch. Another, but probably less important factor is the finite response time of a wind turbine controller. It ranges in seconds or minutes and therefore the correlation structure does not respond instantly to fluctuations and changes in wind speed [181]. One example are the epochs sorted into cluster 3 whose wind speeds seem to lie in the range of cluster 1. They are all moved into the fourth cluster if we take another step in the algorithm. Its centroid is shown in the dendogram Fig. 4.2 and exhibits a structure very close to cluster 1. Such mismatches occur due to the heuristic nature of the algorithm and noise and fluctuations in the data, which results in matrices lying on the edge between two clusters. We will see in the following section that we can use the silhouette coefficient to identify them. One might also imagine high turbulence intensities, i.e. the standard deviation of wind speed divided by the mean wind speed in an epoch, leading to strange behavior in an epoch. We have tested filtering the epochs based on this turbulence intensity and did not find significant changes in the results. The dependency on wind speed is also in accordance with the stability in time as periods with stable wind speeds are common [182].

We have seen that the correlation matrix is non-stationary in time. The clustering has confirmed a primary influence of the control strategies in dependence of the wind speed. While the existence of different control regimes is not new, our analysis proves that they have strong influence on the structure of the correlation matrix. This automatic separation of states is a vital first step to account for non-stationarities when performing any analysis on high frequency SCADA data.

4.3.2 Additional Results with Pitch Angle and Tip Speed Ratio

The inclusion of the tip speed ratio does not affect the number of calculable epochs as it is directly derived from two other observables. Including the pitch angle observable, we only get 623 epochs, for which it is possible to calculate a correlation matrix. This is 123 epochs less than before. As missing values in the time series of pitch angle were filled as described in Sec. 4.1 this can only be due to standard deviations in the pitch angle being zero for the pitch angle data. We want to point out that this can be a direct result of the filling mechanism used for missing values. This goes to show that the results with pitch angle while being interesting have to

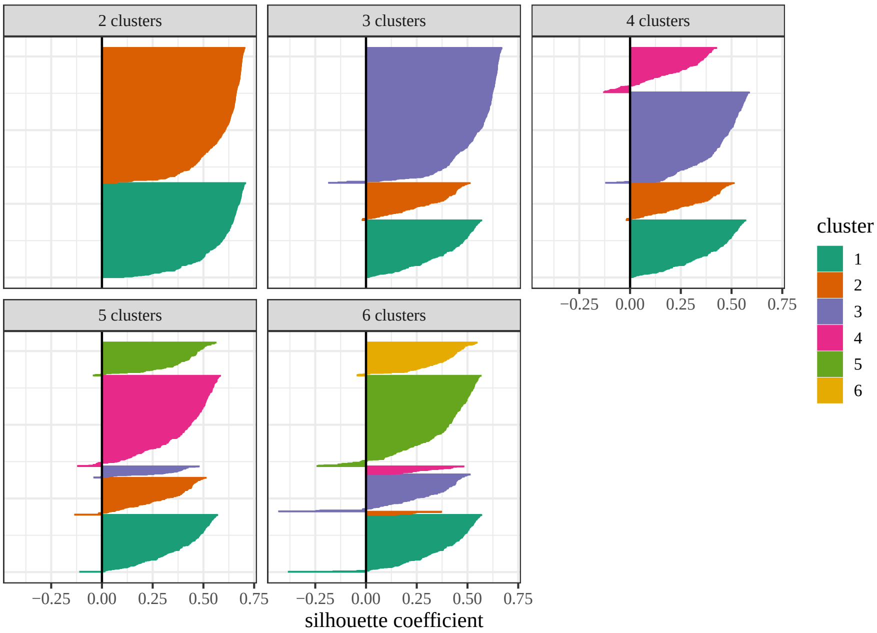


Figure 4.5: Silhouette plots for clustering solutions with 2-6 clusters with pitch angle. Each clustered element (matrix) is represented by a horizontal line the length of which is the silhouette coefficient for that element. Different clusters are color coded.

be treated with caution.

The calculation of matrices and the clustering are carried out in exactly the same way as before.

As we cannot assume that the number of relevant clusters stays the same when looking at a different set of observables, we look again at silhouettes in Fig. 4.5 and Tab. 4.2 and the cluster dendogram in Fig. 4.6.

The silhouette coefficients are again largely positive with some expected negative values from imperfect clustering heuristics. On average the values of the silhouette coefficients are smaller than in the analysis with only five observables. They still indicate a good grouping. The minimum and first quartile even increase in comparison to before, which indicates less poorly sorted matrices. A slight overall decrease

Table 4.2: Minimum, first quartile, median, mean, third quartile and maximum of silhouette coefficients for the clustering solutions with 2-6 clusters for correlation matrices of WT1 with pitch angle.

clusters	min	1st Qu.	median	mean	3rd Qu.	max
2	0.071	0.521	0.631	0.580	0.672	0.706
3	-0.183	0.380	0.500	0.461	0.609	0.668
4	-0.129	0.239	0.396	0.353	0.490	0.587
5	-0.133	0.270	0.418	0.366	0.495	0.582
6	-0.429	0.243	0.402	0.347	0.479	0.569

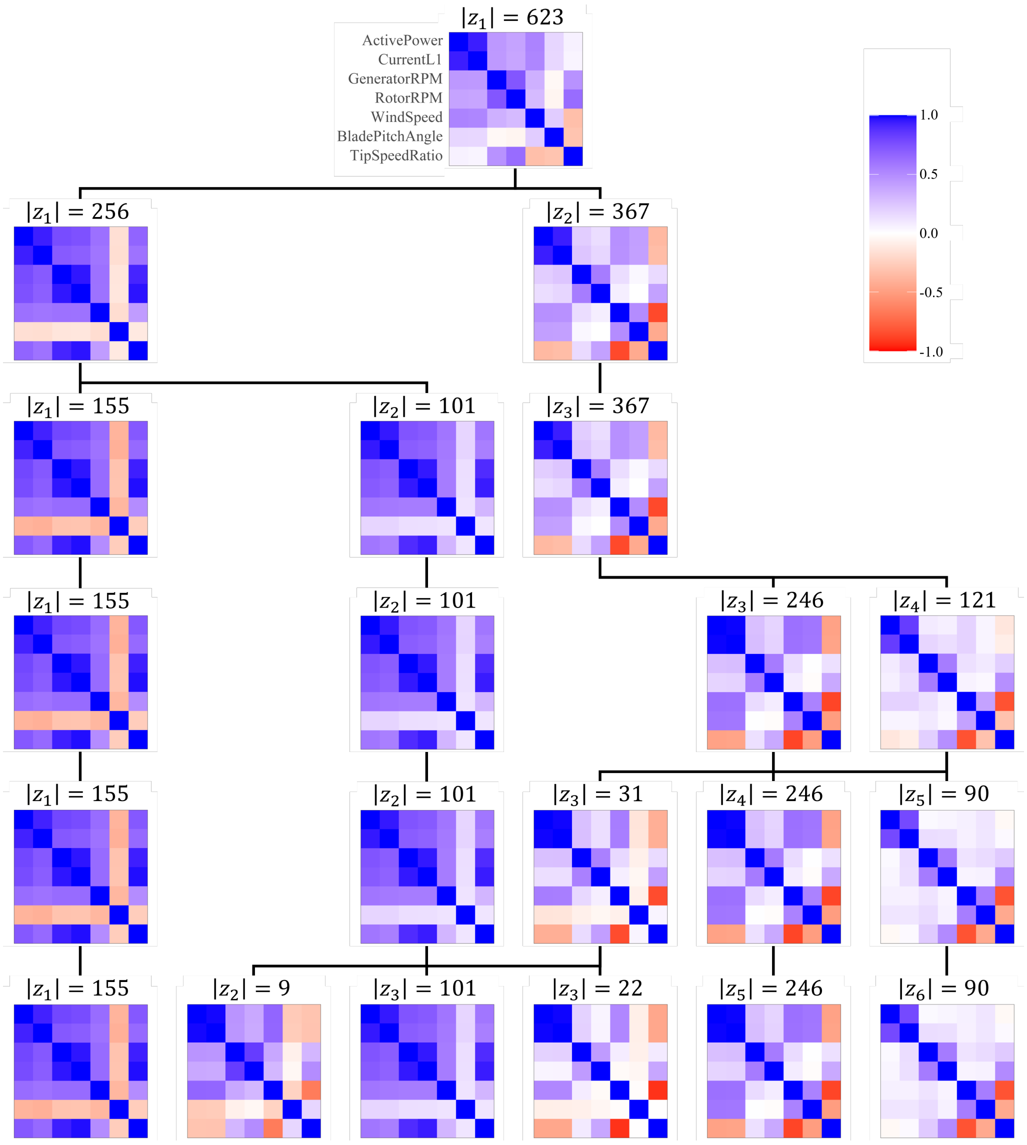


Figure 4.6: Cluster centroids as calculated in Eq. (4.4) for WT1 for different numbers of clusters with pitch angle. The color indicates the value of the correlation coefficient. Black lines connect child and parent clusters of the hierarchical algorithm and the number of cluster elements is given as $|z_s|$. Each cluster solution is ordered from low wind speeds (left) to high wind speeds (right) according to the average wind speed in a cluster.

in silhouettes is to be expected when clustering larger matrices as more pairs of single correlation coefficients need to be compared and each of them adds fluctuations. Once again, we see decreasing silhouette coefficients for larger numbers of clusters while still indicating that the grouping is reasonable.

Looking at the centroids in Fig. 4.6 we see that with four clusters the three cluster solution for with five observables has reemerged. Only cluster one from the previous solution is already split again because of the pitch angle. The numbering of clusters is done based on the average wind speed in each cluster, i.e. a low cluster number indicates low wind speeds.

Interestingly, with pitch angle and tip speed ratio considered we cannot stop at three clusters. The four and five cluster solution still show centroids that are structurally different. The sixths cluster distinguishes stronger and weaker values (mainly in the pitch angle) of the same type of structure and is already quite close to outsider classification with only nine inhabitants. Considering this as well as the sharply falling minimal value of silhouette coefficients from the five cluster solution to the one with six, we will take a closer look at five clusters. The cluster number over wind speed is shown in Fig. 4.7.

For low wind speeds the five main observables and the tip speed ratio are always correlated, while the pitch angle is either anti-correlated to all other observables or decouples from them. For the second case, it is likely that it simply stays constant in this regime as the intake of energy from the wind does not need to be reduced here. The anti-correlations might stem from a turbine being turned on underlined by wind speeds being slightly lower for cluster 1 than cluster 2. While it is turned off, large pitch angles are used to minimize strain on a still standing rotor and must then be reduced as wind speed increases above the cut-in point. It is not clear, however, if this effect is strong enough to produce anti-correlations for a period of 30 minutes. Cluster 3 of the five cluster solution appears to be an intermediate state, where the rotations are already decoupled from the rest of the system, but the pitch angle is not changed. Tip speed ratio is now anti-correlated to wind speed, active power and current as the rotation - and therefore the tip speed - does not increase any longer. Contrary to the average wind speed sorting cluster 3 shares a parent cluster with cluster 5 instead of 4. This might be because the decoupling of pitch angle from active power and current is a clearer distinction than the decoupling of the rotations from the rest. As the turbine controller tries to keep rotation constant, it still fluctuates creating some weak correlations whereas the pitch angle usually stays constant when it is decoupled from everything else. For intermediate wind speeds in cluster 4 the pitch angle is coupled to wind speed, active power and current. In cluster 5 at high wind speeds active power and current decouple from wind speed and pitch angle as well. In both states, the pitch angle is used to decrease the intake of power of the turbine. The tip speed ratio behaves as expected from Eq. (4.1).

Figure 4.7 shows that the clusters are not as clearly separated over wind speed alone as with only five observables. This is mainly true for low wind speeds. The three regimes we identified in our main analysis can again be distinguished. Furthermore, we see in Fig. 4.8 that the small range intermediate cluster 3 can be distinguished from cluster 4 when looking not only at the average wind speed in the epoch but also at the standard deviation of wind speed in the epoch. Cluster

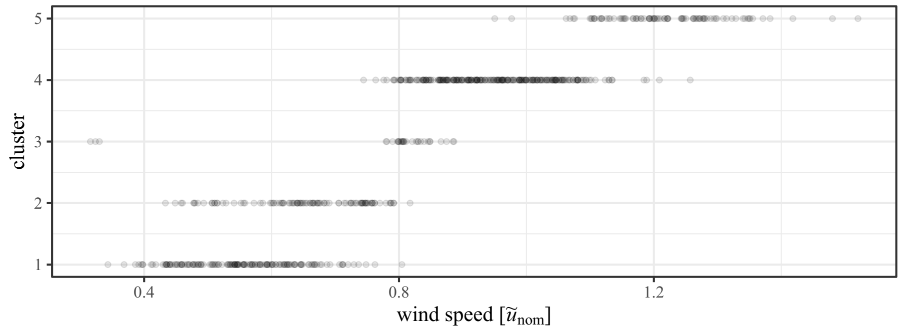


Figure 4.7: Cluster identifier s over wind speed for WT1 and included pitch angle. Each dot represents a 30 minute epoch.

3 exists for small standard deviations. One possible explanation is that the wind changed so little that the controller did not change the pitch angle even though it would already do so in this regime as seen in cluster 4. There could also be a small wind speed regime where the controller already tries to keep rotation constant, but does not change the pitch angle to this end. We must also mention that the amount of filled missing values for the pitch angle time series is quite high in cluster 3 as can be seen in Fig. 4.9. This is also true for the overlapping clusters 1 and 2, for which we did not find a clear distinction criterion. The high amount of filled missing values in clusters 2 and 3 could be reasonable as the pitch angle is decoupled from other observables. This would lead to a constant value which would lead to many missing values in the time series, if the reasoning in Sec. 4.1 is correct. However, without knowing for certain that this reasoning is correct, the decoupling of pitch angle from the rest that we find could also be an artifact of the data manipulation.

In conclusion, we have seen that non-stationarity can also be detected with our clustering when including pitch angle and tip speed ratio. The primary influence still stems from the control strategies in dependence of the wind. In general, it can be necessary to have more than one observable, on which the distinction between clusters is based. It could also happen that some overlap cannot be distinguished and more than one normal state would need to be considered for analysis of the system under those operating conditions.

Further analysis of matrices with more and different observables as well as an attempt to distinguish other influencing factors is interesting for future work. Regulatory impacts on the correlation structure as, for example, curtailment should also be considered. We will continue the current work with an analysis of the possibility to predict the state solely based on wind speed for our five main observables.

4.4 Cluster Prediction by Wind Speed

Having established a strong influence of the control system on the structure of the analyzed correlation matrices, we will now try to predict which correlation matrix state, i.e. operational state, the turbine should be in based on the wind speed. We have seen in the analysis with additional observables that overlaps between states can happen when differentiating solely based on wind speed. In such cases, more

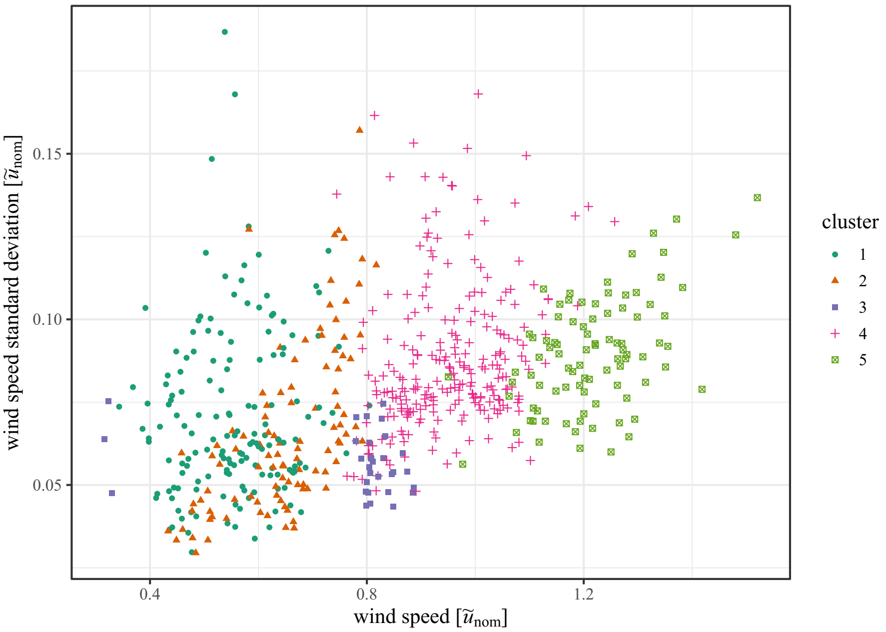


Figure 4.8: Cluster allocation as colored points over the average wind speed in an epoch on the x-axis and the standard deviation of wind speed during an epoch on the y-axis.

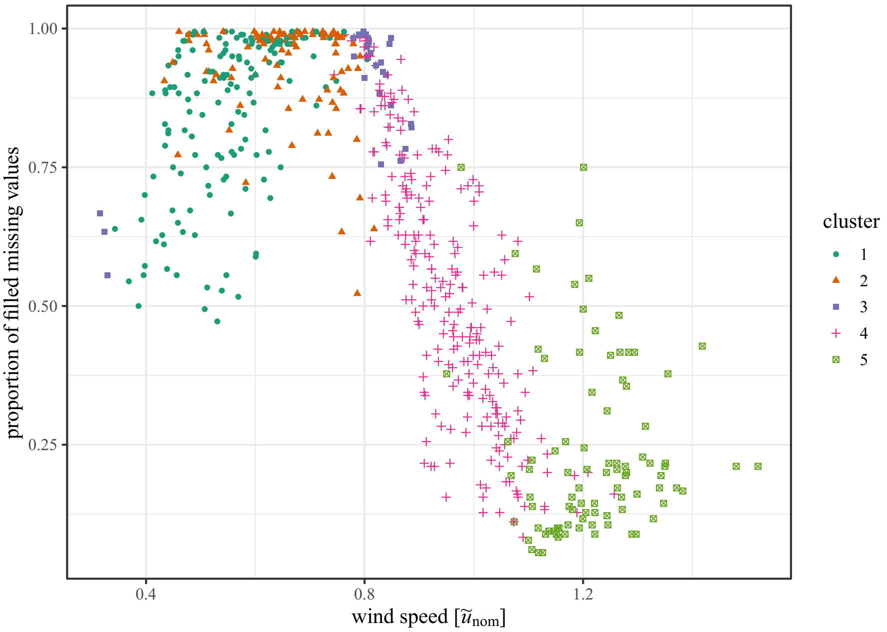


Figure 4.9: Cluster allocation as colored points over the average wind speed in an epoch on the x-axis. The y-axis shows the proportion of filled missing values in the pitch angle time series relative to all values in an epoch.

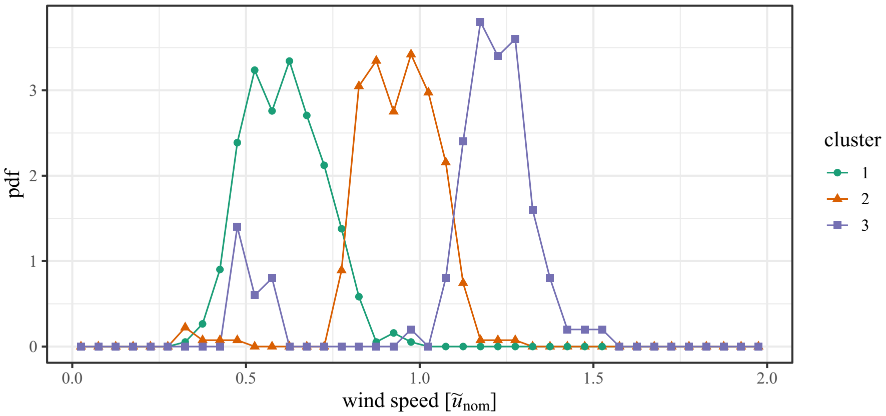


Figure 4.10: Probability density functions for the 30 minute epoch mean wind speed per cluster. The width of bins is $0.05\tilde{u}_{\mathrm{nom}}$.

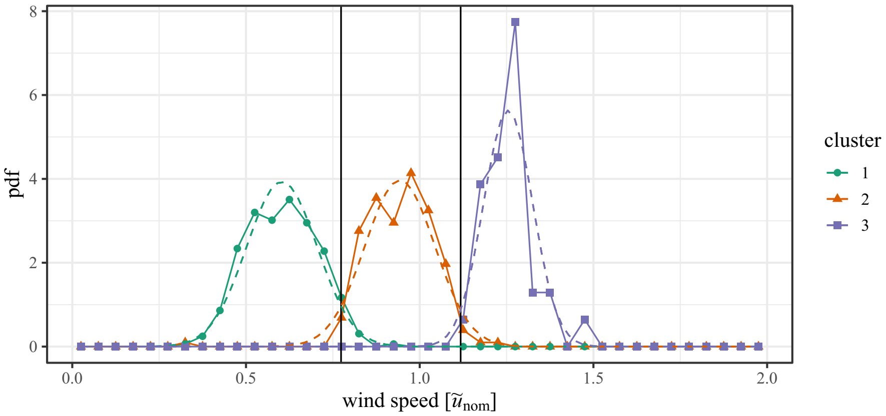


Figure 4.11: Probability density functions for the 30 minute epoch mean wind speed per cluster using only epochs with a silhouette coefficient above the first quartile of all silhouette coefficients. For each cluster a normal distribution was fitted. Then black vertical lines indicate the intersections of these distributions and thereby the boundary wind speeds. The width of bins is $0.05\tilde{u}_{\mathrm{nom}}$.

observables or external parameters might be necessary for state prediction. For the current work, we will stick with the simpler example of five observables. Here, we are confident when separating the clusters solely based on wind speed and show that such a distinction can, in general, work. This makes the example easy to follow. Also, this way we do not need to worry about the filled in pitch values. We now present a method that allows separation of the three states found based on wind speed. To this end, we look at the distribution of wind speeds in the different states, analyze them and then predict the boundary wind speeds that separate the distinct groups. This will show that it is possible to account for found non-stationarity, even though small adaptations are likely to be necessary when considering different sets of observables or turbines. In Fig. 4.10 one can see the empirical probability density functions (pdf) for wind speeds per cluster state. As already expected from Fig. 4.4, we can clearly distinguish the different regimes. However, we identify much more clearly what we are calling mismatches: epochs that are sorted into cluster 3, but have mean wind speeds in the range associated with cluster 1. They make up the left peak of the distribution for cluster 3. Furthermore we can see a small peak in the probability density function for cluster 2 lying at very low wind speeds beneath the distribution for cluster 1. These could be reasonable as the rotation of the generator shaft is kept at a minimum rotational speed needed for operation of the turbine for very low wind speeds as discussed in the previous section. However, the data in the very low wind regime is sparse and not as reliable as the turbines often move in and out of operation during these times due to shutting off below a certain minimal wind speed, therefore we will disregard this first boundary u_1 for now.

Before modeling the boundaries between the distributions, we try to compensate for mismatches due to matrices lying at the edge of two clusters, matrices being wrongly sorted, or singular outliers. This can easily be done by using the silhouette coefficient we have introduced before. It gives an indication of how good a member of a cluster fits into this cluster compared to the other clusters. This means that any 30 minute epoch that has been sorted into cluster s but should rather be in cluster r will have a very small or negative silhouette coefficient. We can use this fact and remove from the calculation of the probability density function all epochs with a silhouette coefficient below the first quartile of all silhouette coefficients, which can be seen in Tab. 4.1. The resulting probability density functions can be seen in Fig. 4.11. The second peak at low wind speeds for cluster 3 disappears. This indicates that our reasoning of a mismatch was correct. The persistence of the small peak at very low wind speeds for cluster 2 on the other hand shows that it indeed points toward a control of the rotational speeds in this regime.

The empirical distributions are noisy due to the finite amount of data points. This is especially true for cluster 3, which contains the least epochs. However, it is very clear that every cluster is representing a wind speed regime. There are now two basic approaches to defining the boundaries between these regimes. One can simply look at the empirical data and define for each value of wind speed the maximum likelihood state based on the empirical probability density function. Secondly, one can fit a distribution to the data and calculate the intersections of these, which represent the boundaries. For now, we choose to fit distributions as it turns out that a normal distribution is a good choice for each wind speed regime (see Fig. 4.11) and the other method could be heavily influenced by noisiness in the empirical data. Of course, this will dismiss the smaller peak of cluster 2. It should be taken into account if enough data exists in this regime (cf. Sec. 4.5). For now, cluster 1 has a mean of $0.603\tilde{u}_{\mathrm{nom}}$ and a standard deviation of $0.101\tilde{u}_{\mathrm{nom}}$. Clusters 2 and 3 are centered at $0.943\tilde{u}_{\mathrm{nom}}$ and $1.255\tilde{u}_{\mathrm{nom}}$ with standard deviations of $0.101\tilde{u}_{\mathrm{nom}}$ and $0.071\tilde{u}_{\mathrm{nom}}$ respectively. These values lead to boundaries at $u_2 = 0.774\tilde{u}_{\mathrm{nom}}$ and $u_{\mathrm{nom}} = 1.118\tilde{u}_{\mathrm{nom}}$. Interestingly, the last value shows that u_{nom} as calculated in our analysis is larger than $\tilde{u}_{\mathrm{nom}}$. The reason for this discrepancy lies in realistic operational conditions. The power curve (active power in dependency of wind speed) as given by the manufacturer is one line. Accordingly, there is exactly one value $\tilde{u}_{\mathrm{nom}}$ which marks the starting point for nominal power production. In reality, especially when looking at high-frequency data, there will always be an area around this line

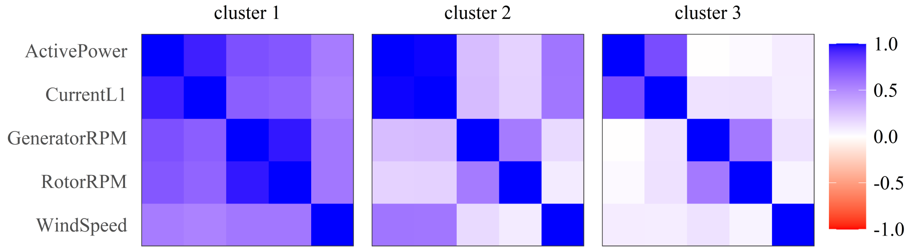


Figure 4.12: Matrices corresponding to the group centroids after sorting with the epochs according to the calculated boundaries for WT1. The mean matrices were calculated for all epochs, not only those used to determine the boundaries. The color indicates the value of the correlation coefficient.

which is realized. The value $\tilde{u}_{\mathrm{nom}}$ lies in the middle of this smeared out power curve. At this wind speed nominal power output can be reached but is not yet constant. With even higher wind speeds, it becomes less and less likely that the actual power produced lies beneath the nominal value. Only when this probability nearly vanishes, a change in correlation structure is detectable by our method. It is therefore reasonable that our value u_{nom} lies higher than $\tilde{u}_{\mathrm{nom}}$. While our value is therefore well suited to distinguish correlation states, it cannot be directly compared with the nominal wind speed given by the manufacturer of a turbine. We have confirmed this by looking at scatter plots of our data but cannot show them in this thesis due to data confidentiality.

We want to point out two things. First, when using our method as a pre-processing for an analysis it needs to be run on the same observables that are to be considered in the analysis (compare Sec. 4.3.2). Some adaptations for additional influences from the external conditions, regulatory influences (e.g. curtailment, de-ration) or overlaps of clusters might be necessary. Second, it is not proven that a normal distribution will always provide the best fit. For example, if much larger wind speeds exist in the data set, the distribution for cluster 3 would have much heavier tails in the large wind speed regime unless cut off by the cut-out wind speed of the turbine.

In the current case for WT1 the model fitted works very well. Compared to the clustering solution we have 9.9% of changes in group assignment. If one only looks at epochs with silhouettes above the first quartile this number reduces to 3.7%. This is clear as the epochs previously characterized as mismatches obviously change their cluster allocation when applying the model to all epochs. The mean correlation matrices of the states as sorted by the model are shown in Fig. 4.12. They clearly exhibit the different structures discussed above produced by the control system of the turbine showing that our state prediction works. This state prediction is an essential first step for using the different operational regimes as a pre-processing for data analysis. Using the fitted criterion, one can predict what state the turbine should be in and run the analysis for the corresponding operational regime. This is necessary if the analysis itself (e.g. principal components) directly involves the correlation matrix. Otherwise, it is also possible to make direct use of the clustering and simply identify the state by comparing the current correlation matrix to the previously identified cluster centers. A direct application of both that we want to test in the future is monitoring the correlation structure. Using the fitted criterion we predict a correlation matrix and compare it to the current one. With sufficiently labelled data we plan to analyze if deviations might signify operational problems or failures.

As we have seen in our analysis with additional observables in Sec. 4.3.2 for low wind speeds, large overlaps between multiple states can occur when differentiating by wind speed. If this is the case in an analysis at hand, one could look for other distinguishing factors. However, it is not a given that these exist. Another possibility is to accept that more than one control state is normal for the given conditions and compare new data to all possible states. If for example the goal is to minimize false alarms in a failure detection procedure, one could run failure analysis in all likely states and then choose the one that gives the least indication for failure. An alternative could be weighing the failure indicators with the likeliness calculated for each state under the given conditions.

4.5 Application to Multiple Turbines

For a single wind turbine we identified different operational states in the correlation structure and presented a model to distinguish these states based on wind speed. To be useful for applications, our findings need to be general characteristics and not be specific for one turbine. We proceed and test our methodology for all turbines in the data set. We want to emphasize that in a first step this does not mean assuming one model with fixed wind speed boundaries and applying it to all turbines. Rather, we test if the procedure described in previous chapters can be automatically transferred to other turbines without supervision. Hence, we perform cluster analysis and fit the boundaries for each turbine separately.

An easily comparable indicator for the quality of the proposed methodology is the relative numbers of cluster allocation changes from the model compared to the clustering itself. We already discussed that for WT1 at the end of the previous section. This number will drastically increase if either of the two steps in the calculation does not work well on a turbine: If the clustering algorithm returns a solution that is not grouped by wind speed, sorting on the basis of wind speed will change the allocation of many epochs. If they are clustered according to wind speed, but the boundaries are less sharp than for WT1, it will again result in many changed allocations.

A histogram over all calculated allocation changes for the 98 turbines is shown in Fig. 4.13. We can see that the changes for WT1 lie in the lower end of changes, as expected due to it not showing any alarms or failures in the chosen time span. This does not hold true for the other turbines, most of which exhibit a few more changes. However, there are multiple turbines which show no more changes than WT1. Also, for those that do the fitted boundaries still work remarkably well. Some allocation changes are always expected for the reasons discussed above and especially near the

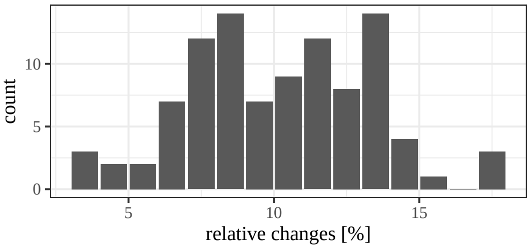


Figure 4.13: Histogram counts showing the frequency of relative changes in state allocation when comparing clustering and individual models per turbine.

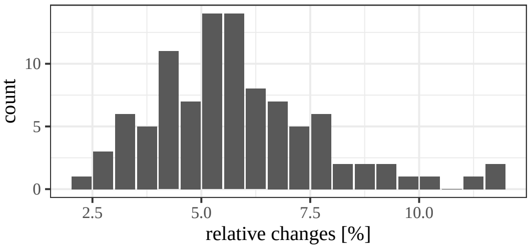


Figure 4.14: Histogram counts showing the frequency of relative changes in state allocation per turbine when comparing the clustering solution and the maximum likelihood model based on data from all turbines.

boundaries the distinction between two states is not always perfect. Concluding that our proposed method works well for all turbines in our dataset, we proceed with an optimization.

It stands to reason that the probability density functions for wind speeds per cluster should be much smoother, if we look at all turbines at the same time. Combining the data from all turbines before fitting the model, we have approximately 98 times more data points than for a single turbine. The resulting distributions are shown in Fig. 4.15. They are indeed much smoother. Furthermore, we can see that the previously assumed Gaussian fit would not work well anymore. Especially the distribution shown for cluster 1 is skewed. Also, the peak in the distribution for cluster 2 beneath wind speeds of $0.4\tilde{u}_{\mathrm{nom}}$ becomes more explicit. This second point is explained in sections 4.3 and 4.4 and actually underlines our reasoning: For very low wind speeds rotation is kept constant and with more data from all turbines we can distinguish this regime more clearly than before.

At a first glance, the skewness of the probability density functions does not fit our theory so well. If the controller would work perfectly and instantly and the wind speeds were constant during each epoch, we would expect the distinctions between the operational states to be represented by rectangular functions as probability densities. In non-perfect conditions these would overlap and smooth out to be something similar to a Gaussian distribution as assumed before, but they should not become skewed. However, the reason can be found in the underlying distribution of wind speeds in the inlay in Fig. 4.15. It follows roughly the expected Weibull distribution. The deviations could be explained by combinations of influences in the environment of the wind farm, overlying of different Weibull distributions for different wind directions or maybe even measurement effects due to the sensor being placed behind the rotor. Some differences might also be introduced by the removal of low silhouette coefficients as these will appear often in the regimes of the boundaries between states where two correlation structures might be mixed during an epoch. This is of interest for future studies. For now, we can take away that the skewness of this underlying distributions might lead to the skewness in the cluster probability density functions. To check this we replace the histogram count of epochs $h_s(u_w)$ for wind speed u_w and cluster s with the rescaled count

$$\tilde{h}_s(u_w) = \frac{h_s(u_w)}{h_{\text{total}}(u_w)}$$

by dividing with the total histogram count of epochs $h_{\text{total}}(u_w)$ for that wind speed and all clusters. This basically removes the effect of the underlying wind speed distribution by transforming it into an equal distribution. The resulting probability density functions for each cluster can be seen in Fig. 4.16. Indeed, we can now see symmetric areas, showing behavior very similar to rectangular functions for cluster 1 and 3 with cluster 2 being smoothed out to a more Gaussian curve, because its wind regime is quite narrowly bounded by overlaps with the other states. This strongly underlines the existence of three regimes corresponding to operational states of the turbines.

As the functions for all turbines are much smoother and the bin size can be reduced, we can apply the direct maximum likelihood method instead of fitting a continuous curve to decide cluster allocation based on the epoch wind speed. This leads to three instead of the previous two boundary wind speeds to account for the appearance of cluster 2 in the very low wind regime. The values of the boundaries are $u_1 = 0.38\tilde{u}_{\text{nom}}$, $u_2 = 0.80\tilde{u}_{\text{nom}}$ and $u_{\text{nom}} = 1.10\tilde{u}_{\text{nom}}$. The resulting histogram of changes due to model allocation compared to the clustering is presented in Fig. 4.14 and shows less changes compared to Fig. 4.13. One reason for this is the taking into account of the very low wind speed regime in cluster 2. Such a method without need for fitting could be easily transferred to other data and turbine types providing high usability as a pre-processing step for data analysis. It will need to be tested in future work, how best to deal with wind speeds where the clusters overlap. The simplest method proposed above is an all-or-nothing approach choosing the likeliest cluster. Contrary to that, if one wants to minimize false alarms in a failure detection for example, it could be useful to compare current data to all clusters which are possible for the current wind speed and choose the one indicating the least likelihood for a failure. Additional filters alongside the current wind speed can also be necessary as seen in Sec. 4.3.2 where the standard deviation of wind speed helped separating clusters. Also, operating measures such as curtailment might lead to temporary changes in the boundaries between clusters and thereby create the need for an additional filter. In general, when applying this method, one should always check the clustering results beforehand.

Overall we conclude that the results formerly shown for WT1 are easily transferred to multiple wind turbines. Furthermore, the model to decide state allocation based on wind speed can be optimized by taking into account more turbines and thereby more data.

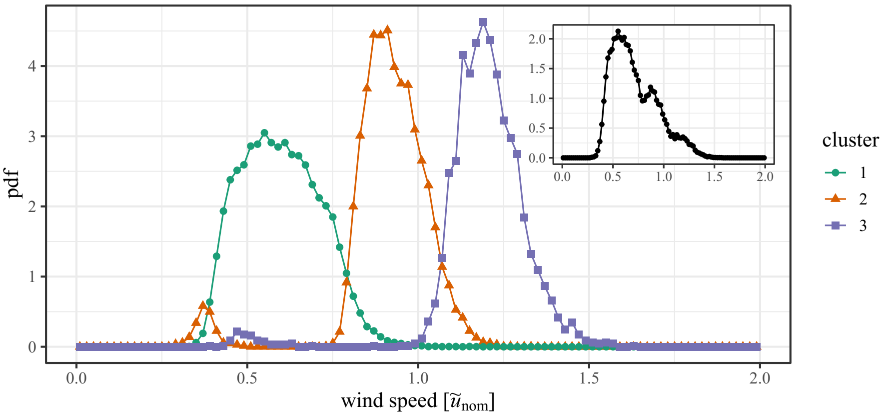


Figure 4.15: Probability density functions for the 30 minute epoch mean wind speed per cluster considering all turbines. Only those epochs with a silhouette coefficient above the first quartile of all silhouette coefficients for each turbine were used (calculated separately per turbine). The underlying wind speed distribution without cluster separation is shown as inlay. The width of bins is $0.02\tilde{u}_{\mathrm{nom}}$.

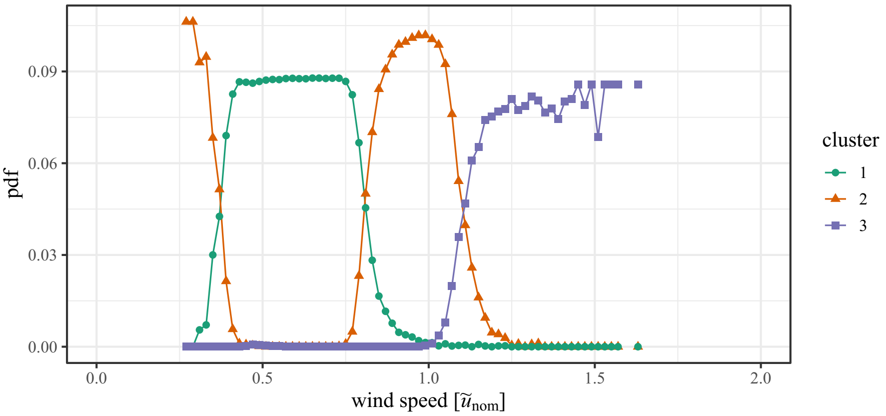


Figure 4.16: Probability density functions for the 30 minute epoch mean wind speed per cluster considering all turbines after dividing by the total number of counts per wind speeds to rescale the underlying distribution to an equal one. Only those epochs with a silhouette coefficient above the first quartile of all silhouette coefficients for each turbine were used (calculated separately per turbine). The width of bins is $0.02\tilde{u}_{\mathrm{nom}}$.

4.6 Conclusions

Using a matrix distance measure and clustering algorithm formerly applied to other complex systems we were able to identify different operational states in 30 minute correlation matrices of high frequency wind turbine data without prior knowledge of the control system. This demonstrates the non-stationarity of the correlation matrix for wind turbines and its automated detectability. While the states quite often exhibit stability over a certain time period, the real dependency lies with wind speed. This is expected for wind turbine control regimes. In the analysis with additional observables, the standard deviation of wind speed during a 30 minute epoch was also shown to have an influence. Furthermore, it was possible to model the boundary wind speeds between the different states for the main analysis of five observables - again without knowledge about the actual parameters used in the control system. This allowed us to recreate the cluster allocation solely based on the 30 minute average wind speeds. Being developed on one turbine, the method is transferred easily to multiple turbines. Results were improved by this increased amount of data. Our study shows clearly that the control system causes detectable non-stationarity in the correlation structure of high frequency wind turbine SCADA data. The automatic separation of states is important to account for this non-stationarity when analyzing such data, for example to monitor a turbine during operation.

While it is of course known that the control system of the turbine changes its operational behavior based on the external influences, our analysis proves that the influence on the correlation structure of the SCADA data is significant and an automatic distinction based on the correlation matrix is possible. Therefore, assuming a stationary correlation matrix, e.g. when applying principal component analysis for dimensionality reduction on a dataset, is unjustified.

Furthermore, it could potentially be important for monitoring with high frequency SCADA-data, e.g. when applying failure detection. Especially methods directly dependent on the correlation matrices such as principal components [140] and Mahalanobis distance [183] might benefit from the definitions of multiple, distinct normal states in the correlation behavior as they usually assume stationarity. They are commonly applied to wind turbines [99, 111, 116–118, 184], see also the reviews mentioned at the beginning of this chapter. We show in Ch. 7 that for generic correlated systems with distinct normal states, the knowledge of these states increases the sensitivity of change detection based on principal components. Incorporating this knowledge into an analysis is possible via pre-processing: Using a criterion based on historical data – wind speed in the presented case – new or live data could be compared to the respective operational states. Charmingly, the proposed ansatz does not require changes in established techniques, it just requires their application to multiple subgroups and is therefore easily implemented.

CHAPTER 5

Dynamics of Operational States

We have seen in the previous chapter that SCADA data and their correlations are significantly influenced by the turbine's control system. In Sec. 2.3.6 we discussed the basics of wind turbine control. The complex turbines of today each have their own variation of sophisticated control strategies to regulate their operation [180, 185]. Menezes et. al. [186] provides a general review of common control strategies. Several studies have explored the analysis of such control systems in wind turbines, including both theoretical and experimental investigations [96]. One approach to analyze wind turbine control systems is by means of modeling and simulation. For example, Pustina et al. [187] presented a nonlinear predictive model approach for power maximization and tested it in simulations using the OpenFAST environment. Another approach is the usage of optimization techniques. For example, Fernandez-Gauna et al. [188] proposed a control strategy for wind turbines based on a predictive machine learning model. The authors managed to combine control of the pitch system and the generator torque in one optimization problem. In addition to modeling and simulation, experimental studies have also been put forward to analyze wind turbine control systems. For example, Pöschke et al. (2022) [189] conducted a validation study of a model based control strategy in a wind tunnel. Some control strategies are also developed and analyzed for specific types of wind turbines. For example, Lopez-Queija et. al. [190] review state of the art control systems for floating offshore turbines.

In this chapter we pursue a new approach. The presented research is a modified version of paper [2]. We study the dynamics of operational states based on real SCADA-data (Supervisory Control And Data Acquisition). We introduce a method based on short-term Pearson correlation matrices and Langevin analysis that two of the present authors developed for financial data [168, 169]. In Ch. 4 we showed that clustering Pearson correlation matrices calculated over short time periods facilitates automatic distinction between operational states of the wind turbine controller. This requires the usage of high-frequency SCADA-data. Langevin analysis is a powerful approach that uses a combination of deterministic and random processes to model the dynamics of a system. It is already used in the wind energy field to study power curve behavior [191]. By employing Langevin analysis on the correlation matrices, we gain a direct way to analyze the dynamics of the control system during real operation. We test the applicability of the method by analyzing data of one year from an offshore VESTAS wind turbine.

We present the methodology in Sec. 5.1. Here, we first very shortly reiterate the calculation of a correlation matrix time series in 5.1.1 and then illustrate the Langevin analysis thereof in 5.1.2. Afterwards, we introduce the data set used for our analysis in Sec. 5.2. We present the results from applying the introduced methodology on our dataset in Sec. 5.3. Here, we will also adapt the Langevin analysis to better resolve hysteresis features observed in our dataset. In Sec. 5.4 we discuss the applicability of the presented method to wind turbines and the findings on our data.

5.1 Theoretical Background

5.1.1 Correlation Matrix Clustering

To obtain the operational states, whose dynamics we want to study, we proceed exactly as described in Ch. 4. We calculate Pearson correlation matrices in epochs of 30 minutes, i.e. on a moving time window. Thereby, we obtain a time series of matrices $C(\tau)$ with τ labeling the epoch as before. Then, we perform cluster analysis using a bisecting k-means algorithm. We sort the matrices into groups that are alike according to the distance measure defined in Eq. (4.2)and obtain $S = 3$ clusters. The center matrix for each cluster s is denoted C_s according to Eq. (4.4). We have dropped the turbine identifier l in the notations as it is not needed any longer. For each cluster, we define a dynamical observable describing the total current correlation structure, i.e. operational state, of the turbine as

$$d_s(\tau) = ||C(\tau) - C_s|| \quad . \tag{5.1}$$

It is the distance between the matrix for epoch τ (current operational state) and the center for cluster s (operational state expected for a given cluster s).

5.1.2 Langevin Analysis

To analyze the time series $d_s(\tau)$ we utilize Langevin analysis. We assume that the time evolution of $d_s(\tau)$ is described by

$$\frac{\mathrm{d}}{\mathrm{d}\tau} d_s \bigg|_{x(\tau)=x} = D^{(1)}(d_s, x) + \sqrt{D^{(2)}(d_s, x)} \cdot \Gamma(\tau), \tag{5.2}$$

with conditioning on an arbitrary additional observable x. The first Kramers-Moyal coefficient is the drift, denoted $D^{(1)}(d_s, x)$, and the second Kramers-Moyal coefficient the diffusion, denoted $D^{(2)}(d_s, x)$. Additionally, we introduce a delta-correlated Gaussian noise $\Gamma(\tau)$ with a variance of two. We focus on the drift estimation for d_s conditioned on the placeholder x (representing the time and wind observables in later analysis). If we had an infinite number of measurements at every point (d_s, x) and any possible time increment ϑ, we would be able to compute the related increments

$$\Delta_\vartheta d_s(\tau) = d_s(\tau + \vartheta) - d_s(\tau). \tag{5.3}$$

and use these to calculate the n-th order conditional moment

$$\mathcal{M}^{(n)}(d_s, x, \vartheta) = \langle (\Delta_\vartheta d_s(\tau))^n \rangle |_{d_s(\tau)=d_s, x(\tau)=x}, \tag{5.4}$$

where $\langle . \rangle$ indicates the expectation value over all measurements at this point [192, 193]. Then, the n-th Kramers-Moyal coefficient would be given by

$$D^{(n)}(d_s, x) = \lim_{\vartheta \to 0} \frac{\mathcal{M}^{(n)}(d_s, x, \vartheta)}{n! \cdot \vartheta}. \tag{5.5}$$

However, the dataset retrieved from the correlation analysis described in Sec. 5.1.1 only consists of a finite number of equidistantly sampled data points with a sample interval of T. We use this as the smallest possible time step $\vartheta = T$ for the stochastic analysis. Moreover, we define $\vartheta_q = q \cdot \vartheta$, where $q = 1, \ldots, Q$, allowing us to compute the increments of the observable d_s over a time lag ϑ_q as

$$\Delta_{\vartheta_q} d_s(\tau) = d_s(\tau + \vartheta_q) - d_s(\tau). \tag{5.6}$$

Then, with our finite number of points (d_s, x), we employ the Nadaraya-Watson [194, 195] estimator to approximate the n-th conditional moment $\hat{\mathcal{M}}^{(n)}(d_s, x, \vartheta_q)$ at the point (d_s, x) over a time lag ϑ_q. We use the hat to indicate quantities estimated from data.

$$\hat{\mathcal{M}}^{(n)}(d_s, x, \vartheta_q) = \sum_{\tau=1}^{T_\text{end}-\vartheta_q} \left(\Delta_{\vartheta_q} d_s(\tau)\right)^n \cdot \frac{\kappa_{a,b}\left(\frac{d_s(\tau)-d_s}{h_d}, \frac{x(\tau)-x}{h_x}\right)}{\sum\limits_{\tau=1}^{T_\text{end}-\vartheta_q} \kappa_{a,b}\left(\frac{d_s(\tau)-d_s}{h_d}, \frac{x(\tau)-x}{h_x}\right)} \tag{5.7}$$

Here, $\kappa_{a,b}(y_1, y_2)$ represents a two-dimensional kernel, and h_d and h_x correspond to the bandwidths utilized for the estimation. We calculate the two-dimensional kernel as the product of two one-dimensional kernels, i.e. a D-kernel,

$$\kappa_{a,b}(y_1, y_2) = \kappa_a(y_1) \cdot \kappa_b(y_2). \tag{5.8}$$

There are plenty of different kernel functions which are useful for different scenarios. The Epanechnikov, Gaussian, and rectangular kernels are three commonly used kernel functions in non-parametric estimation and smoothing techniques [196]. Each of these kernels has distinct properties that impact their use and the resulting estimation or smoothing outcomes. An overview is shown in Tab. 5.1.

The choice between these kernels depends on the specific characteristics of the data and the desired properties of the estimation or smoothing procedure [197]. The Gaussian kernel is popular for its smoothness and computational efficiency. The Epanechnikov kernel is often favored when efficiency and a localized smoothing

Table 5.1: Description of different kernel functions

Kernel	Shape	Efficiency	Tails and Robustness
Epanechnikov	flat and symmetric shape resembling a parabola, with its maximum value at the center	considered efficient, providing accurate estimates with the same amount of data	finite tails, making it more robust to outliers and extreme values
Gaussian	bell-shaped curve, characterized by a smooth and continuous decline in values away from the center	mathematically tractable and computationally efficient, especially in high-dimensional problems	infinite tails, making it sensitive to outliers and potentially leading to issues when extreme values are present
Rectangular	constant value within a fixed interval and drops abruptly to zero outside that interval	computationally efficient due to its simple shape and properties	finite tails but a constant value within the fixed interval, making it less robust to outliers and extreme values towards the edges of the interval

effect are desired. The rectangular kernel is suitable when simplicity and computational efficiency are prioritized, but it may not handle outliers or extreme values as effectively as the other kernels. We use a Gaussian kernel function

$$\kappa_{\mathrm{G}}(y) = e^{-\frac{1}{2}y^2}. \tag{5.9}$$

for the one-dimensional kernels [197, 198].

The bandwidth associated with the kernel function is just as crucial as the kernel function itself. When examining large-scale structures, it is useful to employ wider bandwidths. However, larger bandwidths may hamper resolution of small-scale structures. For the analysis of our data we found that the following values give good results and used them unless stated otherwise:

- for wind speed $h_{\mathrm{WindSpeed}} = 0.5\mathrm{m/s}$
- for time $h_t = 1.6\mathrm{h}$
- for the inter matrix distance $h_{d_s} = \dfrac{\max\limits_{\tau}(d_s(\tau))}{30}$.

We make the assumption that for small time increments ϑ_q, the conditional moments $\hat{\mathcal{M}}^{(n)}(d_s, x, \vartheta_q)$ exhibit linearity, and there is no additional measurement noise, implying $\hat{\mathcal{M}}^{(n)}(d_s, x, 0) = 0$ [199]. By averaging the conditional moments divided by the employed time increment ϑ_q, as depicted in Eq. (5.10), we estimate the n-th

Kramers-Moyal coefficients based on the provided estimations of the n-th conditional moments: [192, 193]

$$\hat{D}^{(n)}(d_s, x) = \frac{1}{Q} \sum_{q}^{Q} \frac{\hat{\mathcal{M}}^{(n)}(d_s, x, \vartheta_q)}{n! \cdot \vartheta_q} \tag{5.10}$$

As the minimal time step ϑ already is 30 minutes, we only consider $Q = 1$, so that ϑ_q will not be too large. Here, again a shorter epoch length $T = \vartheta$ would be desirable, but is not feasible due to the correlation matrix calculation.

Sometimes a more intuitive description is the potential

$$\hat{\Phi}(d_s, x) = -\int \hat{D}^{(1)}(d_s, x) \mathrm{d}d_s \tag{5.11}$$

calculated from the first Kramers-Moyal coefficient. Local minima of the potential correspond to stable fixed points in the system. When looking at the drift, which is the derivative of the potential, these points are indicated by values of zero and a negative slope at the zero crossing. Throughout our analysis, we will use whichever is best suited for understanding a particular issue.

5.2 Data

The data we use stems from the Supervisory Control and Data Acquisition (SCADA) system of a VESTAS turbine. It is situated in an offshore wind farm off the coast of Great Britain. The data were measured approximately every 5 second for the year 2017. To obtain consistent time stamps and a stable frequency, i.e. a consistent sampling interval Δt, the data were aggregated on 10 second time intervals by averaging. If no data were measured in the original 5 second set during a 10 second time interval, then there will also be missing data in our aggregated set.

The analyzed data contains measurements of five observables:

- generated active power (ActivePower)
- generated current (CurrentL1)
- rotation per minute of the rotor (RotorRPM)
- rotation per minute of the high speed shaft at the generator (GeneratorRPM)
- wind speed (WindSpeed)

As there are no deviations between the three phases of the generated current, we simply choose one of them. We expect from the VESTAS turbine a control shift from a low wind speed regime with variable rotation speed to an intermediate regime with constant rotation to a rated region with constant rotation and produced power. The studied observables are suitable to analyze these changes. The two rotational speed observables as well as current and active power are usually strongly coupled observables. We include all of them in our study as group correlations are important for the characterization of correlation structures. When trying to detect anomalies, for example, such structures might break up. It is also consistent with Ch. 4.

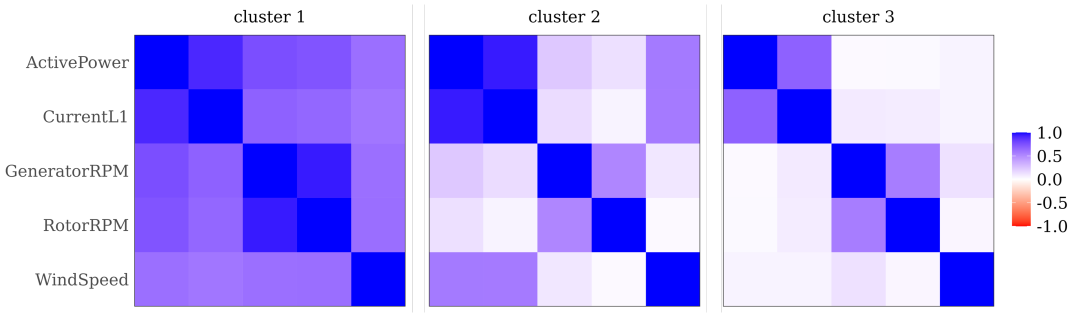


Figure 5.1: Cluster centers of Pearson correlation matrices calculated according to Eq. (4.4). The x- and y-axes are identical and display the different observables, for presentation we only labeled the y-axis. The matrix entries are represented by color.

5.3 Results

5.3.1 Correlation Matrix States

As a first step to analyze the operational dynamics of wind turbines, we must automatically distinguish the different operational states of the wind turbine. We apply the method developed in Ch. 4. For each epoch, we calculate the Pearson correlation matrix. The resulting set may be viewed as a time series of matrices. After applying clustering to this set, we obtain the three cluster centers shown in Fig. 5.1 according to Eq. (4.4). Their number is decided based on visual inspection as well as the silhouette coefficient. They represent different operational states of the turbine: At low wind speeds exists a regime where stronger winds lead to faster rotation, which in turn leads to more generated power. This is represented by Cluster 1. Intermediate wind speeds are best used by keeping the rotation at a constant, optimal value. This is indicated by the vanishing correlations of the rotational observables in Cluster 2. Here, more power is generated by increasing the torque. For high wind speeds, the turbine operates at rated power output, i.e. it has reached its upper power production limit. This is seen in Cluster 3. Rotation and produced power are both decoupled from the wind speed. The clustering procedure and wind speed distributions for each cluster were presented in detail in Ch. 4. When choosing different observables, the clusters and possibly their number will change.

The cluster centers represent typical operational states. However, with switching between different states and constantly changing external conditions, the correlation matrix for any epoch will usually fluctuate around the identified states. The current correlation structure is described in an aggregated way by the distance between the current correlation matrix and a chosen cluster center as calculated in Eq. (5.1). Thereby, we effectively look at the system from the viewpoint of one cluster center. An example time span for this matrix distance measure is shown in Fig. 5.2. We only plot d_1 as an example. This time series appears to be a stochastic process with noise fluctuating around fixed points (the cluster centers). Therefore, we now try to extract the deterministic components of its behavior to study the switching dynamics of the operational states.

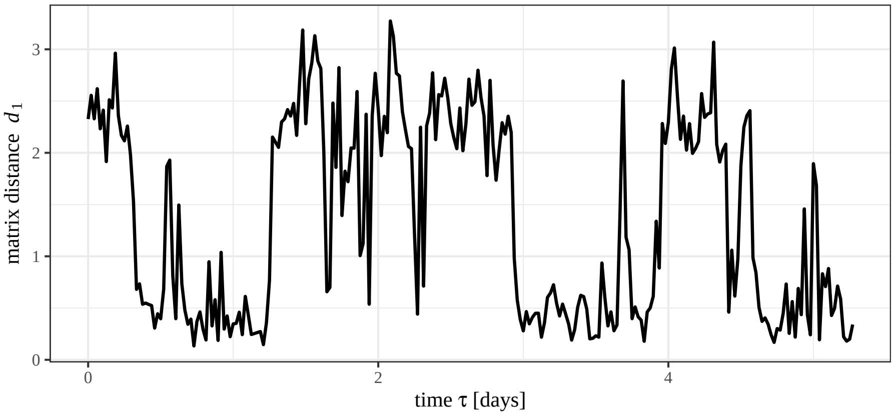


Figure 5.2: Matrix distance d_1 versus time τ for an arbitrary time span.

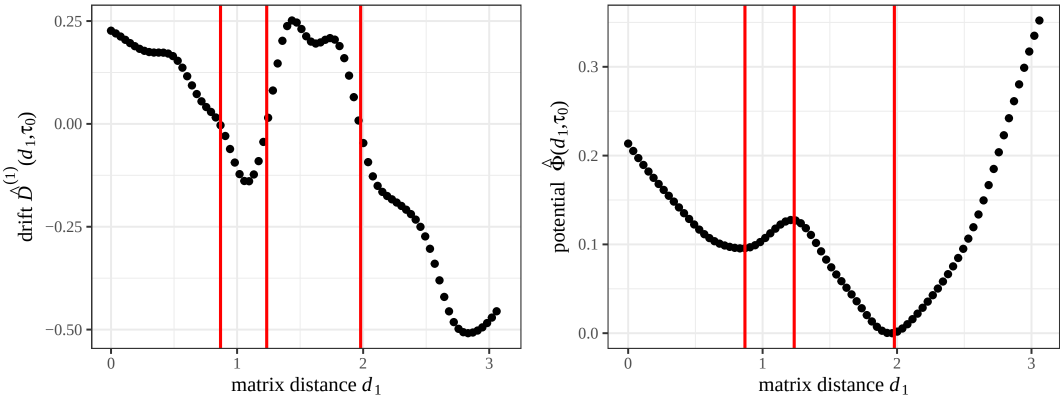


Figure 5.3: Drift $\hat{D}^{(1)}(d_s, t_0)$ and corresponding potential $\hat{\Phi}(d_1, t_0)$ versus matrix distance d_1 at time $\tau = \tau_0$. Here, the bandwidth for the distance is chosen as $h_{d_1} = \max_{\tau}(d_1(\tau))/10$. Red lines approximately indicate the values of d_1 where the drift crosses zero and thus stable or unstable fixed points, also seen in the potential.

5.3.2 Dynamics versus Time

Combining the identified clusters with the Langevin analysis as described in Sec. 5.1, we are now able to analyze the dynamics of operational states and their transitions. First, we look at the dynamics versus time, i.e. $x \widehat{=} \tau$. At a fixed point in time $\tau = \tau_0$, we obtain a drift $\hat{D}^{(1)}(d_s, \tau_0)$ and corresponding potential $\hat{\Phi}(d_s, \tau_0)$ versus matrix distance d_s. As an example, we show this for the matrix distance d_1 to Cluster 1 in Fig. 5.3. Any positive value in the drift means that the system tends to move to larger d_1, a negative value means the opposite. Therefore, a zero crossing in the drift with negative slope indicates a stable fixed point. In the potential this is represented as a minimum. If the drift crosses zero, but with a positive slope, the potential has a local maximum indicating, in general, an unstable fixed point. Here, one deep minimum exists representing the dominating operational state at $\tau = \tau_0$. The second, less deep, minimum indicates the presence of a second operational state in the vicinity (cf. the bandwidth in Eq. (5.7)) of $t = t_0$.

Now, we look at the non-stationarity over time. In Fig. 5.4 the potential is shown

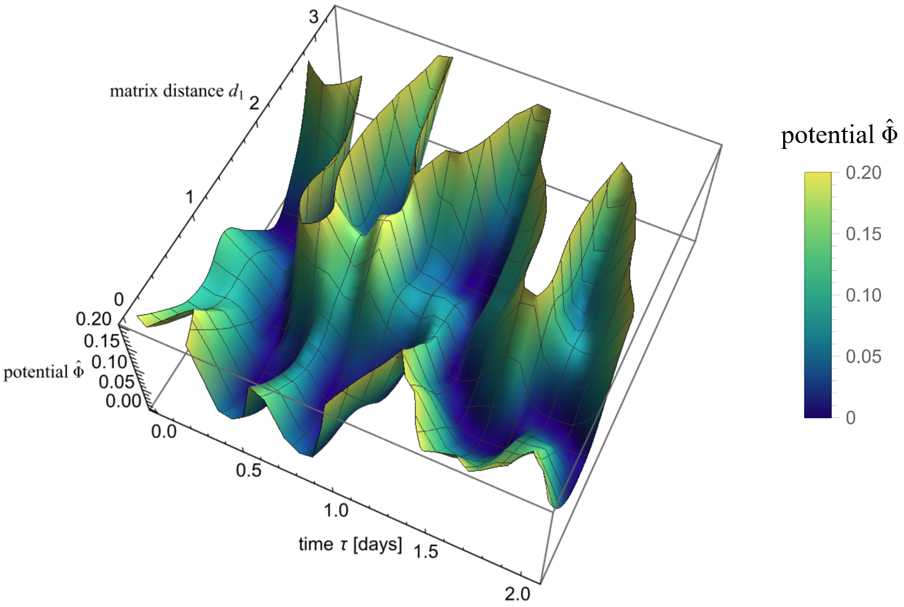


Figure 5.4: Potential $\hat{\Phi}(d_1, \tau)$ versus matrix distance d_1 and time τ. The shown total time span is approximately two days. The range of shown values is restricted for increased readability. Here, the bandwidth for the distance is chosen as $h_{d_1} = \max_{\tau}(d_1(\tau))/10$.

for a time span of two days. We see that it changes quite quickly over time. As expected the correlation matrix, i.e. the operating state, is not stable over time and therefore the potential changes. It is clear that the transitions between the states happen quite often (here, about 5 times in two days). However, we see that for the short periods where no changes happen, a clear minimum in the potential exists, indicating stability during these periods.

5.3.3 Dynamics versus Wind Speed with Standard Drift Estimation

In Ch. 4 we found that the operational states primarily depend on wind speed. Hence, another way to look at our system is by studying the non-stationarity not versus time, but versus wind speed, i.e. $x \widehat{=} u$. Thereby, we do not find the quick changes due to environmental conditions any longer. We rather see the transitions between operating states as they change with wind speed. Figure 5.5 shows the drift as viewed from cluster centers 1, 2 and 3. In Fig. 5.6 we also show the corresponding potentials for comparison. Essentially, at all times the drift is zero and the crossing occurs with negative slope, we find a minimum in the potential. These are the points where the system is stable. For a given wind speed, the system usually drifts to the matrix distance where the drift is zero. As the matrices in each cluster still fluctuate around their center, the stable fixed point for a cluster s when analyzing the matrix distance d_s in relation to that same cluster does not lie at zero, but at small positive values.

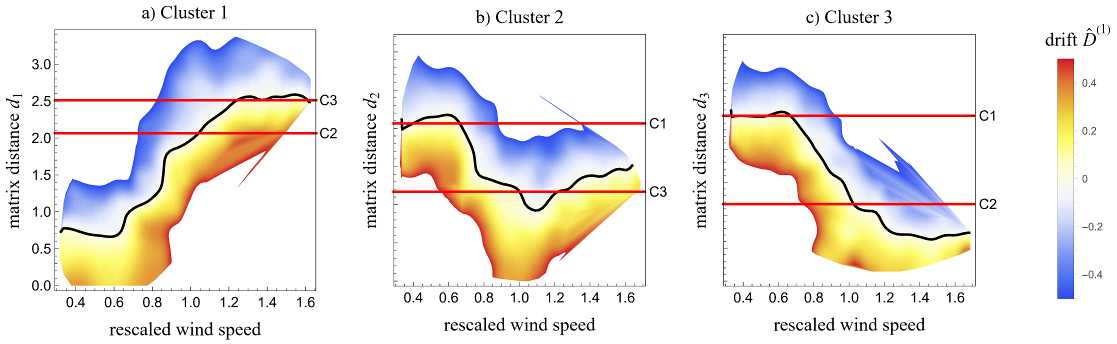


Figure 5.5: Color profile of drift $\hat{D}^{(1)}(d_s, u)$ versus matrix distance d_s and rescaled wind speed. For panels a), b) and c) the cluster s is chosen to be 1, 2, and 3, respectively. The conditional moments $\hat{\mathcal{M}}^{(1)}(d_s, u, \vartheta_q)$ for the drift estimation were calculated according to Eq. (5.7). The range of shown values is restricted for increased readability. The black line indicates where the drift is zero. Red lines indicate the distance to the other two cluster centers.

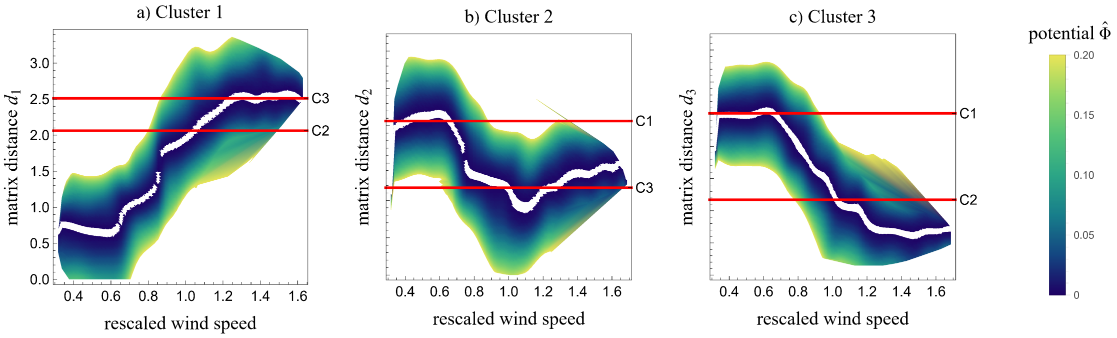


Figure 5.6: Color profile of potential $\hat{\Phi}(d_s, u)$ versus matrix distance d_s and rescaled wind speed. For panels a), b) and c) the cluster s is chosen to be 1, 2, and 3, respectively. The conditional moments $\hat{\mathcal{M}}^{(1)}(d_s, u, \vartheta_q)$ for the drift and potential estimation were calculated according to Eq. (5.7). The range of shown values is restricted for increased readability. The white area indicates the minimum of the potential, i.e. $\hat{\Phi}(d_s, u) < 0.0005$. Red lines indicate the distance to the other two cluster centers.

While the zero values of the drift do not exactly coincide with the distances between clusters at all times, the qualitative behavior is generally as expected. Clusters 1, 2 and 3 develop from low to high wind speeds. Cluster 1 dominates for rescaled wind speeds (RWS) between 0.3 and 0.7. Cluster 2 is the system's desired state between 0.85 and 1.15 on the RWS axis as is Cluster 3 between RWS=1.05 and RWS=1.6. We see, that Cluster 2 and 3 share the interval from RWS=1.05 to RWS=1.15. Here, the transition between these two clusters is visible. Clusters 1 and 2 do not seem to overlap. Their transition appears more complicated. When viewed from Cluster 1 in Fig. 5.5 a) it seems, there might be another, not yet defined state in the region from 0.6 to 0.9 on the RWS axis. It appears close to Cluster 1 in the matrix distance values, but shows a rather sharp transition into Cluster 2 at around RWS=0.9. However, viewed from Cluster 2 in Fig. 5.5 b) it appears to be the other way around. The intermediate state is close to Cluster 2 and shows a sharp transition to Cluster 1 at around RWS=0.7. Here, the state seems to also exist for higher RWS up to 1. Viewed from Cluster 3 in Fig. 5.5 c) one is inclined to infer two intermediate states between Cluster 1 and 2. Using the clustering algorithm to distinguish more than three states does not reveal such intermediate states, rather, smaller clusters of outliers with very few matrix elements split off Cluster 3.

One possible explanation is the occurrence of changing environmental conditions during the 30 minute intervals used for the calculation of the correlation matrices. This might then lead to correlation structures, which represent an average between different clusters. It is unclear if they are prominent enough to cause the appearance of a new intermediate stable point in the Langevin analysis. Another possibility is hysteresis in the control behavior, i.e. depending on previous conditions the controller does not always choose for the same operational conditions the same system behavior. Coming from low wind speeds, the turbine does not switch from State 1 to State 2 quickly, rather only if higher wind speeds persist for a certain amount of time. It is the other way for a transition from higher to lower wind speeds. This might lead to two fixed points for the same wind regime. Usually, this is resolved in Langevin analysis. However, in our case these two fixed points do not coexist at the same time. At any given time, depending on the operational state and its hysteresis, only one fixed point is present. If so, in the averaging process over the single increments in the data as calculated in Eq. (5.7) these two different behaviors are mixed. That this is true - at least to a certain extent - is seen in Fig. 5.7. We show three smooth kernel probability density functions for the increments in different regions of the RWS and matrix distance space. The overlap of different drift behaviors is obvious. In Fig. 5.7 we see for small matrix distances d_1 (blue curve) a large peak at approximately zero, which stems from the fixed point for Cluster 1. However, large values for the increments exist also, which result from times when the controller tries to realize operational state 2. The mean of the distribution therefore lies at small positive values as seen in Figs. 5.5 a) and 5.6 a). The distribution in Fig. 5.7 for large d_1 (green curve) shows the same effect, but the other way around. For intermediate d_1 (red curve) we see positive and negative values of the increments in Fig. 5.7, which lead to an average close to zero. Effectively, each of these probability densities is an overlap of two different densities stemming from the two different clusters.

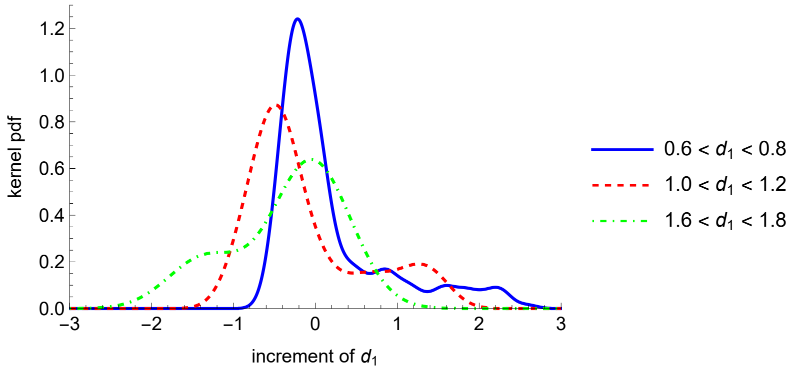


Figure 5.7: Kernel probability density functions (pdf) of increments in matrix distance d_1 measured in the data. The distributions are calculated for the RWS area: $0.65 < \text{RWS} < 0.75$. The covered area of matrix distance d_1 is indicated by color.

5.3.4 Dynamics versus Wind Speed with Modified Drift Estimation

To better resolve the issue of overlapping increment densities stemming from different clusters, we introduce a new estimation method for the drift. We attempt to determine the drift associated only with the density of the cluster, whose fixed point is closest in the space of matrix distance and wind speed. This cluster should be more likely to appear at this point as, in general, a system might be more frequently around its fixed point than very far away from it. Therefore, this cluster should be the origin of the highest peak in the pdf as seen in Fig. 5.7. In fact, Fig. 5.7 indicates that the hysteresis might be better visible if one estimated the drift as the maximum of the shown distributions instead of the average. This means we effectively trim the densities by excluding outliers to a symmetric distribution around the highest peak. Then, mean value and highest peak occur at the same value, which should be close to the drift value associated with only one cluster. Thus, we proceed from the estimation of conditional moments in Eq. (5.7) to a new method. For the estimation at point (d_s, x), we consider all increments of d_s to the power n over a time lag ϑ_q such that

$$\left(\Delta_{\vartheta_q} d_s(\tau)\right)^n, \ \tau = 1, \ldots, T_{\text{end}} - \vartheta_q. \tag{5.12}$$

For these quantities, we calculate a smooth kernel probability density function considering the same Gaussian weights as before in Eq. (5.7). We choose its maximum as the new $\hat{\mathcal{M}}^{(n)}(d_s, x, \vartheta_q)$. As we are interested in the deterministic drift, we set $n = 1$. All remaining calculations to derive the drift and potential remain as before. We display results for this new estimation of the drift in Fig. 5.8. We see that the general behavior is similar to the standard drift estimation shown in Fig. 5.5, but we resolve more details.

At very low wind speeds, we identify a regime, where Cluster 2 appears to be dominant. This was not seen before at all in Fig. 5.5. However, for the analyzed turbine this behavior is correct. A non-vanishing probability for Cluster 2 to occur in this regime was also seen in the extended cluster analysis in Ch. 4. The system tries to keep the rotation constant at a minimal viable rotation frequency for very low wind speeds. In the correlation structure this looks the same as the later on appearances of Cluster 2 even if the fixed value of the rotation is different here.

For higher wind speeds, between RWS$\approx$ 0.3 and RWS$\approx$ 0.8 Cluster 1 is dominant. We see good alignment of the distance to the cluster center (red lines) and the zero values of the drift when viewed from Clusters 2 or 3 in Figs. 5.8 b) and c). Viewed from Cluster 1 the drift is zero at small positive values in the matrix distance d_1. This is because the distance can only take positive values and the system still fluctuates around the center of the cluster. Starting from RWS$\approx$ 0.77 to RWS$\approx$ 0.99 we now resolve the complicated transition between Clusters 1 and 2. This is best seen when the system is viewed from Cluster 1 in Fig. 5.8 a). Here, a bistable region exists where both clusters exist for the same wind speeds due to hysteresis. We show this transition in detail in Fig. 5.9. We see that for RWS=0.77 one clear minimum exists representing Cluster 1. With increasing wind speed a second local minimum appears, which represents Cluster 2. At the beginning of transition the global minimum lies clearly at Cluster 1. For these low wind speeds State 2 is possible, but most often the turbine will be in State 1. Then, at RWS= 0.87 both minima have approximately the same depth, before with even higher wind speeds the first minimum representing Cluster 1 starts to become less deep than the one for Cluster 2. Here, at higher wind speeds, Cluster 1 might appear when the turbine did not yet switch to Cluster 2, but it becomes ever less likely. The potential minimum associated with Cluster 1 finally vanishes at RWS= 0.99.

Next, Cluster 2 is dominant at around RWS= 1.0. This is best seen when viewed from Cluster 2 in Fig. 5.8 b), but there is also a clear overlap between the zero values of the drift with the distance to Cluster 2 in Fig. 5.8 a). It does not take much higher wind speeds for the system to start shifting to Cluster 3. Interestingly, this transition is apparently not affected by hysteresis. We see a steady shift in the drift, also found in the potentials shown in Fig. 5.9. In a physics interpretation, the transitions from Cluster 1 to 2 and from 2 to 3 resemble phase transitions of first and second order, respectively. From RWS$\approx$ 1.25 onward, the system operates in State 3, i.e. with nominal power output. The drift values of zero and the distance to Cluster 3 coincide well, especially viewed from Cluster 1 in Fig. 5.8 a).

Overall, the regimes seem to be best resolved when viewed from Cluster 1, where the distances to Clusters 2 and 3 match the drift values of zero. When looking at matrix distances d_2 and d_3, Cluster 1 is resolved well whereas it appears hard to resolve the other cluster. This might be due to cluster centers 2 and 3 not being very far apart and the necessity for a non-vanishing bandwidth due to limited amounts of data. The qualitative behavior is visualized and estimated well in all cases. In the hyperspace of the correlation matrices, the different cluster centers apparently do not lie on a straight line. This means that the matrix distances d_s contain different information for each s. For example, the hysteresis between States 1 and 2 is best seen from Cluster 1.

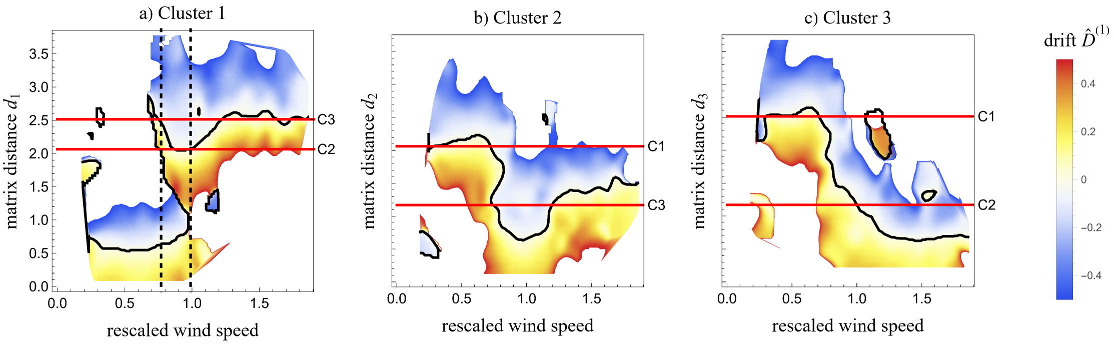


Figure 5.8: Color profile of drift $\hat{D}^{(1)}(d_s, u)$ versus matrix distance d_s and rescaled wind speed. For panels a), b) and c) the cluster s is chosen to be 1, 2, and 3, respectively. The conditional moments $\hat{\mathcal{M}}^{(1)}(d_s, u, \vartheta_q)$ for the drift estimation were calculated according to our newly proposed peak determination. The range of shown values is restricted for increased readability. The black line indicates where the drift is zero. Red lines indicate the distance to the other two cluster centers. Black, dashed lines indicate the bistable region in panel a) for $s = 1$ between RWS=0.77 and RWS=0.99 according to Fig. 5.9.

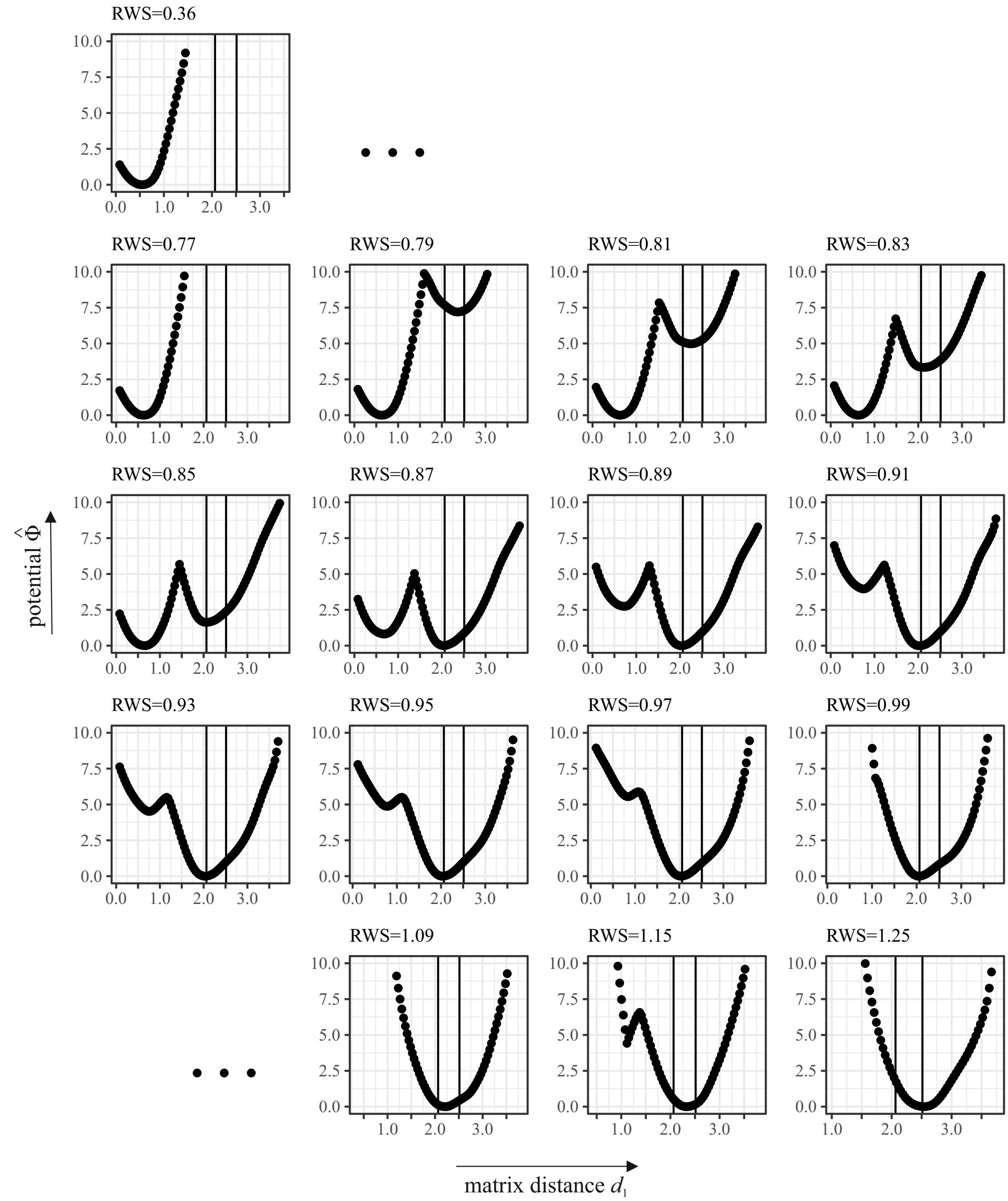


Figure 5.9: Potential $\hat{\Phi}(d_1, \tau)$ versus matrix distance d_1. Each subplot represents the potential at the indicated RWS. From upper left to lower right the RWS increases. The vertical line represents the distance to the center of Cluster 2. The individual plots are vertical slices of the potential corresponding to Fig. 5.8 a).

5.4 Conclusions

We studied the dynamics of operational states in wind turbines on the basis of real data. First we calculated Pearson correlation matrices on a moving time window, which facilitates studying the non-stationarity of wind turbine systems. Clustering the correlation matrices for one year of an offshore wind turbine, we identified three main operational states analogously to Ch. 4. Then, we combined our cluster analysis with Langevin analysis. The distance between a correlation matrix at any time t and one of the cluster centers proved to be a good indicator of the current operational status. Hereby, we effectively reduced the multidimensional system, in our case correlations of five different observables, to a one-dimensional time series. Langevin analysis of this new time series provided information on the operational dynamics. Hence, it was possible to describe the complex dynamics of the multidimensional system in a simplified way. This is, in general, also possible for more observables and larger systems. We visualized the drift and corresponding potential for an intuitive interpretation of the system dynamics.

For our data, a VESTAS turbine and five observables, we found three correlation matrix clusters, i.e. operational states of the wind turbine. The Langevin analysis allowed us to identify the wind regimes, where these states are stable. Furthermore, the regimes where transitions happen are also identified. While the presented method was previously applied in Econophysics [168, 169], we developed some adaptations of the method to the peculiarities of wind turbine data. Thereby, we succeeded in studying the nature of the state transitions. Between the states with constant rotation and operation at rated power we saw a smooth transition. The transition at lower wind speeds from a variable rotation state to constant rotation on the other hand, shows hysteresis behavior. In a physics interpretation they resemble second and first order phase transitions, respectively. We identified the existence of multiple fixed points in the low wind speed transition regime due to control hysteresis. Here, two fixed points exist in the system, but they never occur at the same time. This effectively leads to an overlap of different drift fields. With our new estimation method we were able to resolve these two fields. Thereby, we resolved the hysteresis effect as a bistable wind speed regime. Our new method is suited for analyzing transitions with hysteresis, where the potential minima never coexist at the same time.

Overall, while the control systems of a specific turbine are known at least to the manufacturer, our approach allows the analysis and visualization of their dynamics during real operation. The method is transferable to other wind turbine models and different choices of observables. Two caveats are in order: First, the observed operational states can also depend on more external factors than the wind speed. It might be necessary to study the drift and potential conditioned on multiple observables, but this does not constitute a problem in principle. Second, one has to carefully consider the number of states. Here, the Langevin analysis is helpful. We saw in the present study that it potentially reveals states that were not identified in the clustering. In Sec. 5.3 we saw the possibility for another state between Clusters 1 and 2, which we identified as a transition regime with hysteresis.

Apart from a direct study of non-stationary dynamics in turbine operation, our results are interesting also for normal behavior modeling in the context of failure prediction. Here, if one wants to detect anomalies, one must first define what is normal. This, of course, includes different behaviors introduced by the controller. Our analysis helps with the identification of wind speed regions where one can be sure of what is normal and transition periods where this is potentially unclear.

CHAPTER 6

Including Non-Stationarity: Langevin Power Curve Analysis

After having seen, that non-stationarity is surely present in high-frequency SCADA data, we now make use of that knowledge. In this chapter we employ the suggested pre-processing and run an established analysis method per cluster state. This has been done in paper [3], of which this chapter is a modified version.

Recently, the Langevin equation approach has been used to study the dynamics of wind turbines [200] [201] [202]. This approach utilizes temporally highly resolved measurements of wind speed and power output to determine the drift and diffusion coefficients of the energy conversion process, which characterize the deterministic and stochastic behavior of the system. However, this approach assumes a quasi-stationary system and does not account for the potential impact of different operational states.

As we have seen in Ch. 4, this assumption is not generally true. By clustering Pearson correlation matrices we identified different operational states. These states distinguish non-stationary behavior in the mutual dependencies and represent different turbine control settings. In this chapter, we employ the same clustering methods – including the pitch angle observable – to identify five operational states, which are then used to condition the Langevin analysis. The analysis reveals unique behavior patterns in the power conversion process corresponding to each operational state. The study also successfully resolves hysteresis effects commonly observed in the dynamics of wind turbines [201] [203].

In Sec. 6.1 we introduce the data used for this chapter. Then, we quickly summarize the methodology for calculating the correlation matrix states in Sec. 6.2.1. While the general concept of the Langevin analysis is the same as presented in Ch. 5, we introduce it nonetheless in Sec. 6.2.2 as this chapter can be read and understood independently from Ch. 5. Furthermore, we present the methodology for conditioning Langevin analysis on the cluster states in Sec. 6.2.3. Then, we present the results in Sec. 6.3 before drawing conclusions in Sec. 6.4.

6.1 Data

The data utilized in this study is sourced from the SCADA system of a Vestas wind turbine located in an offshore wind farm off the coast of Great Britain. These measurements were recorded at approximately 5 second intervals throughout the year 2017. To ensure consistent time stamps and a stable frequency, i.e. a consistent sampling interval Δt, the data were aggregated by averaging over 10 second intervals. It is important to note that if no measurements were obtained within the original 5 second interval, the aggregated dataset may contain missing data during the corresponding 10 second interval.

The dataset under analysis comprises six observables, namely:

- generated active power (ActivePower)
- generated current (CurrentL1)
- rotation per minute of the rotor (RotorRPM)
- rotation per minute of the high speed shaft at the generator (GeneratorRPM)
- pitch angle of the blades (BladePitchAngle)
- wind speed (WindSpeed)

As there are no deviations between the three current phases in our data, we simply choose one of them. Our expectation for the present turbine is a shift in control strategy as the wind speed changes. This shift includes transitioning from a low wind speed regime with variable rotation speed, to an intermediate regime with constant rotation, and finally to a rated region with constant rotation and produced power. The selection of these observables allows us to effectively analyze these operational state changes.

In our figures, we explicitly only show the wind speed u and the active power output P. To comply with confidentiality agreements, we rescale their values when displaying them. The wind speed u is divided by the nominal wind speed u_{nom} of the turbine. This yields the rescaled wind speed (RWS) already used in previous chapters. The active power output is divided by its maximal value. Hence, in the figures it only takes values between zero and one.

6.2 Theoretical Background

6.2.1 Correlation Matrix States

In order to automatically calculate the operational state of a turbine, we employ the method presented in Ch. 4. Pearson correlation matrices are calculated for non-overlapping time intervals of 30 minutes called epochs to obtain a time series of correlation matrices. These are clustered to find structurally different operational states and thereby a time series $s(\tau)$, which labels the current operational state for every epoch. It is easily extended onto all times t by $s(t) = s(\tau)$ for $\tau \leq t < \tau + T$, i.e. all times in an epoch receive the same state label as the epoch.

Here, we use for this calculation the channels presented in Sec. 6.1. Therefore, the results of the clustering in terms of centroid matrices look very similar to Fig. 4.6. In the context of the Langevin analysis it is more relevant to visualize the states in the power curve, i.e. in the dependency of the power output on the wind. This is shown in Fig. 6.1.

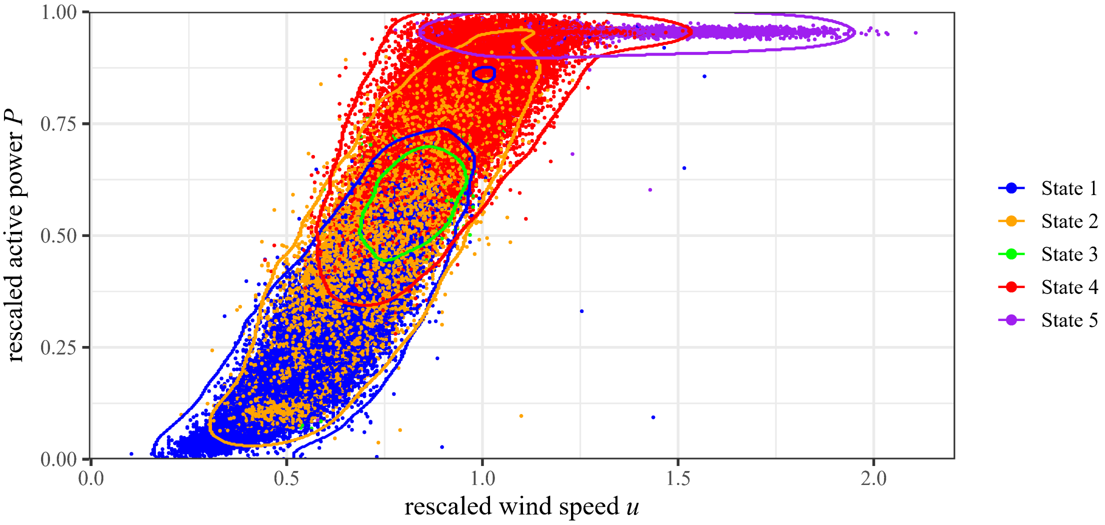


Figure 6.1: Rescaled active power P versus rescaled wind speed u. The color indicates the operational state as identified via clustering. Lines indicate the area of each operational state based on the density of data points (see App. C for details).

6.2.2 Estimation of the Kramers-Moyal Coefficients

The underlying methodology is the same as described in Sec. 5.1.2. We start with the traditional approach to model the power conversion process

$$\left.\frac{\mathrm{d}}{\mathrm{d}t} P(t)\right|_{u(t)=u} = D_P^{(1)}(P(t), u) + \sqrt{D_P^{(2)}(P(t), u)} \cdot \Gamma(t) \tag{6.1}$$

of a wind turbine in terms of stationary Langevin equation [191–193]. Here, the power output $P(t)$ is modeled as a one-dimensional stationary stochastic process for a fixed wind speed u. We assume a Gaussian distributed, delta correlated noise $\Gamma(t)$ with a mean value of zero and a variance of two. The analytical considerations are the same as in Sec. 5.1.2. However, again we only have a limited set of data and proceed analogously to Sec. 5.1.2.

Here, we consider a two-dimensional dataset $(P,\ u)$ from $t = 1$ to $t = T_{\text{end}}$ with an equidistant sample interval Δt. Furthermore, we define $\vartheta_q = q \cdot \Delta t$, where $q = 1, \ldots, Q$, allowing us to calculate the increments of the power output P over a time lag ϑ_q as

$$\Delta_{\vartheta_q} P(t) = P(t + \vartheta_q) - P(t). \tag{6.2}$$

To estimate the n-th conditional moment $M_P^{(n)}(P, u, \vartheta_q)$, we again employ the Nadaraya-Watson estimator with a two-dimensional D-kernel $K_{a,b}(y_1, y_2) = k_a(y_1) \cdot k_b(y_2)$ [196, 198, 204]. This D-kernel is represented as the product of two one-dimensional kernels. Effectively, the Nadaraya-Watson estimator functions as a means of calculating the weighted, data-driven average of increments. In our specific case the weights and increments are determined by the states P and u, as well as the kernel functions $k_a(y)$ and $k_b(y)$, along with the bandwidths for power output (h_P) and wind speed (h_u). For the first conditional moment, we calculate the weighted average of the power output increments using

$$\hat{\mathcal{M}}^{(n)}(P, u, \vartheta_q) = \sum_{t=1}^{T_{\text{end}}-\vartheta_q} \left(\Delta_{\vartheta_q} P(t)\right)^n \cdot \frac{\mathcal{K}_{a,b}\left(\frac{P(t)-P}{h_P}, \frac{u(t)-u}{h_u}\right)}{\sum\limits_{t=1}^{T_{\text{end}}-\vartheta_q} \mathcal{K}_{a,b}\left(\frac{P(t)-P}{h_P}, \frac{u(t)-u}{h_u}\right)}. \tag{6.3}$$

Different possible kernel functions were introduced in Sec. 5.1.2. Here, we employ an Epanechnikov kernel function

$$\kappa_{\mathrm{E}}(y) = \begin{cases} 1 - y^2 & , |y| \leq 1 \\ 0 & , |y| > 1) \end{cases}. \tag{6.4}$$

At least as important as the kernel function is the related bandwidth. For the analysis of large structures (macro-scale structures), large bandwidths should be used. However, with larger bandwidths, the small structures (micro-scale structures) are no longer visible. To estimate the Kramers-Moyal coefficient of our specific dataset, we used the bandwidths according to the IEC 61400-12-1 [205]. The bandwidth h_u for the wind speed is 1 m/s, and the bandwidth for the power h_P is 100 kW. These bandwidths should be adjusted based on the given dataset (larger bandwidths for a smaller dataset, smaller bandwidths for a larger dataset). For the dataset we used, we found that these bandwidths, in conjunction with the Epanechnikov kernel, yield sensible results.

We assume that the n-th conditional moments $\mathcal{M}^{(n)}(P, u, \vartheta_q)$ are linear for small time steps ϑ_q. We estimate the Kramers-Moyal coefficients

$$\hat{D}_P^{(n)}(P, u) = \frac{1}{Q} \sum_{q=1}^{Q} \frac{\hat{\mathcal{M}}^{(n)}(P, u, \vartheta_q)}{n! \cdot \vartheta_q}. \tag{6.5}$$

by averaging the n-th conditional moments divided by the used (small) time step ϑ_q times n factorial [200]. We choose $Q = 3$ in our calculations.

Furthermore, we can determine the fixed points $P_0(u)$ of the system [191]. These fixed points correspond to values of P at which the drift term becomes zero

$$\hat{D}_P^{(n)}(P_0, u) = 0, \tag{6.6}$$

indicating an equilibrium state. In order to assess the stability of these fixed points, we examine the derivative of the drift at the fixed point. If the derivative is negative, it signifies that the fixed point is stable:

$$\frac{\mathrm{d}}{\mathrm{d}P} \hat{D}_P(P_0, u) < 0 \tag{6.7}$$

The derivative of the drift at the fixed point plays a crucial role in understanding the stability of the fixed point as well as providing valuable insights into the mean reversal time.

When studying the stability of a fixed point, we are interested in how the system responds to small perturbations from its equilibrium state. The derivative of the drift provides information on the local behavior of the system near the fixed point. As said before, if the derivative of the drift evaluated at the fixed point is negative, it indicates that the fixed point is stable. In this case, any small disturbances from the equilibrium will eventually dampen out, and the system will return to its steady state. On the other hand, if the derivative is positive, it suggests that the fixed point is unstable, and even the slightest perturbations will cause the system to diverge from the equilibrium. Additionally, the derivative of the drift defines the mean reversal time of a system, i.e the average duration it takes a perturbed system to return to its equilibrium state.

6.2.3 Separation per Operational State

The operational states $s(t)$ of the wind turbine can only take discrete values in our analysis, specifically $s(t) \in [1, 2, 3, 4, 5]$. Furthermore, we consider that both the drift and diffusion coefficients depend on the turbine's operational state. By incorporating this additional condition, we reformulate the Langevin equation for the power conversion process to

$$\frac{\mathrm{d}}{\mathrm{d}t} P(t)\bigg|_{u(t)=u,s(t)=s} = D_P^{(1)}(P(t), u, s) + \sqrt{D_P^{(2)}(P(t), u, s)} \cdot \Gamma(t). \tag{6.8}$$

The numerical approach is derived analogously to the case without states. The only distinction is that we employ a 3-dimensional Kernel $\kappa_{a,b,c}(y_1, y_2, y_3) = \kappa_a(y_1) \cdot \kappa_b(y_2) \cdot \kappa_c(y_3)$. Due to the discrete values of the operational state, we can utilize a dedicated Boolean kernel function

$$\kappa_{\mathrm{B}}(y) = \begin{cases} 1 & y = 0 \\ 0 & y \neq 0 \end{cases} \tag{6.9}$$

We apply

$$\hat{M}_P^{(n)}(P, u, s, \vartheta_q) = \sum_{t=1}^{T_{\mathrm{end}}-\vartheta_q} \left(\Delta_{\vartheta_q} P(t)\right)^n \cdot \frac{\kappa_{a,b,\mathrm{B}}\left(\frac{P(t)-P}{h_P}, \frac{u(t)-u}{h_u}, s(t)-s\right)}{\sum_{t=1}^{T_{\mathrm{end}}-\vartheta_q} \kappa_{a,b,\mathrm{B}}\left(\frac{P(t)-P}{h_P}, \frac{u(t)-u}{h_u}, s(t)-s\right)} \tag{6.10}$$

to estimate the n-th conditional moment at a specific state (P, u, s). With these conditional moments we are able to obtain the Kramers-Moyal coefficients in a similar manner as shown above.

6.3 Results

In this section, we present the outcomes of our investigation into the wind turbine power conversion process using the Kramers-Moyal coefficients, considering both scenarios with and without separation per operational state. We focuson analyzing the drift and diffusion values governing the power output of a wind turbine. To deepen our understanding, we extend the analysis to include the computation of fixed points, their associated stability and the diffusion values at these fixed points. Thereby, we reveal the nuanced dynamics intrinsic to distinct operational states.

The calculated drift values of the power output, as depicted in Fig. 6.2, reveal familiar patterns observed in prior studies without operational state separation [191, 201, 202]. The top-left plot illustrates typical behavior of the power conversion process. Significant differences emerge when comparing drift maps for distinct operational states. A clear contrast is evident when comparing State 2 and State 4, particularly at rescaled wind speeds u (RWS) of approximately $0.7 - 0.9$. We also observe variations when comparing State 4 and State 5.

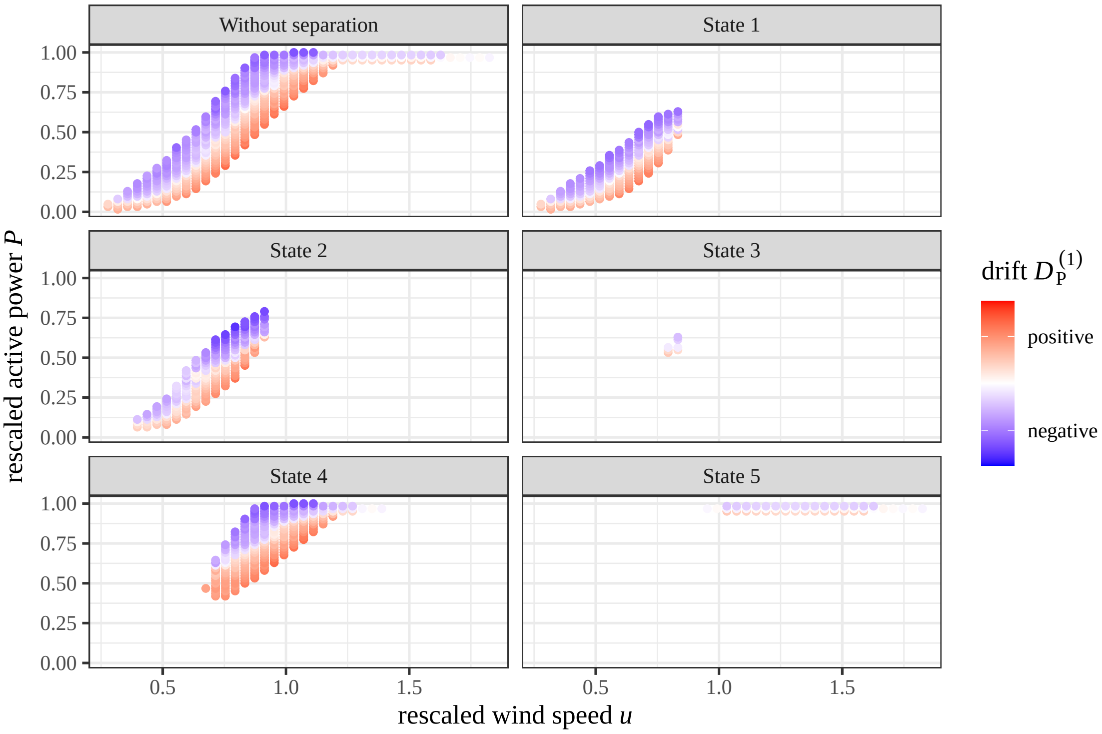


Figure 6.2: Color profile of drift $D_P^{(1)}(P(t), u, s)$ versus rescaled active power output P and rescaled wind speed u. The upper left panel shows the drift without differentiation per cluster. The others show the drift for each cluster separately.

To deepen our analysis we calculate stable fixed points and their derivatives. Figure 6.3 depicts stable fixed points per wind speed. We see disparities across different operational states, particularly at RWS values around $0.6-1.2$. Multiple stable fixed points are identified for a given wind speed, with States 1 and 2 displaying relative similarity. In contrast, significant differences are observed in other states, confirming the presence of hysteresis effects within the system dynamics [201, 203]. This also corroborates the results of Ch. 5. The absence of multiple fixed points per wind speed without operational state separation is attributed to the choice of a relatively high bandwidth during the estimation of the Kramers-Moyal coefficients. This, coupled with the use of a kernel function and the distribution of operational states, may have led to a more aggregated representation of the system dynamics. Here, also the effects discussed in Ch. 5 that led us to change the drift estimation might play a role.

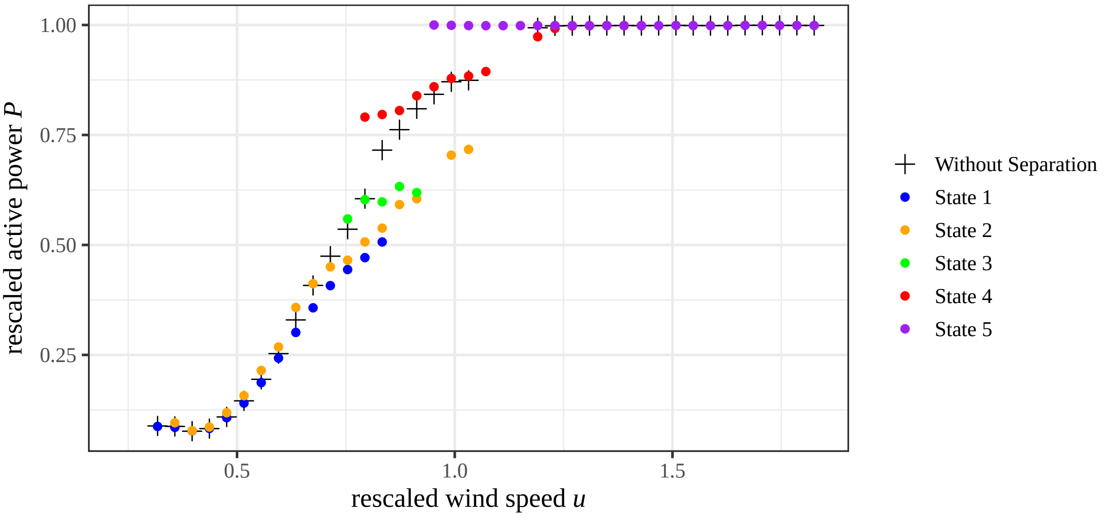


Figure 6.3: Rescaled fixed points P_0 of the active power output versus rescaled wind speed u. The colors distinguish different clusters as well as the solution without clusters.

We further explore the stability of these fixed points through the derivatives of the drift at the fixed points. Figure 6.4 illustrates the derivatives of the drift at the stable fixed points per wind speed. Negative values signify stable fixed points, with larger absolute values indicating a shorter mean reversal time towards the fixed point. Comparing derivatives for different operational states reveals similarities for States 1, 2, 3, and 4 across RWS values of approximately $0.0 - 1.1$. However, a significant change occurs for State 4 around RWS≈ 0.25, aligning it with State 5. At $RWS \approx 0.9 - 1.25$, State 5 shows distinctly lower values in the derivative than the other states.

We extend our analysis from the deterministic parts of the behavior the calculation of diffusion coefficients. The results are shown in Fig. 6.5. Without operational state separation, diffusion values are generally smaller near fixed points than further away from them. Small diffusion values are observed at rated wind speed (rescaled $u > 1.0$) and lower power values (rescaled $P < 0.3$). Differences in the diffusion values for different states are identified across various wind speeds. States 1 and 2 exhibit qualitative similarities. In Fig. 6.6 we display diffusion values at the stable fixed points conditioned on wind speed. Here, the diffusion values for the same wind speeds might represent different power values. States 1 and 2 exhibit similarity with a slightly steeper increase for State 2. Significant differences are observed across all other states, especially for RWS values between 0.75 and 1.2. The diffusion values

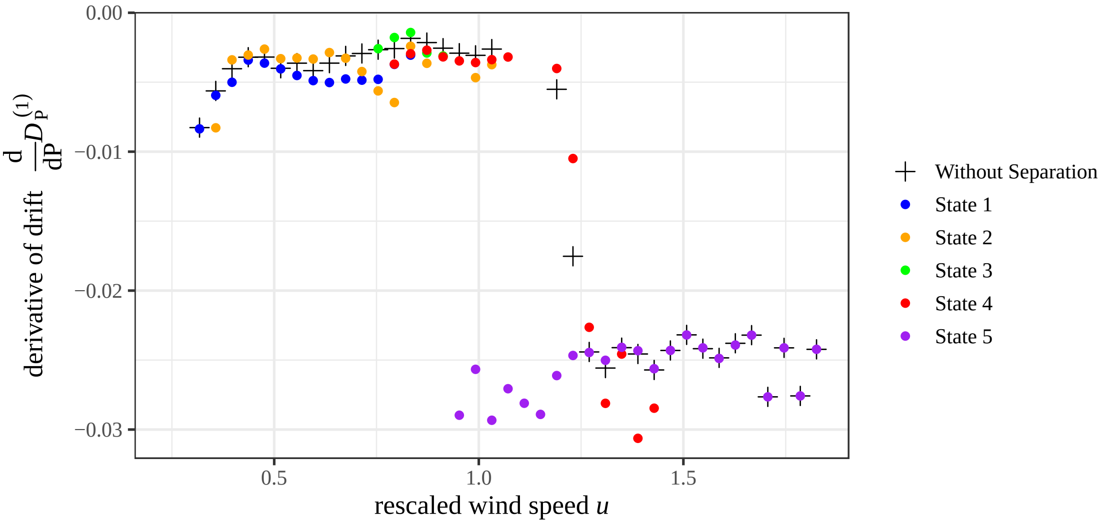


Figure 6.4: Derivative of the drift $\frac{\mathrm{d}}{\mathrm{d}P}D_P^{(1)}$ at the stable fixed points of the power output versus rescaled wind speed u. The colors distinguish different clusters as well as the solution without clusters.

are especially small for State 5 in comparison to other states.

The diffusion analysis further underscores distinctions in the behavior of State 5 compared to other states already seen in the derivatives of the drift. The smaller diffusion values for State 5, coupled with the reduced derivatives at the fixed points, contribute to diminished fluctuations around these stable fixed points of the power time series of State 5. In contrast, State 2 exhibits higher diffusion values at the fixed points for RWS values between 0.4 and 0.55 than the other states. Having higher diffusion values at the fixed points coupled with similarities in derivative values with States 3 and 5 result in higher fluctuations for these wind speeds in State 4.

6.4 Conclusions

In this chapter we successfully extended a method to estimate the dynamics of the power conversion process by taking non-stationarity into account. This was done by accounting for the different operational states identified according to Ch. 4. Our analysis revealed distinct dynamics associated with each operational state in the power conversion process. The operational states have a significant influence on the deterministic and stochastic behavior of the system. This highlights the importance of accounting for non-stationarity – here operational states – when studying and optimizing systems such as wind turbines.

We successfully resolved hysteresis effects within the power conversion process. When separating per operational state distinct fixed points per wind speed are visible. Without accounting for states, these are averaged out into one fixed point per wind speed.

The presented analysis also allows to identify differences in the dynamic behavior of states. State 5, representing rated power production, displayed a much more

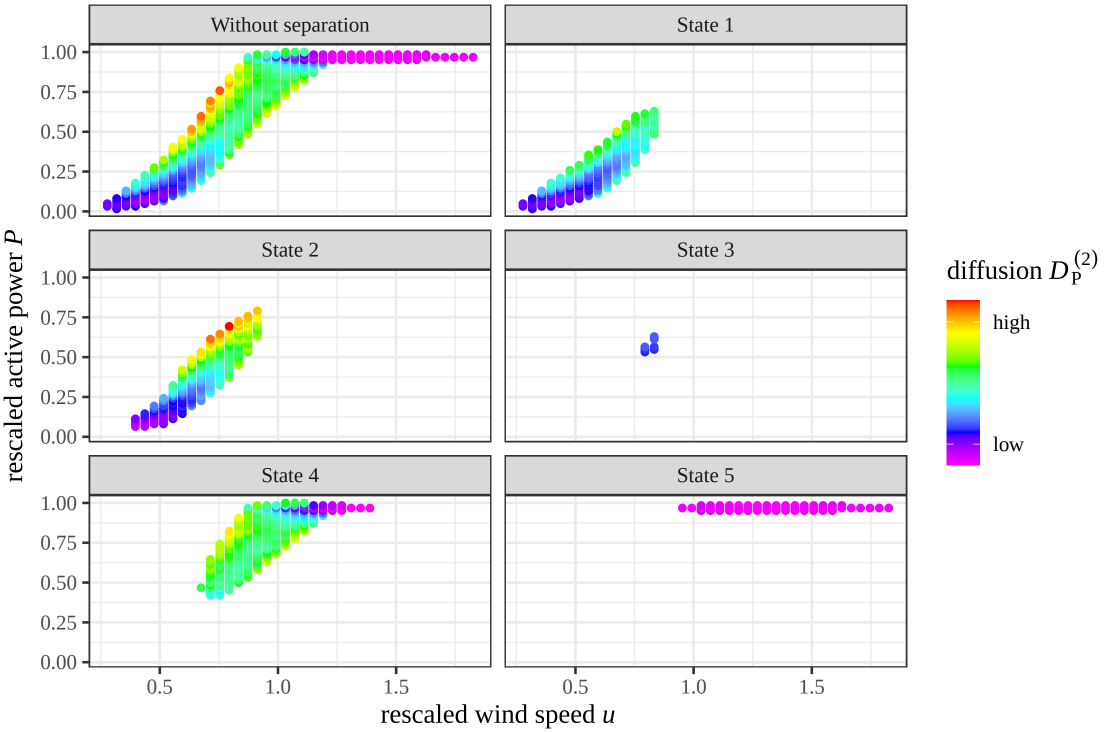


Figure 6.5: Color profile of diffusion $D_P^{(2)}(P(t), u, s)$ versus rescaled active power output P and rescaled wind speed u. The upper left panel shows the drift without differentiation per cluster. The others show the drift for each cluster separately..

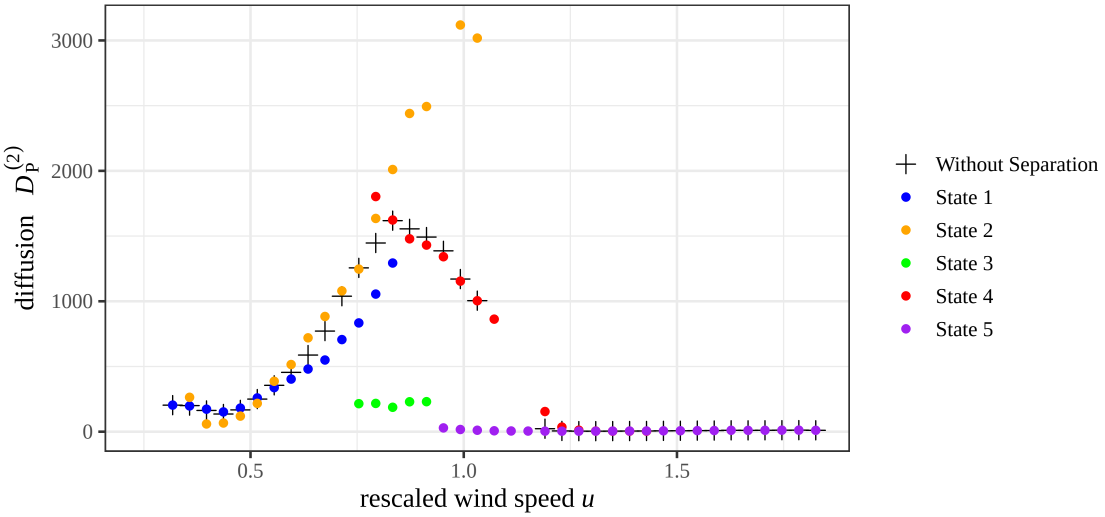


Figure 6.6: Diffusion $D_P^{(2)}(P_0, u, s)$ at the stable fixed points of the power output versus rescaled wind speed u. The colors distinguish different clusters as well as the solution without clusters.

stable behavior with less fluctuations than other states. This remained true even for wind speed values where State 5 overlaps with other states.

The results in this chapter clearly show that it is possible to enhance existing methods by considering the described operational states. The analysis concept does not need to change much, but rather only takes the automatically detected operational state as a distinction parameter for multiple subanalyses with the original method.

CHAPTER 7

Detecting Changes with Principal Components in the Presence of Non-Stationarity

While we have now shown that at least one established analysis method is enhanced by consideration of the found non-stationarity, we have not actually performed failure detection on wind turbines. To do this, we would need to be able to reliably label time periods as healthy and faulty in our data. If faulty, it would be even better to know the nature of the fault. As discussed in Sec. 2.7 this is problematic in our dataset. Therefore, we present in this chapter a modified version of paper [4], in which we study how non-stationarity influences change detection with PCA in general correlated systems. Thereby we also analyze the importance of non-stationarity for change detection in any given system, apart from providing an indication of how failure detection in wind turbines could be improved by the results from previous chapters. This is important as the detection of changes, novelty, failures or faults is a crucial task in numerous real world systems, such as industrial monitoring, energy generation, IT- and road-traffic, sensor networks, image processing and many more [154, 206–208]. Early or even preemptive detection can help operations, avoid down-times and reduce costs in general. Techniques for such analysis can be roughly categorized into three groups: model-based methods, knowledge-based methods and data-driven methods [151]. While a lot of research is done in all directions, the first two groups require extensive knowledge about the monitored system and are therefore challenging and time-consuming. This has made the data-driven methods especially interesting for researchers in all fields [147].

The problem of non-stationarity is also of on-going interest for other novelty detection methods [159–164].

Non-stationarity exists in different forms. First, a system can slowly evolve due to influences such as climate [161] or economic situation [42]. On the other side, external conditions can vary on shorter time-scales causing non-stationarity without an evolutionary trend [1, 51, 164]. Some systems also exhibit periodic non-stationarity [162]. A second distinction can be made in the type of non-stationarity: changes can occur in the absolute values of measured observables, their individual distribution or the relation between different observables. In simple terms these three categories represent novelty as changes in mean values, standard deviations or correlation structure respectively. Of course, this does not capture all possibilities. For example, distribution changes can also effect higher moments only. Furthermore, relations can be non-linear. However, this simplification already provided valuable insights in the sensitivity of principal components in stationary systems [158]. We study the importance of non-stationarity accordingly.

Of those three categories non-stationarity in the correlation structure is especially important for PCA (or any other method depending directly on the covariance or correlation matrix). The purpose of PCA is to find new coordinates that already incorporate the linear correlations in the system. If they change, i.e. the eigenvectors of the correlation matrix change their structure, the projections into PCA space are substantially different. Such a non-stationarity in the correlation matrix has been thoroughly investigated, for example, in financial markets [42, 43] and road traffic [51, 209]. In Ch. 4 we have found it also for wind turbine data. We assume that such a distinction into well-separated individual states is most likely to happen for a difference in correlation structure. At a certain point internal or external causes lead to a behavioral change. Changes in mean or standard deviation of observables might also exhibit such distinct states, but are more likely than a correlation structure to change continually. Therefore, our analysis will focus on correlation changes, but we will nevertheless study all types of change.

We aim with this study for a structured and comprehensive analysis of the sensitivity of principal components in the presence of non-stationarity, while trying to keep it simple enough that mechanisms can be easily understood. The non-stationarity is modeled as multiple possible normal states of the system. We study all three types of possible changes, but do not mix them. This means, if the change occurring at a certain point in time is of a certain type (mean value, standard deviation, correlation structure), the states will also differ in that same type. We base the study framework on the one used by Tveten [158] and extend it to incorporate non-stationarity. This facilitates comparison with the stationary case. We analyze the sensitivity of the principal components under the assumption that we know about the non-stationarity and compare it to the results obtained, if we did not know about the non-stationary behavior. Thereby, we can study how important knowledge about such a state-wise non-stationarity is for novelty detection with PCA.

The main focus of our study is a general analysis of the influence of non-stationarity on change detection with PCA as well as the presentation of a method to account for this non-stationarity. Additionally, we illustrate our simulation results with an example using real traffic flow data. Thereby, we show how the proposed methodology can be applied to measured data. In traffic, the non-stationarity is common knowledge: traffic volumes are not at all constant during a day. Furthermore, traffic is very different on workdays as compared to weekends or bank holidays. These well known facts make traffic a good example to understand the application of the method. We first classify the non-stationarity during single work days. Then, we define weekends and the onset of a bank holiday as a change and try to detect these with and without taking the classified non-stationarity into account.

We describe the idea of the experiment with a metaphor in Sec. 7.1. In Sec. 7.2 we introduce the problem and notations and set up the simulation framework in a general form. Section 7.3 contains the results of our simulation studies divided into sections for the different change types. In these we also present the detailed simulation set up for each case. We apply the method to real data by working out the traffic example in Sec. 7.4. A summary of results and findings is found in Sec. 7.5.

7.1 Analysis Concept

To facilitate easy understanding of the concept of our analysis before going into detail, we outline the simulation experiment with a simple example. Imagine a lamp with a light bulb that is burning with a certain color. Let us say it emits blue light. At a certain point in time $t = t_1$ the light bulb changes its color slightly and we study from which viewing angle we can best detect this change. As a measure we take a color chart and estimate the difference h in points between the color before and after the change. Then we repeat this experiment time and again with different bulbs, different colors and different color changes while keeping track of the change detectability from each viewing angle. This is an easily imagined example for the stationary simulation experiment.

Now, let us assume that before $t = t_1$ the light bulb burns with S different colors. At random intervals it changes its color, but all colors are equally likely over time. This represents non-stationarity in the way we are going to analyze. At time $t = t_1$ a change occurs once again. However, the light bulb can be in any state $s \in 1, ..., S$, i.e. burn with one of the possible colors. If we know of the non-stationarity and have a criterion to determine which color the lamp should have, we detect the change compared to the realized color s^*. Once again we use our color scale and determine a point score h_1 from different viewing angles. However, we want to analyze how much improvement in the detectability of that change is caused by our knowledge about this non-stationarity. Therefore, we also calculate a detectability without it: We compare the color after the change occurs with the average color before the state. So, if our lamp would normally burn with either blue or red, we compare to a violet light. Again we take the color chart and note difference h_2 in points between changed color and average color before change from each angle.

To measure the increase in detectability with knowledge of the non-stationarity we have to compare the two measurements h_1 and h_2. The first one is easy enough, because it should produce the same results as the stationary case if our knowledge about the original state, i.e. color, of the light bulb is perfect. When measuring without knowledge about the non-stationarity we have to take into account that even without a change we will detect deviations from our assumed normal (the average state): At any given time the light will be red or blue, but we will compare it to violet and observe a difference $h_{s,\text{norm}}$ between state s and the average state. This means that changes at $t = t_1$ are detectable only if they differ stronger from the violet than our normal states (blue and red) do. We have to correct our measurement of difference by subtracting the maximum difference we measure without a change present: $h_2^{\mathrm{corr}} = h_2 - \max_s(h_{s,\mathrm{norm}})$. As in the stationary case we perform multiple Monte-Carlo simulations. The importance of knowledge about non-stationarity for change detection is then measured as $h_1 - h_2^{\mathrm{corr}}$.

7.2 Theory and Setup

In our analysis, the light bulb from the example above is a K-dimensional system with T_{end} observations $\boldsymbol{X}(t) \in \mathbb{R}^K$ at times $t = 1, \ldots, T_{\mathrm{end}}$. We do not look at the time series themselves, but rather describe the system with mean values, standard deviations and Pearson correlations. This description is complete only for Gaussian systems. We assume this to be the case for our study. At the time $t = t_1$ a change occurs in the system, that can either influence mean, standard deviation or correlation structure of the observables. This is analogous to the change in color of the light bulb. For the stationary case the sensitivity to change in principal components was already analyzed by Tveten [158]. We extend it by the inclusion of non-stationarity. We assume that the system has S possible states and at every time is in one $s \in \{1, \ldots, S\}$ of them. For each of these states s we have a vector of means $\boldsymbol{\mu}_0^{(s)}$, a vector of standard deviations $\boldsymbol{\sigma}_0^{(s)}$ and a correlation matrix $C_0^{(s)}$. The index 0 indicates that these values refer to the system before the change at $t = t_1$ occurs. In our example with the light bulb these three values together describe the color of the state. For further analysis, we standardize each state separately to mean zero and standard deviation one and denote standardization with a hat. This results in $\hat{\boldsymbol{\mu}}_0^{(s)} = \mathbf{0}$ and $\hat{\boldsymbol{\sigma}}_0^{(s)} = \mathbf{1}$. Without knowledge about the S different states, two ways of measurement present themselves. One either measures just one set of parameters for the time span $t = 1, \ldots, t_1$ as was assumed in [158] or one measures the parameters in epochs of a length $T < t_1$ and takes the averages. The second option offers the advantage that new data, in which one wants to test for change, is most likely also only available for a time span $T_{\mathrm{new}} < t_1$. The comparability of new and old data is therefore increased if epoch length and time span of the new data coincide, i.e. $T \approx T_{\mathrm{new}}$. Correlations, for example, might be largely positive on a large time scale due to a global trend in values, but show more structure when measured on shorter time spans. Therefore we define the average state (analogous to the violet light of the light bulb example) as averages over epochs. Without loss of generality, we assume that each state s occurs for the same amount of time before the change and define the parameters of the average state as element-wise averages,

$$\overline{\boldsymbol{\mu}}_0 = \sum_{s=1}^{S} \boldsymbol{\mu}_0^{(s)}, \quad \overline{\boldsymbol{\sigma}}_0^2 = \sum_{s=1}^{S} \boldsymbol{\sigma}_0^{2(s)}, \quad \overline{C}_0 = \sum_{s=1}^{S} C_0^{(s)}. \tag{7.1}$$

We average the variance instead of the standard deviation. The average over the correlation matrix is also meant element-wise and is according to the centroid calculation in real data clustering in e.g. [42, 43, 51] or Ch. 4. Standardization with the mean value and standard deviation of this average state will be denoted with a tilde: $\tilde{\overline{\boldsymbol{\mu}}}_0 = \mathbf{0}$ and $\tilde{\overline{\boldsymbol{\sigma}}}_0 = \mathbf{1}$.

When the change happens at time $t = t_1$, it will affect the values we currently measure. In the light bulb example this would be the slight change of color. We denote these new values with an index 1 as $\boldsymbol{\mu}_1$, $\boldsymbol{\sigma}_1$ and C_1. Next, we need to decide how the data after the change should be standardized. In reality this questions is, how the new (or possibly live) data of a system should be standardized before comparing it to the usual behavior. This is independent of whether or not non-stationarity has been accounted for. One can either standardize with the known pre-change values of the usual behavior or with the newly measured ones. If one measures only a single new data point, only the first option is feasible. Without loss of generality, we normalize with pre-change values. However, the correlation matrix after the change will not be a well-defined correlation matrix. Especially the entries on its diagonal will not be one. This does not present a problem for our analysis.

Including non-stationarity in the way of multiple pre-change states, we need to know what normal state the system would have been in after the change at $t = t_1$ to make the right comparison. We will denote this state with s^*. In reality this means that after having found multiple normal states, we need one criterion (or possibly several criteria) marking the states. For the current, theoretical analysis we assume that we simply know the state s^* in the epoch after the change. We can then standardize with the pre-change state parameters to get $\hat{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\sigma}}_1$ and $\hat{C}_1$. The hat denotes standardization with state parameters. To compare the sensitivity of principal components with knowledge of non-stationarity to the case, in which we do not know about the different states, we will also calculate $\tilde{\boldsymbol{\mu}}_1$, $\tilde{\boldsymbol{\sigma}}_1$ and $\tilde{C}_1$, where the tilde denotes standardization with the average state parameters.

To analyze the sensitivity of principal components, we study the projections of the system onto the eigenvectors before and after the change. This is analogous to looking at the light bulb from different angles. For a state s we calculate the eigenvalues $\lambda_j^{(s)}$ and normalized eigenvectors $\boldsymbol{v}_j^{(s)}$, $j = 1, \ldots, K$ of the correlation matrix $C_0^{(s)}$. Ordering is assumed from largest to smallest eigenvalue, i.e. $\lambda_1^{(s)} \geq \cdots \geq \lambda_D^{(s)}$. Note that the eigenvectors are equivalent to the principal components as discussed in sections 3.3 and 3.4. The projection of a data point $\boldsymbol{X}(t)$ onto the j-th component is then calculated according to Eq. (3.18) with inclusion of the state as $X'_j(t) = \boldsymbol{v}_j^{(s)T}\boldsymbol{X}(t)$. Furthermore, we want to compare projections onto these components for epochs rather than a single data point. If we know the vector of means $\boldsymbol{\mu}$ and the correlation matrix C of said epoch, we calculate the mean μ'_j and standard deviation σ'_j of the projection onto the j-th eigencomponent by

$$\mu'_j = \boldsymbol{v}_j^{(s)T}\boldsymbol{\mu} \quad \text{and} \quad \sigma'_j = \sqrt{\boldsymbol{v}_j^{(s)T} C \boldsymbol{v}_j^{(s)}}. \tag{7.2}$$

We have dropped the state identification (s) on the left side. This is possible, because projections are only ever needed into the average system and the system of s^* so that no confusion with other states is possible. This allows us to keep the identifier, if it is actually the mean vector or the correlation matrix of the state, which is being projected. For example, the mean of projection onto component j of the of the original state mean $\boldsymbol{\mu}_0^{(s)}$ would read $\mu_{j,0}^{\prime(s)}$, whereas the projection of the changed state means $\boldsymbol{\mu}_1$ would simply be $\mu'_{j,1}$. Then also the overline of an average state would translate into an overline on the projected notation. If standardization notation is necessary it will also translate. For comparison without knowledge about the non-stationarity projections into the eigensystem $\overline{\lambda}_j$ and $\overline{\boldsymbol{v}}_j$, $j = 1, \ldots, S$ of the average state are also necessary. We denote these in the same way with a double prime,

$$\mu''_j = \overline{\boldsymbol{v}}_j^T \boldsymbol{\mu} \quad \text{and} \tag{7.3}$$

$$\sigma''_j = \sqrt{\overline{\boldsymbol{v}}_j^T C \overline{\boldsymbol{v}}_j}. \tag{7.4}$$

We now have all projections into the eigenvector system, i.e. onto the principal components, we need. In short, we have for every component j the projections of

- the original state into the correct state $\mu'^{(s^*)}_{j,0}$, $\sigma'^{(s^*)}_{j,0}$,
- the changed state into the correct state $\mu'_{j,1}$, $\sigma'_{j,1}$,
- the original state into the average state $\mu''^{(s^*)}_{j,0}$, $\sigma''^{(s^*)}_{j,0}$ and
- the changed state into the average state $\mu''_{j,1}$, $\sigma''_{j,1}$.

Having established the mean and standard deviation projections, we must now compare them. Therefore, we define the sensitivity for change as the Hellinger distance [210] between the marginal distributions before and after the change. In the light bulb example this is the difference in points on the color chart. In general, the squared Hellinger distance between two probability measures P and Q on the Lebesque-measurable space Z with probability densities $p(z)$ and $q(z)$ is defined as

$$H^2(P, Q) = \frac{1}{2} \int\limits_Z \left(\sqrt{p(z)} - \sqrt{q(z)} \right)^2 \mathrm{d}z = 1 - \int\limits_Z \sqrt{p(z) q(z)} \mathrm{d}z. \tag{7.5}$$

For our case, it is easily calculated for two normal distributions $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ as

$$H^2(\mu_1, \sigma_1, \mu_2, \sigma_2) = 1 - \sqrt{\frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp\left\{ -\frac{1}{4} \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \right\}. \tag{7.6}$$

For this theoretical study it is more feasible than a data distribution comparison such as Hotelling T^2-statistics. For the Hellinger distance to be applicable - under the assumption of normal distributions - we do not need the actual distributions. This allows us to simulate mean values, standard deviations and correlation matrices without having to generate the underlying data.

In Eq. (7.6) we have already adopted a notation that is easy to use for our application by writing the Hellinger distance as a function of the means and standard deviations of the normal distributions to compare. The projected μ and σ are calculated as described above and fully describe the comparison. For example, the Hellinger distance between the pre-change projection of the state s^* onto itself and the projection of the changed state into that same system with standardization with state pre-change parameters would read

$$H(\hat{\mu}'^{(s^*)}_{j,0}, \hat{\sigma}'^{(s^*)}_{j,0}, \hat{\mu}'_{j,1}, \hat{\sigma}'_{j,1}) = H_j(\hat{\mu}'^{(s^*)}_0, \hat{\sigma}'^{(s^*)}_0, \hat{\mu}'_1, \hat{\sigma}'_1). \tag{7.7}$$

For the right side we have simply taken the index j, which must always be equal for all arguments in H, and moved it from the arguments to the function for better readability.

We carry out different Monte Carlo simulations for changes in the correlation structure, the mean and the standard deviation. Each time we will assume that the other factors stay constant and the change type is the same for the occurring change at $t = t_1$ and in between the states. The detailed scheme for the simulation will always be given in the sections dealing with the results as these differ for the different change types. In general we will simulate normal states, calculate the average state and then simulate a change at $t = t_1$. We will always perform this for a multitude of normal states and for each of these for a multitude of change scenarios. The after-change data will be standardized and projected into the corresponding state eigensystem as well as the average eigensystem, Hellinger distances are calculated respectively. We then take the expectation value $\overline{H}_j$ over the Monte Carlo runs.

While in real data some fluctuations are always expected, which lead to non-zero Hellinger distances even without a change occurring, this is true for both the state and the average case and can therefore be neglected in a comparison. Remembering our light bulb example however, we have to correct the result without knowledge of the non-stationarity. In a non-stationary system with states there is an additional non-zero part of the Hellinger distance, if we do not know about the states. This is because the system would be in a state s^*, which is unequal to the average state even without change, but projected into the average eigensystem and compared to the average state. This would be additional noise in our change detection, that only occurs when not knowing about the non-stationarity. In our example this was the blue and red light already being different from the average violet color without a change. We therefore calculate the projections of the different states $s = 1, \dots, S$ into the eigensystem of the average state. Of course, to do this, we need to standardize the state observables with the pre-change average parameters. We then calculate the Hellinger distances this would cause even without a change. The maximum of these distances is the threshold distance up to which we have to assume that the system behaves normally. We can only detect a real change at $t = t_1$ if it causes a Hellinger distance larger than the maximum one the states themselves cause. This maximum is subtracted from the Hellinger distance between the system with a change and the average state as only the difference between these two measures the detectability of a change in a non-stationary system without knowledge about this non-stationarity. As detectability can never be less than zero, we set negative values to zero. The subtraction is done inside the Monte Carlo runs. This ensures that the actual noise created by the state average comparison is taken as it can vary strongly depending on how different the original normal states are. We denote this corrected Hellinger distance with $\mathcal{H}_j$.

7.3 Results

The focus of our analysis lies in the difference between knowing and not knowing about the non-stationarity of the system. We present our results separately for the different types of change. This way it is easier to describe the simulation procedure

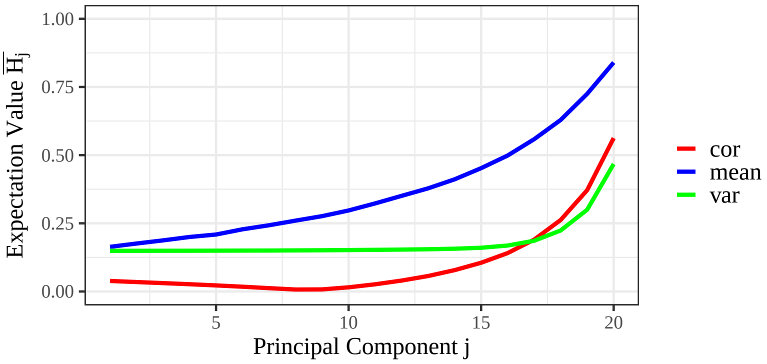


Figure 7.1: Monte Carlo estimates for the sensitivity to changes of the different eigenvectors in a stationary system. The sensitivities are the Hellinger distances between the distributions of data projections onto the eigenvectors before and after the change. Different types of changes are shown as different colors.

including necessary standardization. We have tested our implementation to produce the same results as given by Tveten [158] for the stationary case, i.e. if we assume the number of states S to be one. The results are shown in Fig. 7.1. It shows the sensitivity of the different eigenvectors to changes in the stationary system. As described in Sec. 7.2, the sensitivity is given by the Hellinger distance between the distributions of data projections onto eigenvectors before and after the change. A higher value of the expectation value for the Hellinger distance on the y-axis therefore indicates a higher sensitivity to changes. Clearly, the minor components, i.e. the eigenvectors with small eigenvalues, display a higher sensitivity to change. This type of plot is used frequently in the following analysis including non-stationary. The interpretation of the values on the y-axis stays the same. Here, we show results for the three different types of changes in one plot. The simulation procedures used for each different change type are described in the following sections.

We present the simulation and its results for correlation structure, mean value and standard deviation in Sec. 7.3.1, 7.3.2 and 7.3.3, respectively.

7.3.1 Change in Correlation Structure

We explore the sensitivity of principal components to a change in the correlation structure of the simulated observables. We think that this is the prominent use-case for PCA as changes in mean and standard deviation can also be detected by comparing their values directly. In contrast to the other scenarios, we do not have to worry about standardization issues as mean and standard deviation always stay the same. This is so, because we assume that the change does not influence these parameters and the normal states only differ in correlation structure. Hence, the vector of means $\boldsymbol{\mu}_0^{(s)}$ and the vector of standard deviations $\boldsymbol{\sigma}_0^{(s)}$ are equal for all states. We will therefore refrain from using the standardization notation in this case for the sake of readability.

To obtain results for different combinations of dimension K, number of normal states S and change sparsity Y, we perform Monte Carlo simulations with various change scenarios. Thereby we get the estimate $\overline{H}_j$ as an average over all simulation runs. We simulate two scenarios, which differ in the method to create the different states s that exist before the change. For one scenario they are random and unrelated to each other. For the second one we draw one random correlation matrix and obtain the other $S-1$ states by changing the first state in the same way that change is introduced at $t = t_1$ later on.

In the case of unrelated random states, our simulation follows the steps:

1. Draw S random correlation matrices $C_0^{(s)},\ s = 1, \ldots, S$ of dimension K using the method described in [211].
2. Calculate the element-wise average correlation matrix $\overline{C}_0$ before the change.
3. Calculate the Hellinger distances $H_j(\overline{\mu}_0'', \overline{\sigma}_0'', \mu_0''^{(s)}, \sigma_0''^{(s)}),\ j = 1, \ldots, K\ ,\ s = 1, \ldots, S$ between the occurring states and the average state, which gives the discussed base noise for detection.
4. Draw a change sparsity Y uniformly as an integer number between 2 and K. This gives the number of change affected dimensions.
5. Determine which dimensions are affected by randomly drawing Y integer numbers uniformly between 1 and K.
6. Randomly draw the normal state s^* the system is in after the change at time $t = t_1$ from the available states S states.
7. Draw a multiplicative change in correlation a uniformly between 0 and 1. Then multiply the correlations between all observables i and k from the affected dimensions $\mathcal{K}$ with this change a for $i \neq k$.
8. Calculate $H_j(\mu_0'^{(s^*)}, \sigma_0'^{(s^*)}, \mu_1', \sigma_1'),\ j = 1, \ldots, K$ between changed state and state, which gives the sensitivity for change detection with knowledge about the non-stationarity.
9. Calculate $H_j(\overline{\mu}_0'', \overline{\sigma}_0'', \mu_1'', \sigma_1''),\ j = 1, \ldots, K$ between changed state and average state.
10. Calculate the corrected Hellinger distance. This is done by subtracting the occurring additional noise (the maximum over all the distances between the states and the average state) from the Hellinger distance between changed state and average state:
$$\mathcal{H}_j(\overline{\mu}_0'', \overline{\sigma}_0'', \mu_1'', \sigma_1'') = \max(H_j(\overline{\mu}_0'', \overline{\sigma}_0'', \mu_1'', \sigma_1'') - \max_s(H_j(\overline{\mu}_0'', \overline{\sigma}_0'', \mu_0''^{(s)}, \sigma_0''^{(s)})), 0),$$
$$j = 1, \ldots, K.$$
This gives the sensitivity for change detection without knowledge about the non-stationarity.
11. Repeat steps 4 to 10 for 10^3 times.
12. Repeat steps 1 to 11 for 10^3 times.

In step 7. we only apply decreases in correlation. We do this to avoid many indefinite changed matrices [158]. One can easily imagine that a multiplicative increase could often lead to correlation coefficients larger than one. If any indefinite matrices still occur, we use Higham's algorithm [212] to find the closest positive-definite one. The results of such a simulation for $K = 20$ and $S = 3$ are seen in Fig. 7.2. As mentioned before, correction of the Hellinger distance for the change detection with the average state is done inside the Monte Carlo runs. To obtain the single green line representing the correction values, we average over the Monte Carlo simulations.

To compare states that emerged by applying changes to one random state, we simply change step 1. to the following substeps and present the results in Fig. 7.3:

a) Draw one random correlation matrix $C_0^{(1)}$ of dimension K using the method described in [211].

b) Draw a change sparsity Y uniformly as an integer number between 2 and K. This gives the number of change affected dimensions.

c) Determine which dimensions are affected by randomly drawing Y integer numbers uniformly between 1 and K.

d) Draw a multiplicative change in correlation a uniformly between 0 and 1. Then multiply the correlations between all observables i and k from the affected dimensions $\mathcal{K}$ with this change a for $i \neq k$.

e) Repeat steps b) to d) $S - 1$ times and use the changed correlation matrices as $C_0^{(s)},\ s = 2, \ldots, S$.

In both scenarios the sensitivity is greatest for the minor components in the changed state to state comparison. This is, of course, in accordance with the non-stationary results (see Fig. 7.1) as we always compare to the correct normal state. The changed state to average state distance is larger than the one to the actual state for major components and a crossing point between the two appears towards minor components. This point lies with larger j for the case of related original states. However, as pointed out before, we need to correct this Hellinger distance by the maximum distance between the actual states and the average state. This corrected Hellinger distance indicates the sensitivity of the change detection without knowledge about the non-stationarity. Its values are smaller than the ones with that knowledge for all principal components. In fact, they often lie below zero indicating no possible detection at all. The knowledge about non-stationarity greatly increases the possibility to detect changes.

For the case of unrelated states the blue and green line seem to be almost flat, indicating that all components possess the same sensitivity. This is an inherent feature of the averaging. If the correlation matrices of the states are all entirely random, the correlation structures tend to cancel each other out. Simply put, the off-diagonal elements of the average matrix tend towards zero. This results in meaningless eigenvector structures. This is further underlined by Fig. 7.4 showing the same results for $S = 7$. With more states the average matrix is closer to zeros on

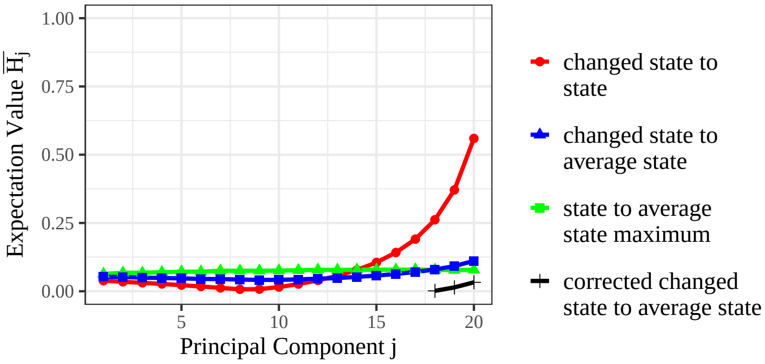


Figure 7.2: Monte Carlo estimates for the Hellinger distances of projections onto the different eigenvectors in presence of a change in correlation structure for $K = 20$ and $S = 3$. The original system states are created randomly and unrelated to each other. In color code the results for change sensitivity with knowledge of the non-stationarity (red), the uncorrected change sensitivity without knowledge (blue), the additional base noise induced by the system being in a state but comparing to the average state (green) and the corrected change sensitivity without knowledge (black) are shown. They correspond to the calculation steps 8, 9, 3 and 10 in the simulation, respectively.

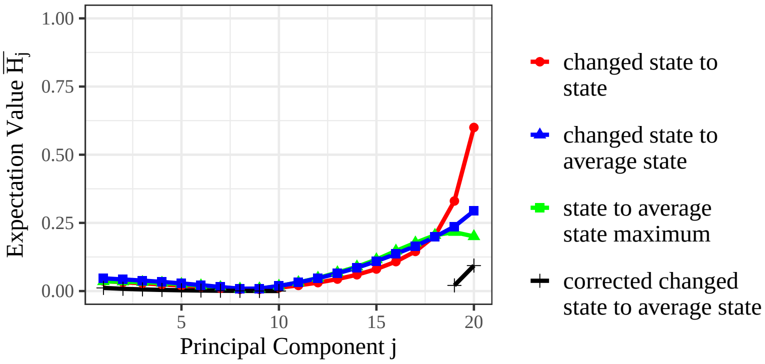


Figure 7.3: Monte Carlo estimates for the Hellinger distances of projections onto the different eigenvectors in presence of a change in correlation structure for $K = 20$ and $S = 3$. In the original system, one state is created randomly and the others are obtained from it by applying a change. In color code the results for change sensitivity with knowledge of the non-stationarity (red), the uncorrected change sensitivity without knowledge (blue), the additional base noise induced by the system being in a state but comparing to the average state (green) and the corrected change sensitivity without knowledge (black) are shown. They correspond to the calculation steps 8, 9, 3 and 10 in the simulation, respectively.

the off-diagonal and the blue and green line are even flatter. It is one of the reasons for the introduction of the scenario with related states. Another, more application oriented, reason is that many systems will not change or reverse their entire behavior, but rather change the behavior of certain groups of observables. In this case, we see in Figs. 7.4 and 7.5 that with an increased number of states the detection without knowledge about the non-stationarity becomes impossible, i.e. the corrected Hellinger distances is always smaller than zero. For $S = 3$ some detectability was given for minor principal components knowing just the average state, but this vanishes completely for $S = 7$.

In general, the increase in change detectability with knowledge about non-stationarity compared to without is measured by the difference between the changed state to state distance and the corrected value for the average state. Results for this difference are shown in Figs. 7.6 and 7.7. With non-stationarity present in the analysis, the minor components remain the most sensitive. Knowledge about the non-stationarity enables a more sensitive change detection for different numbers of normal states and different dimensions. As expected, the increase in sensitivity is larger for more states. It is also larger for unrelated normal states. Basically, these states are more different from each other than in the related case, so it is reasonable that the knowledge is more helpful here. We see, however, that for $S = 7$ for the largest j, where detectability is highest in general, the increase is larger for related normal states. As seen in Figs. 7.4 and 7.5 for $S = 7$ the corrected state to average state distance is always zero for both scenarios. So any difference in the sensitivity increase between the scenarios can only stem from a difference in the changed state to state distance. This difference exists purely due to a technicality: Because correlation coefficients cannot be larger than 1 the changes applied here are multiplicative between 0 and 1, i.e. they only reduce correlation. As this is also true for the changes performed to obtain the related states, the related normal states have weaker correlations on average than in the scenario with random states. This leads to a change of the sensitivity in the changed state to state scenario in Figs. 7.3 and 7.5. Here, the results are different from the stationary case, because the underlying set of matrices is changed. In general however, as long as different normal states exist, the knowledge about non-stationarity increases change detection sensitivity. As this is the main interest of the current study, we did not pursue this effect further, but it is interesting for future studies. Moving on, the difference in K does not change the overall results. For small j the increase is a bit larger for smaller dimensions, whereas a for large j it is the other way around. With the sensitivity for change being greatest for the minor components, the knowledge about non-stationarity is very important also in high-dimensional systems. It is noteworthy that an increase in the dimension of the matrix to $K = 100$ means that the relative change sparsity can be much higher as the smallest value of changed dimensions remains at two independent of K. Our results are also valid for sparse changes.

In summary, the knowledge about state-wise non-stationarity is important for the detection of changes. The sensitivity increase with knowledge is largest for the minor components, where change detection without non-stationarity is already most sensitive. For the major components the increase is still detectable, but much smaller. We could speculate that PCA for dimension reduction in a system is therefore still

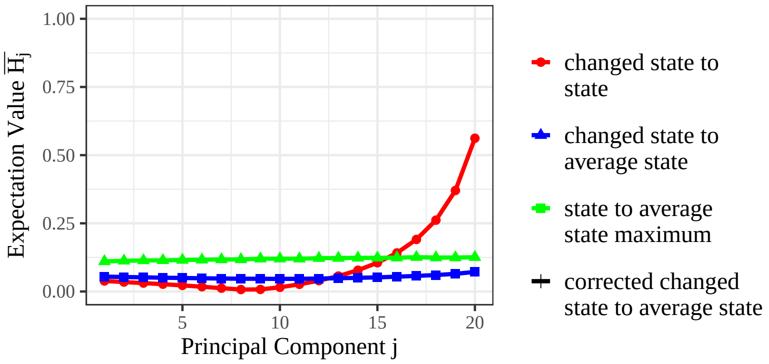


Figure 7.4: Monte Carlo estimates for the Hellinger distances of projections onto the different eigenvectors in presence of a change in correlation structure for $K = 20$ and $S = 7$. The original system states are created randomly and unrelated to each other. In color code the results for change sensitivity with knowledge of the non-stationarity (red), the uncorrected change sensitivity without knowledge (blue), the additional base noise induced by the system being in a state but comparing to the average state (green) and the corrected change sensitivity without knowledge (black) are shown. They correspond to the calculation steps 8, 9, 3 and 10 in the simulation, respectively.

quite possible without knowing about the non-stationarity, but this cannot be tested or verified in our setup and is not in the scope of this thesis.

7.3.2 Change in Mean Values

We now analyze the effect of changes in the mean values of observables. Again, we perform Monte Carlo simulations with various change scenarios to obtain results. The procedure in general is similar to the one used for changes in correlation, but the details change. We also simulate two different scenarios of normal states again: related and unrelated. In the case of unrelated, random states our simulation follows the steps:

1. Draw a random correlation matrix C_0 of dimension K using the method described in [211], which is the same for all states.

2. Draw S vectors of means $\boldsymbol{\mu}_0^{(s)}$, where each element is uniformly drawn as a non-integer value between -3 and 3.

3. Calculate the element-wise average vector of means $\overline{\boldsymbol{\mu}}_0$ before the change.

4. Assume standardization with average pre-change parameters and calculate $H_j(\tilde{\overline{\mu}}''_0, \tilde{\overline{\sigma}}''_0, \tilde{\mu}''^{(s)}_0, \tilde{\sigma}''^{(s)}_0)$, $j = 1, \ldots, K$, $s = 1, \ldots, S$ between the states and the average state, which gives the discussed base noise for detection.

5. Draw a change sparsity Y uniformly as an integer number between 2 and K. This gives the number of change affected dimensions.

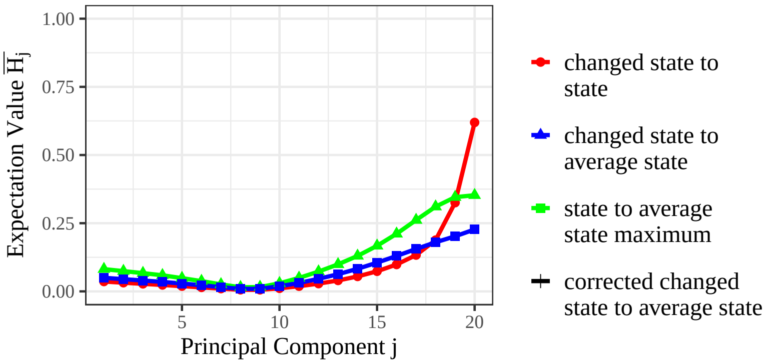


Figure 7.5: Monte Carlo estimates for the Hellinger distances of projections onto the different eigenvectors in presence of a change in correlation structure for $K = 20$ and $S = 7$. In the original system, one state is created randomly and the others are obtained from it by applying a change. In color code the results for change sensitivity with knowledge of the non-stationarity (red), the uncorrected change sensitivity without knowledge (blue), the additional base noise induced by the system being in a state but comparing to the average state (green) and the corrected change sensitivity without knowledge (black) are shown. They correspond to the calculation steps 8, 9, 3 and 10 in the simulation, respectively..

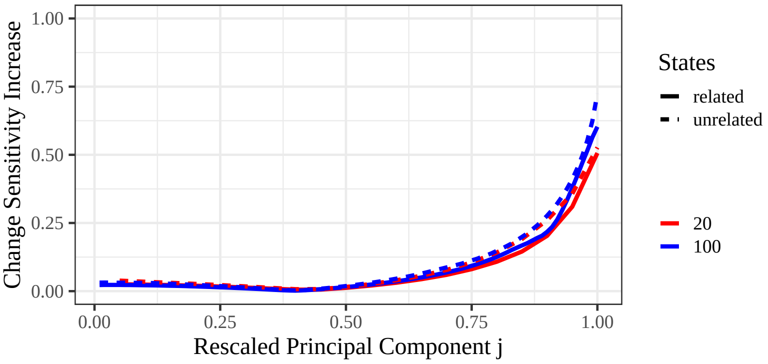


Figure 7.6: Sensitivity increase for a change in the correlation structure when knowledge about the state-wise non-stationarity is available compared to without that knowledge. Results are shown for different, color coded dimensions K, $S = 3$ and unrelated random states (simulation scenario 1) as well as related states, that were generated from one random state (simulation scenario 2).

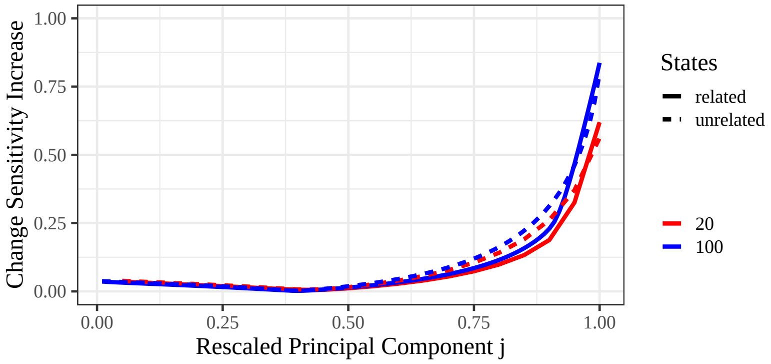


Figure 7.7: Sensitivity increase for a change in the correlation structure when knowledge about the state-wise non-stationarity is available compared to without that knowledge. Results are shown for different, color coded dimensions K, $S = 7$ and unrelated random states (simulation scenario 1) as well as related states, that were generated from one random state (simulation scenario 2).

6. Determine which dimensions are affected by randomly drawing Y integer numbers uniformly between 1 and K.

7. Randomly draw the normal state s^* the system is in after the change at time $t = t_1$ from the available states S states.

8. Draw an additive change in mean $\Delta\mu$ uniformly between -3 and 3. To obtain $\boldsymbol{\mu}_1$, i.e. the vector of means after the change, from the vector of means of the original state $\boldsymbol{\mu}_0^{(s^*)}$ the change value $\Delta\mu$ is added to the elements of the affected dimensions.

9. Assume standardization with known state pre-change parameters, i.e. $\hat{\boldsymbol{\mu}}_0^{(s^*)} = \mathbf{0}$ and

$$(\hat{\boldsymbol{\mu}}_1)_k = \begin{cases} \Delta\mu, & \text{if } k \text{ is an affected dimension} \\ 0, & \text{otherwise} \end{cases},$$

and calculate $H_j(\hat{\mu}_0'^{(s^*)}, \hat{\sigma}_0'^{(s^*)}, \hat{\mu}_1', \hat{\sigma}_1'),\ j = 1, \ldots, K$ between changed state and state. This gives the sensitivity for change detection with knowledge about the non-stationarity.

10. Assume standardization with average pre-change parameters, i.e. $\tilde{\overline{\boldsymbol{\mu}}}_0 = \mathbf{0}$ and

$$(\tilde{\boldsymbol{\mu}}_1)_k = \begin{cases} (\boldsymbol{\mu}_0^{(s^*)})_k - (\overline{\boldsymbol{\mu}}_0)_k + \Delta\mu, & \text{if } k \text{ is an affected dimension} \\ (\boldsymbol{\mu}_0^{(s^*)})_k - (\overline{\boldsymbol{\mu}}_0)_k, & \text{otherwise} \end{cases},$$

and calculate $H_j(\tilde{\overline{\mu}}_0'', \tilde{\overline{\sigma}}_0'', \tilde{\mu}_1'', \tilde{\sigma}_1''),\ j = 1, \ldots, K$ between changed state and average state.

11. Calculate the corrected Hellinger distance. This is done by subtracting the occurring additional noise (the maximum over all the distances between the

states and the average state) from the Hellinger distance between changed state and average state:

$$\mathcal{H}_j(\tilde{\bar{\mu}}''_0, \tilde{\bar{\sigma}}''_0, \tilde{\mu}''_1, \tilde{\sigma}''_1) = \max(H_j(\tilde{\bar{\mu}}''_0, \tilde{\bar{\sigma}}''_0, \tilde{\mu}''_1, \tilde{\sigma}''_1) - \max_s(H_j(\tilde{\bar{\mu}}''_0, \tilde{\bar{\sigma}}''_0, \tilde{\mu}''^{(s)}_0, \tilde{\sigma}''^{(s)}_0)), 0), \qquad j = 1, \ldots, K.$$

12. Repeat steps 5 to 11 for 10^3 times.
13. Repeat steps 1 to 12 for 10^3 times.

To have states that emerged by applying changes to one random state, we simply change step 2. to the substeps:

a) Draw one vector of means $\boldsymbol{\mu}_0^{(1)}$, where each element is uniformly drawn as a non-integer value between -3 and 3.

b) Draw a change sparsity Y uniformly as an integer number between 2 and K. This gives the number of change affected dimensions.

c) Determine which dimensions are affected by randomly drawing Y integer numbers uniformly between 1 and K.

d) Draw an additive change in mean $\Delta\mu$ uniformly between -3 and 3. To obtain $\boldsymbol{\mu}_0^{(s)}$, $s \neq 1$, i.e. the vector of means of a state s, from the vector of means of state 1 $\boldsymbol{\mu}_0^{(1)}$ the change value $\Delta\mu$ is added to the elements of the affected dimensions.

e) Repeat steps b) to d) $S - 1$ times and use the changed vectors of means as $\boldsymbol{\mu}_0^{(s)}$, $s = 2, \ldots, S$.

As we saw in Sec. 7.3.1 the most interesting result is the increase in sensitivity with knowledge about the non-stationarity compared to change detection without said knowledge. These simulation results are shown in Figs. 7.8 and 7.9. We see a clear increase in sensitivity for all simulated parameters. As before with changes in correlation structure, the knowledge about non-stationarity is slightly more important when the original normal states are related to each other and therefore not entirely different for $S = 3$. For $S = 7$ this difference is no longer visible. This is because the corrected Hellinger distances are always zero here and the changed state to state distances are not substantially different between the two scenarios. There is no influential change in the underlying set of states in contrast to the one seen in Sec. 7.3.1.

We again conclude that the non-stationarity is in general important for change detection. The increase seems to be more important for small j, i.e. major components, as compared to the results in Sec. 7.3.1. However, this is most likely due to the fact that the change sensitivity without non-stationarity already exhibits the same changes in their dependency on j (see Fig. 7.1).

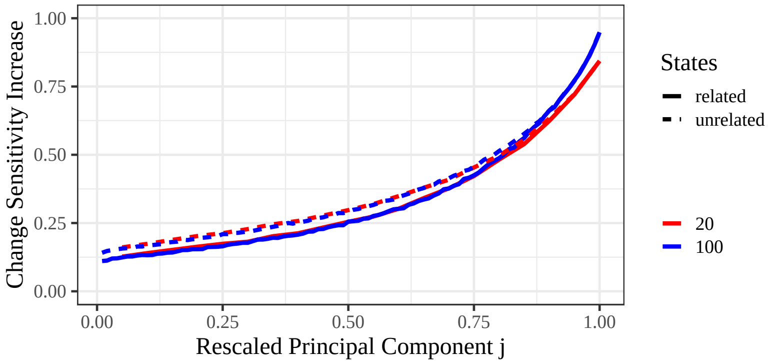


Figure 7.8: Sensitivity increase for a change in mean when knowledge about the state-wise non-stationarity is available compared to without that knowledge. Results are shown for different, color coded dimensions K, $S = 3$ and unrelated random states (simulation scenario 1) as well as related states, that were generated from one random state (simulation scenario 2).

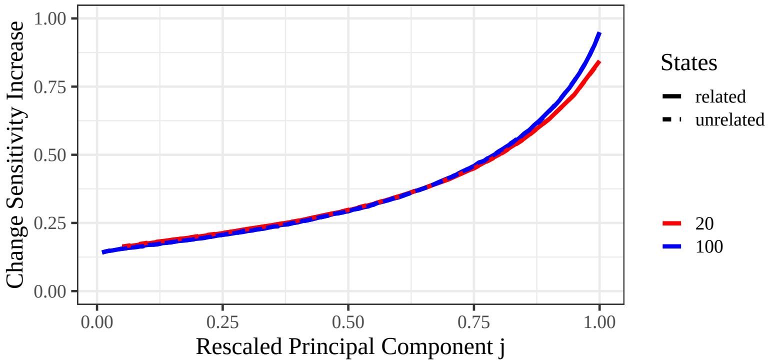


Figure 7.9: Sensitivity increase for a change in mean when knowledge about the state-wise non-stationarity is available compared to without that knowledge. Results are shown for different, color coded dimensions K, $S = 7$ and unrelated random states (simulation scenario 1) as well as related states, that were generated from one random state (simulation scenario 2).

7.3.3 Change in Standard Deviations

Finally, we analyze the sensitivity for changes in standard deviation in the presence of non-stationarity. The basic Monte Carlo process is the same, but once more the details are different. We simulate related and unrelated normal states again. In the case of unrelated, random states our simulation follows the steps:

1. Draw a random correlation matrix C_0 of dimension K using the method described in [211], which is the same for all states.
2. Draw S vectors of standard deviations $\boldsymbol{\sigma}_0^{(s)}$, where each element is uniformly drawn as a non-integer value between 1/3 and 2.
3. Calculate the element-wise average vector of variances $\overline{\boldsymbol{\sigma}}_0^2$ before the change. Then take the square root to get the average standard deviations. (We discussed in Sec. 7.2 that simply taking the average is probably closest to a real use case. In the case of non-changing means between the states and with every state appearing for the same amount time, this actually gives the correct variance over all states.)
4. Assume standardization with average pre-change parameters. Then $\tilde{C}_0$ is a well-defined correlation matrix with ones on its diagonal. To assume normalization with average parameters in the state and calculate $\tilde{C}_0^{(s)}$, we need the state covariance matrix. It is obtained by undoing the correct normalization of the state correlation matrix

   $$\Sigma_0^{(s)} = \operatorname{diag}\left(\boldsymbol{\sigma}_0^{(s)}\right) \hat{C}_0^{(s)} \operatorname{diag}\left(\boldsymbol{\sigma}_0^{(s)}\right) .$$

   It is then wrongly normalized with the average pre-change parameters

   $$\tilde{C}_0^{(s)} = \left(\operatorname{diag}\left(\overline{\boldsymbol{\sigma}}_0\right)\right)^{-1} \Sigma_0^{(s)} \left(\operatorname{diag}\left(\overline{\boldsymbol{\sigma}}_0\right)\right)^{-1} .$$

   $\tilde{C}_0^{(s)}$ is not a well-defined correlation matrix. We then calculate the Hellinger distance $H_j(\tilde{\overline{\mu}}''_0, \tilde{\overline{\sigma}}''_0, \tilde{\mu}_0''^{(s)}, \tilde{\sigma}_0''^{(s)}),\ j = 1, \ldots, K\ ,\ s = 1, \ldots, S$ between the states and the average state, which gives the discussed base noise for detection.
5. Draw a change sparsity Y uniformly as an integer number between 2 and K. This gives the number of change affected dimensions.
6. Determine which dimensions are affected by randomly drawing Y integer numbers uniformly between 1 and K.
7. Randomly draw the normal state s^* the system is in after the change at time $t = t_1$ from the available states S states.
8. Randomly decide if the change increases or decreases the standard deviation. Then, draw a multiplicative change in standard deviation $\Delta\sigma$ uniformly between 1 and 3 or between 1/3 and 1, respectively. $\boldsymbol{\sigma}_1$ is obtained by multiplying the elements of $\boldsymbol{\sigma}_0^{(s^*)}$ with $\Delta\sigma$ for the affected dimensions.

9. Assume standardization with known state pre-change parameters. Then $\hat{C}_0^{\prime(s^*)}$ is a well-defined correlation matrix with ones on its diagonal and $\hat{C}_1$ is calculated analogous to the description in step 4. Then calculate the Hellinger distance
$H_j(\hat{\mu}_0^{\prime(s^*)}, \hat{\sigma}_0^{\prime(s^*)}, \hat{\mu}_1', \hat{\sigma}_1'),\ j = 1, \ldots, K$ between changed state and state. This gives the sensitivity for change detection with knowledge about the non-stationarity.

10. Assume standardization with average pre-change parameters. Then $\tilde{\bar{C}}_0$ is a well defined correlation matrix and $\tilde{C}_1$ can be calculated analogous to the procedure in step 4. Then calculate the Hellinger distance $H_j(\tilde{\bar{\mu}}_0'', \tilde{\bar{\sigma}}_0'', \tilde{\mu}_1'', \tilde{\sigma}_1''),\ j = 1, \ldots, K$ between changed state and average state.

11. Calculate the corrected Hellinger distance. This is done by subtracting the occurring additional noise (the maximum over all the distances between the states and the average state) from the Hellinger distance between changed state and average state:
$$\mathcal{H}_j(\tilde{\bar{\mu}}_0'', \tilde{\bar{\sigma}}_0'', \tilde{\mu}_1'', \tilde{\sigma}_1'') = \max(H_j(\tilde{\bar{\mu}}_0'', \tilde{\bar{\sigma}}_0'', \tilde{\mu}_1'', \tilde{\sigma}_1'') - \max_s(H_j(\tilde{\bar{\mu}}_0'', \tilde{\bar{\sigma}}_0'', \tilde{\mu}_0^{\prime\prime(s^*)}, \tilde{\sigma}_0^{\prime\prime(s^*)})), 0),$$
$$j = 1, \ldots, K.$$

12. Repeat steps 5 to 11 for 10^3 times.

13. Repeat steps 1 to 12 for 10^3 times.

To have states that emerged by applying changes to one random state, we simply change step 2. to the substeps:

a) Draw one vector of standard deviations $\boldsymbol{\sigma}_0^{(1)}$, where each element is uniformly drawn as a non-integer value between 1/3 and 2.

b) Draw a change sparsity Y uniformly as an integer number between 2 and K. This gives the number of change affected dimensions.

c) Determine which dimensions are affected by randomly drawing Y integer numbers uniformly between 1 and K.

d) Randomly decide if new state has increased or decreased standard deviation. Then, draw a multiplicative change in standard deviation $\Delta\sigma$ uniformly between 1 and 3 or between 1/3 and 1, respectively. $\boldsymbol{\sigma}_0^{(s)},\ s \neq 1$ is obtained by multiplying the elements of $\boldsymbol{\sigma}_0^{(1)}$ with $\Delta\sigma$ for the affected dimensions.

e) Repeat steps b) to d) $S - 1$ times and use the changed vectors of standard deviations as $\boldsymbol{\sigma}_0^{(s)},\ s = 2, \ldots, S$.

The results for detection sensitivity increase due to knowledge about the non-stationarity are shown in Figs. 7.10 and 7.11. It seems to be small for a large number of components j, especially the major components. As was the case with a comparison between changes in mean and correlation structure, this is largely due to the different sensitivity in the stationary case (see Fig. 7.1). The curve

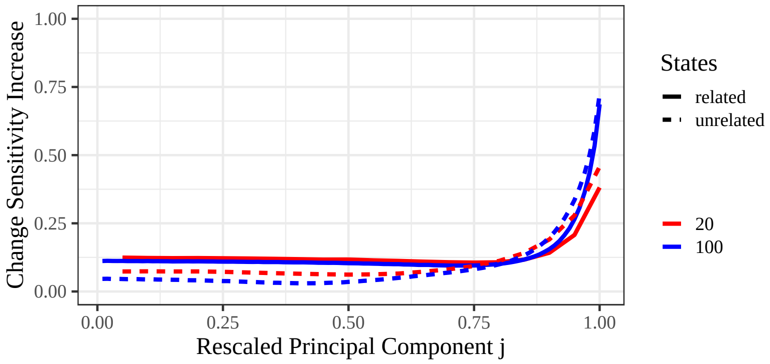


Figure 7.10: Sensitivity increase for a change in standard deviation when knowledge about the state-wise non-stationarity is available compared to without that knowledge. Results are shown for different, color coded dimensions K, $S = 3$ and unrelated random states (simulation scenario 1) as well as related states, that were generated from one random state (simulation scenario 2).

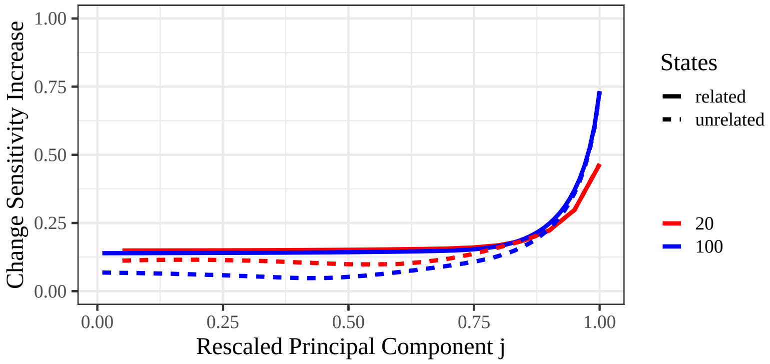


Figure 7.11: Sensitivity increase for a change in standard deviation when knowledge about the state-wise non-stationarity is available compared to without that knowledge. Results are shown for different, color coded dimensions K, $S = 7$ and unrelated random states (simulation scenario 1) as well as related states, that were generated from one random state (simulation scenario 2).

of the increase always shows a similar behavior over j as the stationary sensitivity itself shows. This underlines the importance of knowledge about non-stationarity as one would use these components for practical applications. Changes in standard deviation are almost impossible to detect for small j even in the stationary case. Therefore, our results cannot show a large increase in that regime. For large j we can again conclude that the knowledge about the non-stationarity of the system is quite important and becomes even more so if the original states are similar to each other. Again, as with results for changes in mean, the difference between related and unrelated states is smaller for $S = 7$ and even vanished for large j in that case. This is again due to a vanishing corrected Hellinger distance between state and average stat in this regime.

7.4 Application to Real Data: Traffic as an Example

To demonstrate how the method works, we give an instructive example with real data collected on German motorways. We look at 35 cross-sections (detectors) on the Cologne orbital motorway in 2015. The measured values are flows of vehicles, i.e. the number of vehicles passing a detector per time, in the counterclockwise direction. The time resolution of the data is one minute. As the orbital motorway consists of sections with different numbers of lanes, we do not look at a single lane, but rather at the flow accumulated over all lanes. The flow at time t at cross-section k is denoted $X_k(t)$. The dataset is described in more detail in [51].

This example is chosen because it allows a simple definition of ”normal” and ”non-normal”, i.e. changes in the system. These terms serve only as labels without further interpretation. We consider as normal the workdays Monday through Thursday as they present the typical rush hour behavior one expects from traffic. Fridays are excluded as the afternoon rush hour is stretched in time due to a strong variation in the ends of the working hours. Furthermore, we exclude the North Rhine-Westphalia bank holidays. Accordingly, we can afterwards try to detect weekend or holiday behavior as non-normal. Another advantage of this example is that everybody is aware of the non-stationarity of the system. Traffic is very different depending on the time of day. This means the non-stationarity in this system is not governed by a fluctuating observable but simply by the time of day.

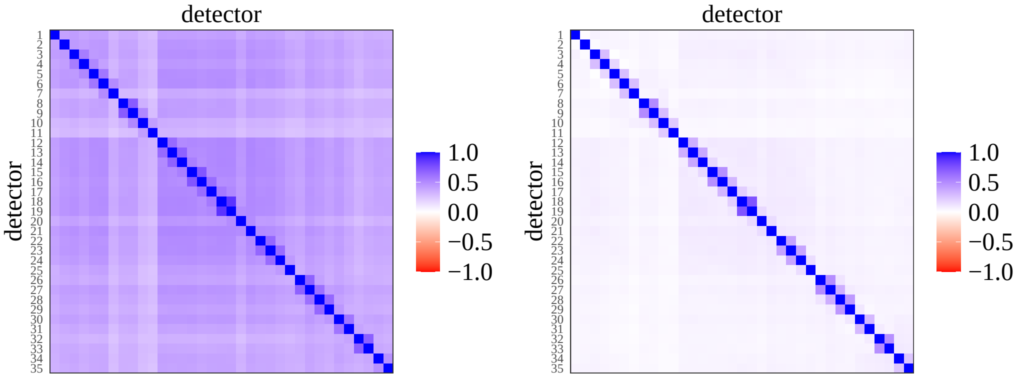
The first thing one has to do, to test the proposed method on real data is classify the non-stationarity in the data. There is, of course, not only one way to do this. Here, we proceed as described in Sec. 3.2 and Ch. 4. We calculate a time series of correlation matrices and apply clustering to them. We restrict ourselves to normal workdays and choose $T = 1\,\mathrm{h}$ for the epoch length. A day will thus contain 24 correlation matrices. Here, we choose to label the epochs with their end time instead of their beginning as in Sec. 3.2 and Ch. 4. The matrix for time $t = \tau$ is calculated from the data $X_k(t),\ k = 1, \ldots, 35\ ,\ \tau - T < t \leq \tau$. We write $|T|$ to denote the number of data points during this epoch.

For each epoch we proceed with the calculation of the correlation matrix time series as described in Sec. 3.2 to obtain $C(\tau)$. This matrix contains as the element at position i, j the Pearson correlation coefficient between $X_i(t)$ and $X_j(t)$ during epoch τ. In case of missing values in the data, we disregard all values of that time stamp to ensure the calculated matrix is a positive-definite correlation matrix. We define a Euclidean distance measure between two matrices in the epochs τ and τ'

$$d(\tau, \tau') = \sqrt{\sum_{i,j} (C_{ij}(\tau) - C_{ij}(\tau'))^2} = ||C^{(}\tau) - C^{(}\tau')||\ . \qquad (7.8)$$

analogous to Sec. 4.2.2 and apply k-means clustering [175] with $k = 2$ to all calculated matrices. This yields the cluster centers (element-wise mean matrices) shown in Fig. 7.12 and the distribution of the cluster appearances over daytime as shown in Fig. 7.13. Clearly, cluster 1 shows times where the traffic flow at all cross-sections is strongly correlated, whereas cluster 2 is mostly uncorrelated. Here, only close to the diagonal some correlations remain, which indicate similar flows on neighboring detectors. Splitting further yields only a separation of the strongly correlated cluster based on the strength of the correlations, no structural differences are detected. Furthermore, we applied also hierarchical k-means clustering [42], hierarchical clustering based on Ward's optimization criterion [213, 214], complete-linkage and single-linkage clustering [215] to the matrices. We have tested all algorithms for solutions of two, three and four clusters. We studied the resulting cluster centers and also calculated the silhouette coefficient [179] as defined in Eq. (4.5) for each matrix. It takes values between -1 and $+1$ with larger positive values representing matrices that are well clustered and negative values showing matrices that are closer to another cluster than to their own. An indicator for the overall clustering is given by the average silhouette coefficient. This is shown in Tab. 7.1 for all calculated clusterings. The hierarchical clustering with k-means splits and the one based on Ward's criterion yield very similar results to the standard k-means. This is not surprising as the optimization criteria for the algorithms are similar, but highlights that the found solution provides a reasonable grouping. The complete-linkage clustering gives a similar solution for two clusters, but starts to classify very small groups from then on. It is noteworthy that the two cluster solution has a higher silhouette coefficient than the k-means solution. However, the matrices and distribution over daytime show no significant differences for two clusters. As one would expect, the single-linkage clustering gives very different results, which do not separate the main groups as well, but rather lead to the detection of outliers. This is, of course, because the optimization criterion is very different from the others. We stick with the standard k-means solution for the following analysis. We have seen that its solution is reasonable and stable. It does not diverge towards outsider characterization for more than two clusters, which makes it a good first choice for this sort of classification also in other systems. In general, one has to tailor the identification of the non-stationarity to the problem at hand. We have found that correlation matrix clustering is useful in many systems [1, 42, 51].

A closer look at Fig. 7.13 allows interpretation of the clusters. The shown histogram counts reveal which cluster is dominant during which time of the day. The strongly correlated times in cluster 1 are caused by an overall increase (5am - 7am) and an overall decrease (8pm) in traffic volume. Another, less pronounced decrease is found during the 0am epoch. While the first two are caused by the major motion between low traffic flows during the night and high ones during the day, the latter is explained by traffic shifts from late evening traffic with low volumes to night time traffic with almost zero volume. As the 0am change is not as strong, cluster 2 remains dominant here. Generally, the appearance of cluster 1 marks transition periods between times with low and high traffic volume. Therefore, we will further split the two correlation clusters: Only consecutive times of a dominating correlation cluster are considered to be one cluster, the next appearance of the same correlation cluster is taken to be a new cluster. Thereby we achieve a good separation of times with different mean values, standard deviations and correlation structures.



(a) Cluster 1 (b) Cluster 2

Figure 7.12: Correlation matrix cluster centers calculated as element-wise means over all matrices sorted into a cluster. X-axis and y-axis both show the different traffic detectors, but labels were removed on the x-axis for better readability of the figure. Each matrix element is the mean Pearson correlation coefficient between the traffic flow signals of two detectors and its value is color coded.

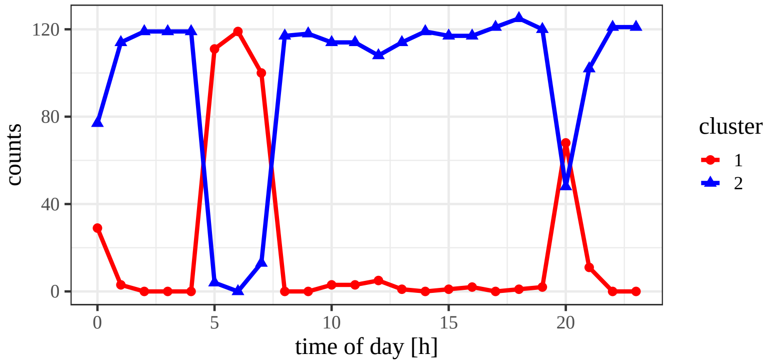


Figure 7.13: Histogram counts of the appearances of cluster 1 and 2 during the 24 hours of a day. The histogram is calculated over all used week days with a bin width of 1 hour.

We arrive at five clusters:

1. 0am - 4am,
2. 5am - 7am,
3. 8am - 7pm,
4. 8pm,
5. 9pm - 11pm.

Of course, we could merge the first and last cluster, but for simplicity we leave the division as it is.

Next, we project all the normal workday data points into the eigenvector systems of the corresponding cluster state analogously to the mean projection in Eq. (7.2).

Table 7.1: Silhouette coefficients for different clustering methods and solutions with 2, 3 and 4 clusters.

method	2 clusters	3 clusters	4 clusters
k-means	0.423	0.317	0.168
hierarchical k-means	0.423	0.354	0.327
Ward's criterion	0.379	0.330	0.168
complete-linkage	0.451	0.452	0.345
single-linkage	0.446	0.235	0.219

This yields the reference normal distributions for the calculation of the Hellinger distance.

We then choose a random normal workday and a random Sunday. The workday is needed to establish which Hellinger distances appear even without any changes due to fluctuations. Hence, each hour of the two days is projected into the corresponding eigenvectors and the resulting distributions are compared to the reference normal distributions by means of the Hellinger distance. Within each cluster s we get a maximum appearing Hellinger distance for the normal data H_s^{norm}. We define this as the threshold a test data point needs to exceed to be called a system change. This is a rather strict definition minimizing false positives for system changes or anomalies. For the weekend data we calculate H_s^{anom} once as the maximum (case (a)) and once as the mean (case (b)) over the daily Hellinger distances. Thus, we introduce the exceedance of the threshold

$$E_s = H_s^{\text{norm}} - H_s^{\text{anom}} \tag{7.9}$$

as a useful mesasure for detectability in cluster s. The overall detectability is then taken as the maximum over all clusters. This yields a detectability performance for each eigenvector. To determine if accounting for the non-stationarity is necessary, we perform the same analysis, but without clustering, i.e. there is only one cluster. The results for these calculations are shown in Figs. 7.14 and 7.15. Detection is obviously easier with knowledge of the non-stationarity. Without, some eigenvectors are not at all usable for detection (values below zero). This is especially true for case (b). With clusters and a careful choice of the correct eigenvectors, we detect the change not only in a single data point but in the average performance of at least one cluster. In general, all eigenvectors show smaller detection performance without clusters. This comparison is not always perfectly fair as there is no warranty that eigenvector j in the no cluster analysis corresponds to the same behavior as eigenvector j in the cluster analysis. This was neglected in the simulation study in previous chapters as it does not matter in the Monte Carlo average. In Figs. 7.16 and 7.17 we show the same type of analysis once more, but instead of a random weekend day, we have taken the start of the summer bank holidays. Here, the mentioned fact is even more evident: For the system as a whole detectabilities are higher with clusters, but for single eigenvectors it might not be so. For example, in this case eigenvector 22 in the analysis with clusters and eigenvector 22 in the one without clusters do not have the same structure, i.e. they represent different behavior patterns of the multivariate system.

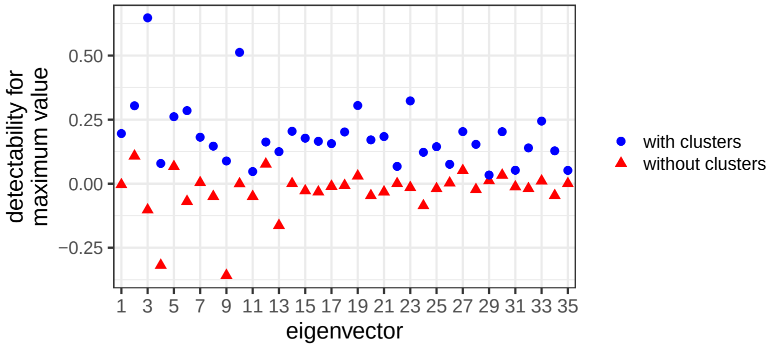


Figure 7.14: Detectability of a weekend day in case (a), where detectability is defined via a single exceedance of the detection threshold during the day. The detectability is shown for each eigenvector (principal component) of the system. Results are shown with knowledge about the non-stationarity during the day, i.e. with clusters, and without.

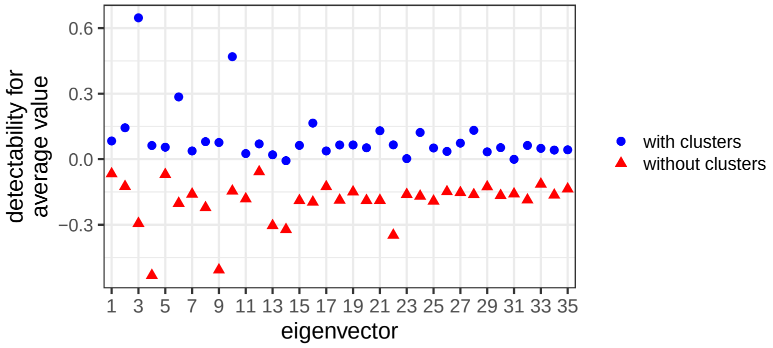


Figure 7.15: Detectability of a weekend day in case (b), where detectability is defined as an average exceedance of the threshold in one cluster. The detectability is shown for each eigenvector (principal component) of the system. Results are shown with knowledge about the non-stationarity during the day, i.e. with clusters, and without.

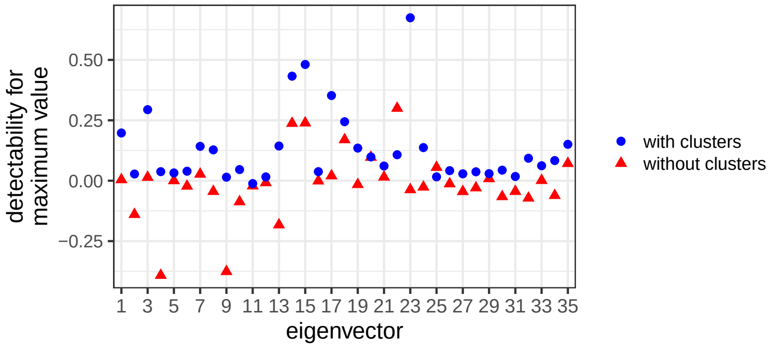


Figure 7.16: Detectability of the summer bank holiday onset in case (a), where detectability is defined via a single exceedance of the detection threshold during the day. The detectability is shown for each eigenvector (principal component) of the system. Results are shown with knowledge about the non-stationarity during the day, i.e. with clusters, and without.

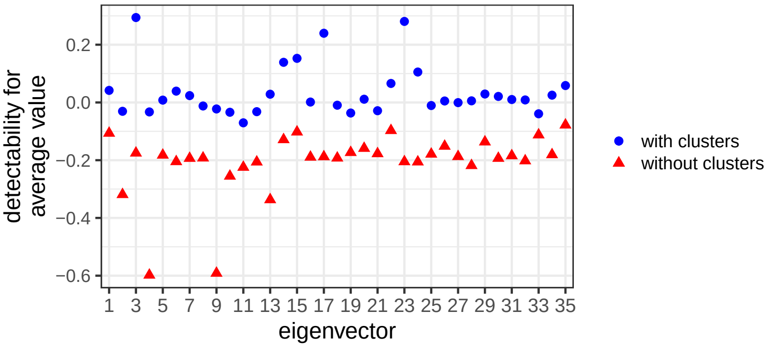


Figure 7.17: Detectability of the summer bank holiday onset in case (b), where detectability is defined as an average exceedance of the threshold in one cluster. The detectability is shown for each eigenvector (principal component) of the system. Results are shown with knowledge about the non-stationarity during the day, i.e. with clusters, and without.

All together, we see a much better change detection performance in this example when the non-stationarity is accounted for. Especially, if we do not want to detect the change only in single data points (case (a)), but rather with a consistently high indicator (case (b)), it is impossible to detect without clusters. This is inferred from all results for the detectability without clusters showing negative values in case (b), cf. Figs. 7.15 and 7.17. While of course in the case with clusters, less data points are used to calculate the mean per cluster, one can clearly identify the change in one of them. Without clusters there might be single points exceeding the detection threshold, but overall the Hellinger distances do not lie above the threshold.

The trend that changes are detected more easily in projections onto eigenvectors with small eigenvalues cannot be confirmed by a single example system. Which eigenvector is most suitable for detection depends on the interaction between correlation structure and system change. The aforementioned trend is therefore only true in the average over many systems.

7.5 Conclusions

We studied the sensitivity for change detection of principal components in non-stationary, correlated systems with multiple time series measurements. The non-stationarity was defined as the possibility for the system to be in multiple, distinctly different normal states prior to the change. Our study was based on the one conducted by Tveten [158] for stationary time series. Accordingly, the changes, which should be detected, were either to correlation structure, mean values or standard deviations of the time series. For simplicity, we constricted the study to those scenarios, where the normal states differ in the same property, in which the change also occurred. We analyzed how the detectability of the change varied for each principal component depending on the knowledge of the non-stationarity.

We found that in general the knowledge about the non-stationarity always increases the sensitivity for change detection. The increase is dependent on the principal component on which the data was projected. This dependency is quite similar to the one that was already found for the pure sensitivity for change detection in a stationary system. This means that where sensitivity for change is highest, also the increase gained by knowledge about the non-stationarity is highest. Usually this is for the minor components, i.e. the eigenvectors associated with small eigenvalues. This is reasonable as they represent behavior that is entirely unusual for a system in its normal state. When a change occurs, it can be seen very easily in those projections. If they are mixed up due to multiple normal states, this clear possibility to see changes is diminished. This underlines the importance of non-stationarity for direct uses of PCA for change detection and is confirmed in the traffic flow example.

Usage of PCA for dimensionality reduction by keeping only projections on major components will probably be less influenced by this. This is an interesting aspect to study in the future. Other methods depending on eigenvectors for change detection (e.g. Mahalanobis distance) were not directly studied, but we think it likely that their sensitivity to change could also be increased by the consideration of multiple normal states. Our results are true for all three different types of changes and states. We further analyzed two scenarios: Unrelated and related normal states. We found that the sensitivity increase is usually greater for unrelated states, which is reasonable as the non-stationarity has a stronger effect in this case. We want to point out once again, that the use of multiple normal states for change detection in real applications is only possible, if a criterion can be found to identify which normal states new data should be compared to. The purpose of the present study was mainly to develop the concepts and to provide the necessary tools. The traffic flow data example shows that it is possible to transfer the idea onto real world data. This opens up applicability in many systems, where PCA and related methods are used for change detection. While traffic is among these systems [153], other prominent examples for fault detection are chemical plants [216] and industrial machinery [116, 217]. A first step to include multiple operational conditions in PCA-based failure detection for heat pumps was already undertaken by Zhang et al. [218]. We intend to test the proposed method on wind turbines, where multiple operational normal conditions based on correlation matrix clustering have already been found [1] and PCA is being used for fault detection [117].

Our results clearly show that non-stationarity should be taken into account if one undertakes change, novelty or failure detection using principle component analysis.

CHAPTER 8

Wind Farm Correlations and Spatial Structures

So far, we have explored the correlations and non-stationarities of data from single wind turbines. Often however, wind turbines do not operate alone, but in wind farms consisting of many turbines. In this chapter we delve into statistical analysis of wind turbine data of whole wind farms, utilizing data from already installed SCADA systems as a simple and cost-effective data source. The chapter is a modified version of paper [5].

Wake effects and turbulences generated by the wind turbines are important for the layout of wind farms. Many sophisticated efforts are undertaken to describe the flows and incorporate them into wind farm design [24, 86]. Detailed physics based Computational Fluid Dynamics (CFD) models are needed to accurately describe the air flows within the wind farms. Models have been developed to calculate wake effects and turbulence generated by the wind turbines themselves [219, 220]. However, these models are often computationally costly and need detailed information on atmospheric flows as well as wind turbine and wind farm geometries. Therefore, we focus on a statistical analysis that is input-free beyond the measured data itself. Previously, data on wind conditions within the wind farm were often only based on single measured values from separate measuring masts and model calculations [221, 222]. Due to the availability of SCADA data, it is now possible to determine electrical power production, wind speed and orientation individually for each wind turbine [223–225]. Especially for large wind farms this might be a significant improvement for investigating wake effects, because the wind conditions may vary at different locations and the measured value of a single separate measuring mast may not be representative for the wind farm as a whole [226].

The interactions between wind turbines are relevant not only for the design but also for control and monitoring of wind farms. To ensure optimal operation, current operating data are evaluated and appropriate adjustments are made. For this, it is essential to have fast analysis procedures that can process real-time data. Computing time can be shortened, for example, by reducing the complexity of the data. This can be done by aggregating several wind turbines, based on wind speed, wind direction and wind farm layout, taking into account wake effects [227–229]. Another approach is to use correlation analyses by determining the correlation structure for the wind farm at one time and comparing it to the structure at other times. This was used in the examination of correlations for return time series in stock markets. Recurring, consistent structures were identified, that can be understood as states [42, 168, 169]. The dynamics can then be represented by the state time series. This significantly reduces the complexity of the system and allows for rapid analysis. Likewise, states in the correlation structure were found in analyses of highway traffic flow [51, 209, 230].

Our goal is to investigate the correlation structure of wind farms. We apply statistical methods to see which information can be gained on the interactions in the wind farm from the SCADA data. We do this without additional input or modeling assumptions. Our focus is on the electrical power production of individual wind turbines and the collective behavior within the wind farm. First, we calculate correlation matrices for the active power. Second, we apply clustering methods and find different states in the correlation structure.

This chapter is organized as follows: In Sec. 8.1 we describe the data used. In Sec. 8.2 we introduce our methods for the calculation of the correlation matrices and the further analysis. In Sec. 8.3 we show our results for the correlation analysis for the active power of the wind turbines for each wind farm. In Sec. 8.4 we summarize our results and provide an outlook.

8.1 Data

We analyze SCADA operational data of the German offshore wind farm Riffgat and the British offshore wind farm Thanet. We introduce the two wind farms and datasets briefly in Sec. 8.1.1, in Sec. 8.1.2 we describe our data pre-processesing.

8.1.1 Description of Wind Farms and Data

The wind farm Riffgat was the first commercial North Sea wind farm when it was built in August 2013 [231]. It is located about 15 km north-west of the island of Borkum [232]. The wind farm consists of 30 Siemens SWT-3.6-120 wind turbines with a rotor diameter of 120 m and a rated electrical power of 3600 kW each. Thus, the wind farm has a total rated electrical power of 108 MW. The wind farm layout is shown schematically in Fig. 8.1, top. It consists of three rows running from west to east. Each row consists of ten wind turbines. The wind turbine spacing within the rows is approximately 550 m (4.6 rotor diameters). The spacing between rows is approximately 600 m (5 rotor diameters). The analyzed operational data span a twelve month period from March 01, 2014 (one month after wind farm launch) to February 28, 2015. The data includes measurements at the wind turbines (e.g. wind speed, nacelle direction, pitch angle, generator speed), grid characteristics (e.g. active and reactive power of the wind turbine, grid voltage, current) as well as status values. The operational data are measured at high frequency, but recorded as the mean values, the standard deviations and minimum and maximum values of ten minute intervals.

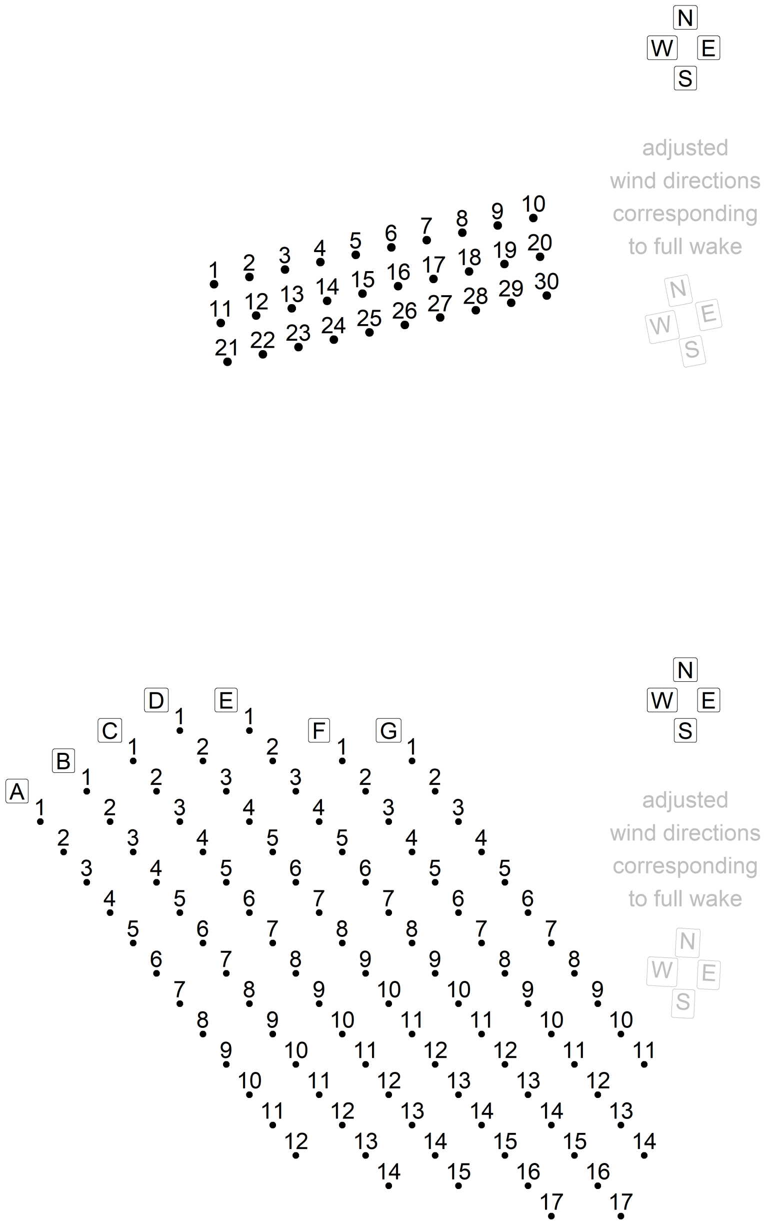


Figure 8.1: Top: Schematic layout of the wind farm RIFFGAT. Bottom: Schematic layout of the wind farm THANET. The adjusted wind directions are used in the later analysis to align the wind farms with the eight wind directions (N, NE, E, ...). For the wind farm RIFFGAT wind direction north corresponds to compass direction 349°N. For the wind farm THANET wind direction north corresponds to compass direction 3°N.

The THANET wind farm was built in 2010 and was one of the largest offshore wind farms in the world [233] at the time. It is located about 12 km off the coast of Kent in the south-east of the United Kingdom [234]. The wind farm consists of 100 VESTAS V90-3.0 MW wind turbines with a rotor diameter of 90 m and a rated electrical power of 3000 kW each. In total, therefore, the wind farm has a rated electrical power of 300 MW. The wind farm layout is shown schematically in Fig. 8.1, bottom. It consists of seven rows running from north-west to south-east. The wind turbine number per row varies between 11 and 17. The wind turbine spacing within the rows is approxiamtely 470 m (5.2 rotor diameters). The spacing between rows is approximately 720 m (8 rotor diameters). However, spacing may vary slightly for individual wind turbines. The various measured observables have different time resolutions. Observables such as wind speed and active power are measured in approximately 5 second increments. For synchronization, we average them to 10 second increments for our studies. Temperature measurements, e.g., of the generator, are made in 10 minute increments. Measured observables such as the nacelle direction are transmitted in each case as changes occur. The THANET dataset is very large due to its high temporal resolution. This has a significant impact on the handling of the dataset and the required computation times. Therefore, the analysis in this chapter is limited to the operational data for February 2017.

For both datasets, the data do not have entries for all observables at all times. The times without measurement vary per wind turbine and observable. Thus, there may be a wind speed measurement for a wind turbine at one point in time, but not its associated electrical power production. Missing points in the data are replaced by *NA* (Not Available).

8.1.2 Data Pre-processing

The data analysis requires preparation. We sort out and correct erroneous values as well as implausible data. We use different preparation steps for the low-resolution RIFFGAT dataset and the high-resolution THANET dataset.

The preparation of the RIFFGAT data is based on Refs. [235–238] and is done according to the following steps: First, we discard data (replace it by *NA*) for which the exact same measurement value was recorded for two consecutive measurements. Two exactly equal measured values are highly unlikely for a specification with such a high resolution (five decimal places). Therefore, we interpret this as a measurement error. Second, we discard all data where the standard deviation for the respective measurement interval has a value of zero. Last, we discard wind speeds above 30 m/s because they are well above the operating range of the wind turbines and are likely to represent measurement errors.

For the THANET dataset, data preparation does not involve any steps other than transforming it into the format used for our analysis, because of its high temporal resolution.

For the calculation of the correlation matrices, missing points in the data are a major problem. Hence, they have to be handled in an appropriate manner. For the RIFFGAT dataset all missing points in the data are replaced by the value 0 kW. We do this because the overall number of missing values is low (2.25 %) and the impact of this filling is therefore minimal. For the THANET dataset, the handling of missing points in the data is more complicated. First of all, the number of missing points is much higher (13.03 %). Furthermore, missing points in the data are not evenly (randomly) distributed across all wind turbines and the entire time period. We observe two main patterns. First, there are long periods with missing data points for individual wind turbines. This is due to failures of these individual wind turbines and is referred to below as *failed wind turbines*. Second, missing data points occur more frequently for multiple wind turbines at the same time when electrical power production and wind speed are low. This is due to shutdowns of the wind turbines due to insufficient wind speeds and is referred to as *wind turbines shut down* in the following.

Failures and shutdowns differ in particular in the fact that the former are individual effects and the latter are collective effects. Therefore, to characterize failures and shutdowns, we first consider the densities of missing data points in the time series of individual wind turbines, NA_{dens}, and the number of simultaneous missing data points in the time series of all wind turbines, NA_{farm}. To ensure proper characterization of failures and shutdowns, the following measures are taken: First, the density of missing data points in the time series of individual wind turbines, NA_{dens}, is determined over a large, sliding time window of twelve hours. This gives greater weight to the long duration of failures and less weight to short periods of shutdowns with many missing data points. Second, to remove collective increases in the density of missing data points, the deviation of the density of missing data points from the mean density of missing data points for all wind turbines is determined and then averaged for each wind turbine over a sliding time window of twelve hours

$$NA_{\text{dens,dev}} = \left\langle NA_{\text{dens}} - \frac{1}{N_{\text{WT}}} \sum_{\text{WT}} NA_{\text{dens}} \right\rangle_{12\,\text{h}} \quad , \tag{8.1}$$

with the number of wind turbines N_{WT}. For the failed wind turbines, we observe another effect in addition to the missing data points. Frequently, individual measured values occur between the missing data points, which deviate strongly from the measured values of all other wind turbines. In our opinion, these are erroneous measured values due to the failure of the respective wind turbine, which should also be replaced. To determine such values, the deviation, Ψ, of the individual wind turbine's electrical power production from the mean electrical power production of all wind turbines is averaged over a ten minute interval for each wind turbine

$$\Psi_{10} = \langle \Psi \rangle_{10\,\text{min}} \quad . \tag{8.2}$$

If a wind turbine produces significantly less electrical power on average than all other wind turbines over a time interval of 10 minutes, its measured values are considered erroneous. The criteria and thresholds chosen to characterize failures and shutdowns are explicitly summarized as:

1. A data point is interpreted as a failure,
   if $NA_{\text{dens}} > 0.6$ and $NA_{\text{dens,dev}} > 0.1$ or
   if $\Psi_{10} < -1000\,\text{kW}$.

2. A data point is interpreted as a shutdown,
   if $u < 4\,\mathrm{m/s}$ and $NA_{\mathrm{farm}} > 20$ and
   it is not characterized as a failure.

A missing data point in the wind speed measurement is considered as a wind speed $u < 4\,\mathrm{m/s}$ for the second condition. The thresholds for failure detection were chosen based on how well they were able to identify short- and long-term failures for some specific days. For the shutdowns, the threshold value for wind speed stems from the operations handbook of the wind turbines. The number of wind turbines is chosen as a fifth of the wind farm to reflect shutdowns as a collective phenomenon, the data of the turbines being shutdown the earliest are kept until at least one fifth of the wind farm is shut down. With this classification, of the 13.03 % missing data points, 3.85 % are due to shutdowns and 6.65 % are due to failures. The remaining 2.53 % cannot be assigned to either category.

Having identified failures and shutdowns, the question is how to handle them. Failures and shutdowns are fundamentally different and therefore their handling should be different as well. Shutdowns occur at low wind speeds and low electrical power productions. Consequently, the simplest option is to fill the shutdowns with values of 0 kW. This is consistent with the logic that no electrical power is produced during a shutdown. We neglect at this point that the wind turbines themselves need a certain electrical power for their operation and the missing data points would have to be filled with negative values accordingly. A second possibility is to fill the missing data points with the last existing value in each case. This is especially suitable for short series of missing data points. Both methods lead to the same qualitative results, but the correlations are stronger in the second case. Hence, for our analysis we use the second method. Failures occur for individual wind turbines. They do not depend on wind speed or electrical power production. Filling the failures with a fixed value such as 0 kW is therefore not useful. Using the last value before the failure is also unsuitable as an option, since on the one hand it is not ensured that this is not already faulty and on the other hand failures usually last for a longer period of time and this value can thus become as unsuitable as a fixed value of 0 kW. Thus, the simplest way to fill up the failures with a variable value is to use the respective current average electrical power production of the remaining wind turbines. We will do this in the following. Please note that the mean value of the wind turbines is calculated after the missing data points due to shutdowns have been filled using the aforementioned method.

8.2 Correlation Analysis

We analyze the correlation of the active power for all wind turbines of a wind farm. Therefore, in accordance with Sec. 3.1 $X_k(t),\ t = 1, \ldots, T$, is the time series of T measurement of an observable X for a wind turbine $k,\ k = 1, \ldots, K$, where K is the number of turbines and T the length of the observed interval. By following the further steps as described in sections 3.1 and 3.2 we calculate the correlation matrix C.

As discussed in sections 3.3 and 3.4, in the eigenvectors to a given eigenvalue, entries with the same sign and comparable numerical value, indicate coherent, i.e. collective, behavior of the corresponding time series. The largest eigenvalue of C and its eigenvalue often measure the collectivity of the system as a whole. Further large eigenvalues usually represent behavior of subgroups of observables [135]. In our analysis we find that the correlation matrices are dominated by collective effects, i.e., we find a large first eigenvalue with an eigenvector that consists of almost equal values in all entries. This is consistent with other complex systems [42, 43, 51, 168, 169, 209, 230]. To subtract the collective part, we use a singular value decomposition, which always exists [239] and has been used in a similar context by Heckens et al. [43]. The normalized data matrix M is decomposed via

$$M = U\Omega V^\dagger \text{ with } \Omega = \begin{bmatrix} \omega_1 & & 0 & 0 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & \omega_n & 0 & & 0 \end{bmatrix}. \tag{8.3}$$

Here U is an orthogonal $K \times K$ matrix with the eigenvectors of $MM^\dagger$ as columns and V is an orthogonal $T \times T$ matrix with the eigenvectors of $M^\dagger M$ as columns. The matrix Ω is a $K \times T$ matrix with the singular values ω on its main diagonal and all other entries being zero. The number of nonzero singular values corresponds to the rank of M. We notice that the $K \times K$ matrix $\Omega\Omega^\dagger$ and the $T \times T$ matrix $\Omega^\dagger\Omega$ contain the (non-negative) eigenvalues of $MM^\dagger$ and $M^\dagger M$, respectively. By setting one or more singular values to zero in Eq. (8.3), we generate reduced time series. For these we calculate the associated reduced-rank correlation matrices with Eq. (3.8). This procedure is equivalent to the construction of reduced-rank covariance and correlation matrices in Heckens et al. [43].

8.3 Results

We first examine the correlation structures of the active power for all wind turbines of each wind farm. Since these are strongly dominated by collective effects, we analyze the correlation structure for the active power without the collective part afterwards. Finally, we relate the correlation structure to the wind farm structure. We perform the same analysis on both wind farms. In Sec. 8.3.1 we present our results for the wind farm Riffgat. Afterwards, in Sec. 8.3.2 we present our results for the wind farm Thanet.

8.3.1 Riffgat

To obtain a first impression of the correlation structure, correlation matrices for the active power for periods of half a day, one day and one week are given in Fig. 8.2. We do not use shorter time windows at this point, as the data basis is too small to obtain reliable results for these due to the coarse temporal resolution of the data. For the correlation matrices shown, it is striking that, besides for Turbine 27 in the matrix over half a day, only positive correlations occur. Along the main diagonals, blockwise associated structures of slightly increased correlations are formed. However, single wind turbines show significantly lower correlations to all

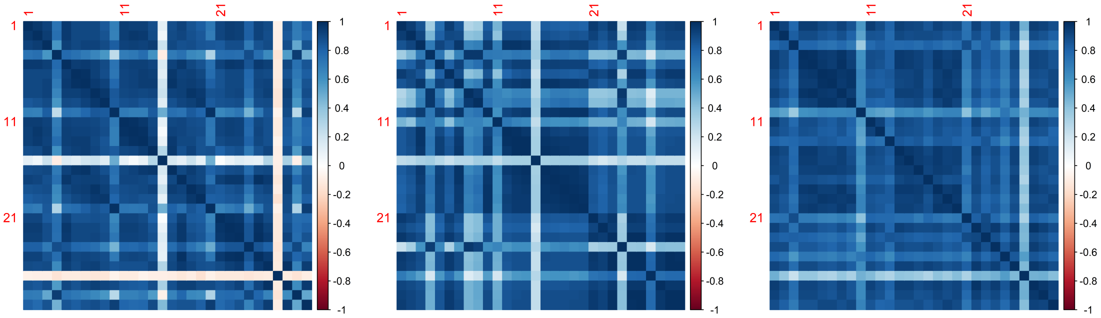


Figure 8.2: Correlation matrices for the active power over half a day (left), a day (middle) and a week (right) for the wind farm Riffgat.

other wind turbines, forming stripes of lower correlations. Overall, the correlation matrices are nevertheless dominated by a uniform positive correlation. The mean correlation increases with the length of the observation period. This is plausible if one considers the common dependence of each power series on one external trend: the wind speed. For long time scales, the wind speed is not constant and the power of all turbines changes strongly in dependence of this. Thus, strong positive cross-correlations are created on these time scales due to trends in the wind speed. With shorter time scales the wind speed varies less strongly allowing for individual variations of the wind turbines to be observed in the correlations. Consequently, the overall correlation for the active power of the wind turbines decreases and shows more interesting structure. For our further analysis of the Riffgat data we only consider correlation matrices over half a day.

The collective behavior observed for the correlation matrices can be illustrated by an analysis of their eigenvalues. In Fig. 8.3, the eigenvalue spectrum and the corresponding eigenvectors of the correlation matrix are shown. In the eigenvalue spectrum, a particularly large eigenvalue with a value of 24 stands out. The associated eigenvector has, with some exceptions, approximately the same value in each component (for each wind turbine). This provides a uniform correlation and indicates a collective behavior for the active power of the wind turbines.

To remove the influence of the collective dynamics, we perform a singular value decomposition of the correlation matrices and take out the contribution of the first singular value/eigenvalue (see Sec. 8.2). Correlation matrices for the active power reduced by the contribution of the first eigenvalue are shown in Fig. 8.4. Now, also negative correlations appear in the correlation matrices. In the first matrix, there are positive correlations in blocks along the main diagonal and otherwise mixed positive and negative correlations. The other two correlation matrices show structures that can be interpreted based on the spatial structure of the wind farm. Positive correlations are found between the first and last ten wind turbines (rows), respectively, in conjunction with negative correlations towards the remaining wind turbines. For correlation matrices other than the ones shown here, positive correlations between the first three wind turbines of each row (1, 2, 3; 11, 12, 13; 21, 22, 23) together with strong negative correlations of these with the last wind turbines of each row

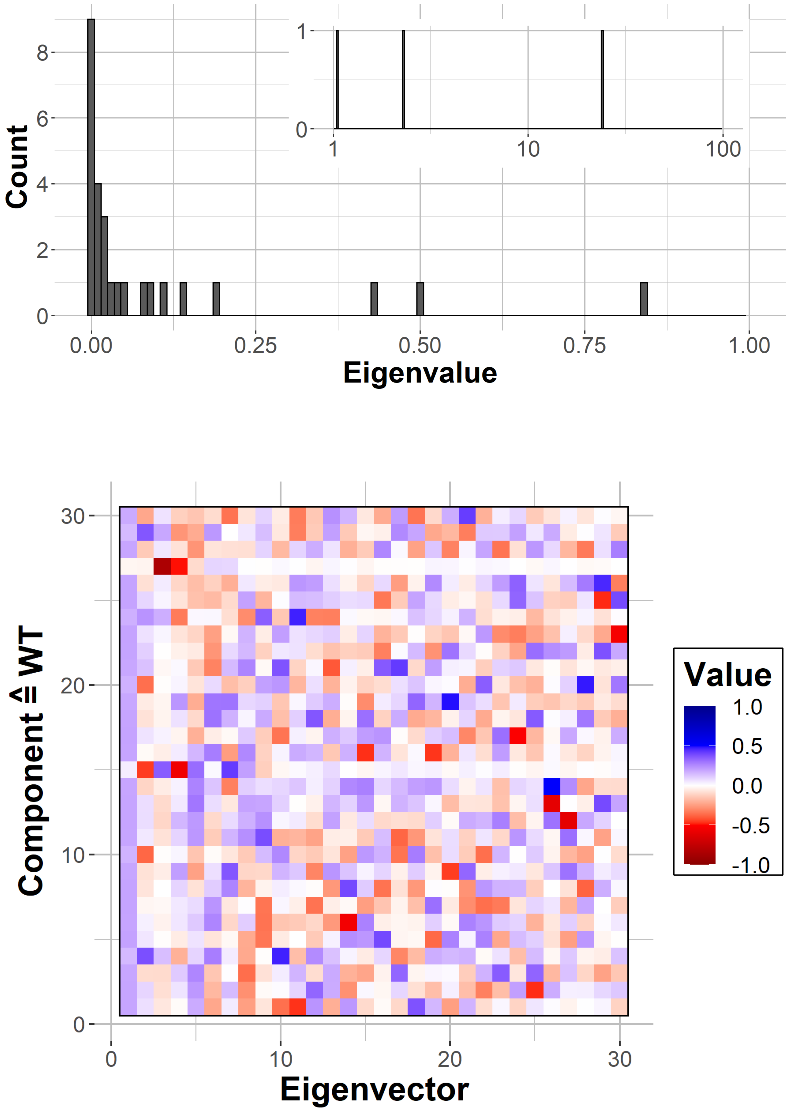


Figure 8.3: Top: Eigenvalue spectrum as histogram counts of the correlation matrix for the active power for half a day. The inlay shows the large values, while the main plot is zoomed in on small eigenvalues. Bottom: Corresponding eigenvectors of the correlation matrix for the active power. Each column represents an eigenvector. From left to right the corresponding eigenvalue decreases. For numerical values of the entries, see color code.

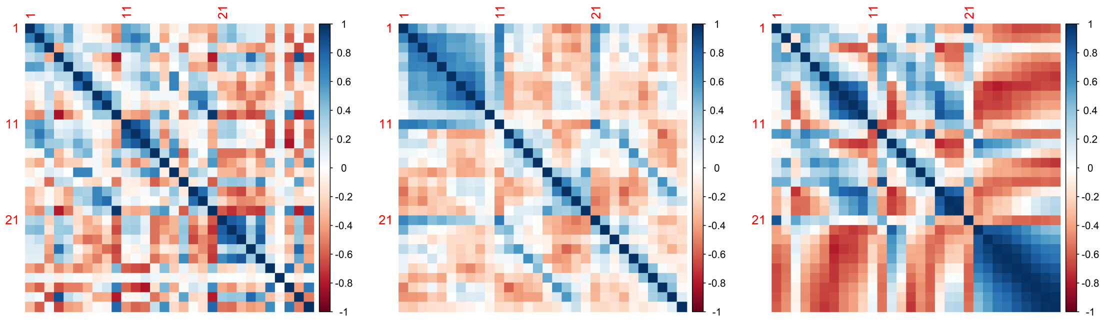


Figure 8.4: Correlation matrices for the active power without the contribution of the first eigenvalue for three different times for the wind farm Riffgat.

(10; 20; 30) also occur. These structures can be interpreted based on the wind farm structure. Depending on the wind direction, different wind turbines form the foremost line, which is the front of the wind farm in the wind. The fluctuations of the active power of these wind turbines are strongly correlated and anti-correlated to the fluctuations of the active power of the wind turbines of the other rows. It should be noted, however, that the wind direction measurement is slightly different for all wind turbines, and can also change significantly over the observation periods of half a day. Thus, it is only approximately possible to assign a wind direction to each correlation matrix.

In order to investigate whether for these long time periods a dependence of the correlation structure on the wind direction can be found in the correlation matrices, we grouped them as a function of the mean wind direction of all wind turbines over the respective observation period. The mean wind direction is calculated as the circular mean of all wind directions. The wind direction is divided into eight ranges of 45° each. The ranges are divided so that they are centered around the wind directions that are orthogonal on the wind farm rows. The mean correlation matrices for the 45° ranges of the wind direction, calculated elementwise, are shown in Fig. 8.5. There are clear structural differences between the mean correlation matrices of the wind directions. For the north direction, the northern wind turbines 1-10 show a strong positive correlation. The correlation for the first and second half of the row is stronger among each other than across the row. The correlation with the other wind turbines is significantly lower, often even negative. Thereby, wind turbines 1-5, 11-15, 21-25 and wind turbines 6-10, 16-20, 26-30 are respectively more strongly (positively) correlated with each other. Wind turbines 11-20 and wind turbines 21-30 show a higher correlation with each other than with wind turbines 1-10. An analogous structure is even more pronounced for the south direction. Here, wind turbines 21-30 show exclusively negative correlations with the other wind turbines. For the east direction, wind turbines $10, 20, 30$ are strongly correlated with each other and show anti-correlations with the other wind turbines. Similarly, wind turbines $1, 11, 21$ are correlated with each other for the west direction. Overall it can be seen that the wind turbines that are directly downwind have high correlations with each other and have a weaker correlation or anti-correlation with the other

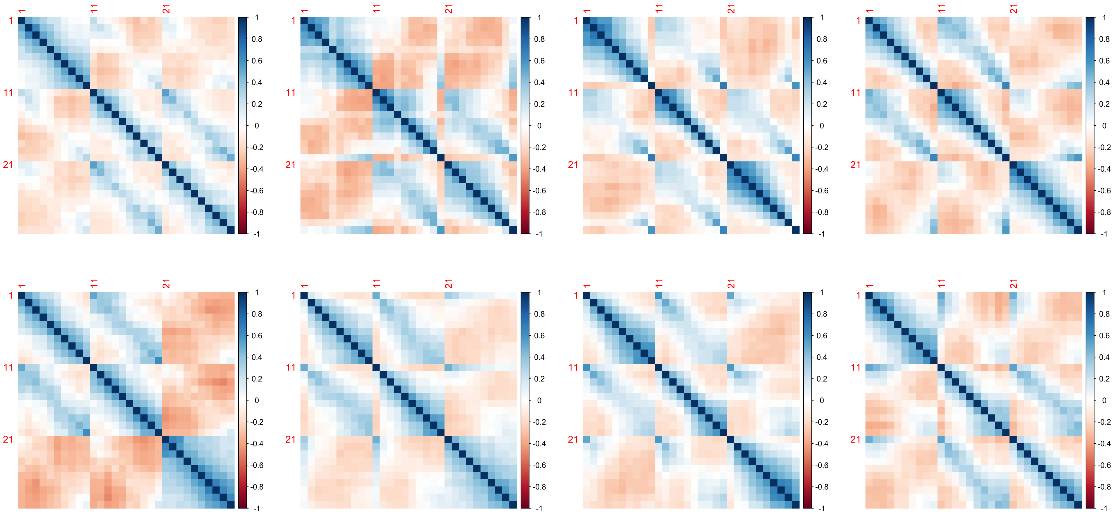


Figure 8.5: Average correlation matrices for the active power without the contribution of the first eigenvalue for the different 45° ranges for the wind farm Riffgat. The wind directions are from left to right N, NE, E, SE and S, SW, W, NW for the two rows.

wind turbines. Thus, information on the spatial structure of the wind farm can be derived from the correlation structure in combination with the wind direction. The observed correlation structures display the total interaction between turbines as an aggregation of complex dynamics such as the wake effect. In all these considerations, it should be noted that the wind direction is still an average over half a day and all wind turbines. The emergence of such a clear pattern despite this fact is noteworthy. It is interesting to investigate to what extent these effects can be found in data with a higher temporal resolution. This will be done in the analysis of the Thanet dataset in the following section.

8.3.2 Thanet

The high temporal resolution of the Thanet dataset makes it possible to study effects on a short time scale. Compared to the Riffgat dataset, the electrical power measurements of the Thanet dataset possess a 60 times higher time resolution. Thus, much smaller observation periods can be chosen with a quantitatively equal data basis. This makes it possible to analyze individual aspects about which no statement could otherwise be made due to a lack of meaningfulness of the data. For the calculation of correlations for the active power, the length of the observation period (the length of the time series of active power) plays a significant role. For short observation periods, the correlation is masked by noise due to individual statistical fluctuations [135, 240].

Therefore, we first discuss the influence of the length of the observation period using different time windows for an exemplary day. In Fig. 8.6 the correlation matrices of the active power over periods of ten minutes, half an hour, one hour, six hours, half a day, and a day are shown. In the correlation matrix over ten minutes, clear correlations and anticorrelations can be seen, forming a recognizable structure. The proportions of correlations and anticorrelations are balanced. The correlation matrix over half an hour holds significantly more correlations than anticorrelations. Interrelated structures of correlations and anticorrelations become larger. They are also more pronounced. The correlation matrix over one hour has essentially the same structure as the matrix over half an hour but weaker anticorrelations. In the correlation matrix over six hours there are only positive correlations. Here, however, large differences for the different observation times emerge. For the example shown all wind turbines are strongly positively correlated with each other. There is only a slight block and stripe structure in the correlation matrix. For another day the correlation is less strong. A stripe structure clearly shows up for it. In addition, there are also some neutral correlations. Yet for another day the correlation matrix consists to the largest part still of neutral correlations (see App. D Fig. D.1 and Fig. D.2). The correlation matrix over half a day does not show much change for the example shown. However, for the other examples mentioned before the matrix is now also dominated by strong positive correlations (see App. D Fig. D.1 and Fig. D.2). Finally, the correlation matrix over a day shows mainly strong positive correlations. For even larger time periods, the correlation increases further and the correlation matrices become more homogeneous. We observe that the correlation matrices all tend towards a uniform structure with increasing length of the observation periods. However, this happens for different observation periods on different time scales. These differences emerge from the range of electrical power production (and wind speed) values within each oberservation period. If the wind speed is similar during the whole observation period, the active power has a narrow range of values and individual fluctuations of the wind turbines dominate. Correlations are therefore much smaller. For large changes in the wind speed during an observation period, the active power has a much broader range of values. The values for all wind turbines also show the same timely variation. Hence, a collective behaviour emerges and dominates the individual fluctuations of the wind turbines. Correlations are therefore much stronger.

We again illustrate the collective behaviour observed for the correlation matrices by an analysis of their eigenvalues. In Fig. 8.7, the eigenvalue spectrum and the corresponding eigenvectors of the correlation matrix over half an hour are shown. The correlation matrix over ten minutes was not chosen at this point because the associated time series include only 60 measurements. This is less than the number of wind turbines considered (100). The rank of the correlation matrix is thus limited by the number of measurements. For this matrix, 41 eigenvalues have a value of zero [43]. For the correlation matrix over half an hour, the associated time series include 180 measurements. Accordingly, the correlation matrix is a full-rank matrix. The eigenvalue spectrum shows many eigenvalues with values smaller than one. Seventeen eigenvalues have values between one and three. However, there are three significantly larger eigenvalues with values of 5.6, 10.1 and 30.5. In the components of the first three corresponding eigenvectors, structures can be recognized that indicate a collective behavior of small groups of wind turbines. The first eigenvector has mostly positive entries of a similar magnitude, but also some neutral and even negative entries. In the second and third eigenvectors, several small coherent groups

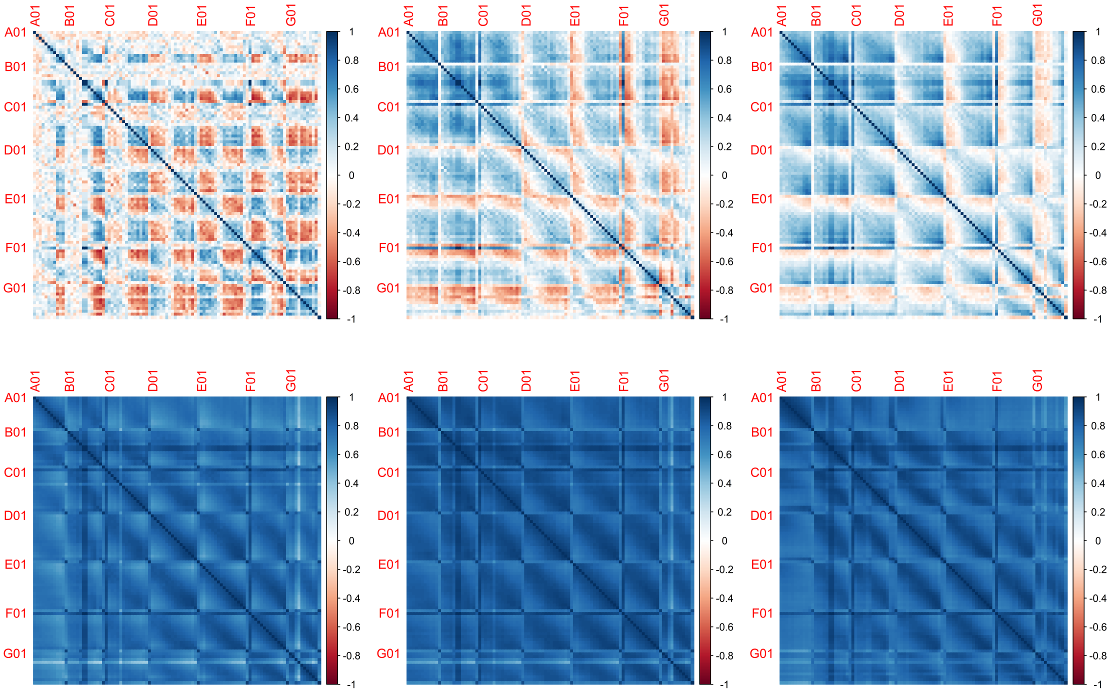


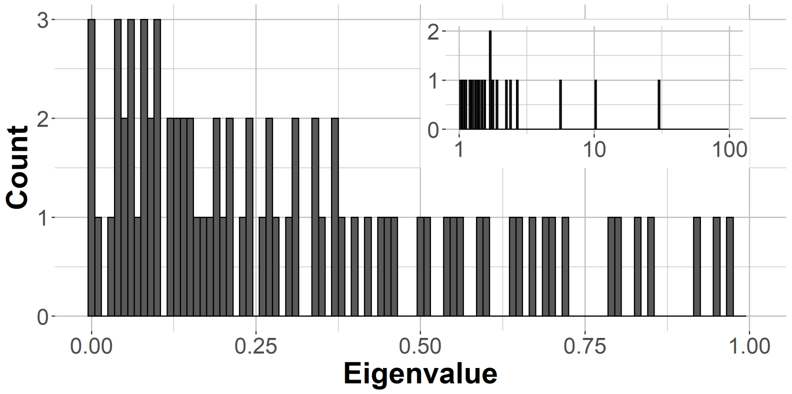
Figure 8.6: Correlation Matrices for the active power over ten minutes (top left), half an hour (top middle), an hour (top right) six hours (bottom left), half a day (bottom middle) and a day (bottom right) for the wind farm Thanet.

of positive and negative entries form. The entries of the remaining eigenvectors do not show a clear structure, they seem to have randomly distributed positive and negative values. The last three eigenvectors each have entries in only four components. However, these are negligible because the associated eigenvalues have a value of zero within machine precision.

Wind speed and wind direction change on short time scales. Long observation periods thus may include phases of different wind speeds and wind directions. This leads on the one hand to stronger correlations between wind turbines and on the other hand to the mixing of possible wind speed and wind direction dependencies in the correlation structure. In order to be able to resolve effects caused by wind speed or direction, the observation period should therefore be chosen as short as possible. The correlation matrices over ten minutes are based on a very small data set and do not have a full rank. For that reason, the correlation matrices over 30 minutes are used for the further analyses.

Despite the short observation period of half an hour, the structures of the correlation matrices are in some cases still strongly characterized by collective effects. To remove the contribution of the collective behavior of all wind turbines, we now consider the time series Ψ of deviations for the individual wind turbines' active power from the mean active power of all wind turbines. A similar approach was used in Ref. [241]. At this point, the active power is not reduced by the contribution of the first singular value/eigenvalue as in Sec. 8.3.1, because the eigenvalue spectra of the correlation matrices show huge differences for the different observation periods. While for some correlation matrices the first eigenvalue is very large and describes the collective dynamics, for other correlation matrices it is smaller and does not describe the collective dynamics. This indicates that sometimes there are influences that are stronger than the collectivity. Taking out the contribution of the first eigenvalue would thus have strongly different effects for the different correlation matrices.

To investigate again a dependence on the wind direction, the correlation matrices are grouped as a function of the mean wind direction of all wind turbines over the respective observation period. For this purpose, the wind direction is divided into eight ranges of 45° each. The ranges are divided so that they are centered around the wind directions that are orthogonal on the wind farm rows. The mean correlation matrices for the 45° ranges of wind direction are shown in Fig. 8.8. There are clear structural differences between the mean correlation matrices of the wind directions. For the structure to be formed in a manner analogous to the observations for the RIFFGAT dataset (see Fig. 8.5), the first wind turbines of each row should be correlated with each other for the north-west direction. This can be seen in the mean correlation matrix. However, the correlations are only very weak. For the south-east direction, the last wind turbines of the rows should be correlated with each other. This can be found in the mean correlation matrix. Finally, for the south-west and north-east directions, the first and last rows, respectively, should be correlated. For the south-west direction, an increased correlation of the first rows can be seen, but for the north-east direction, no increased correlation of the last rows can be found. The stronger expression of the structure for the south-east to south-west direction probably stems from the fact that this is the main wind direction for the observation



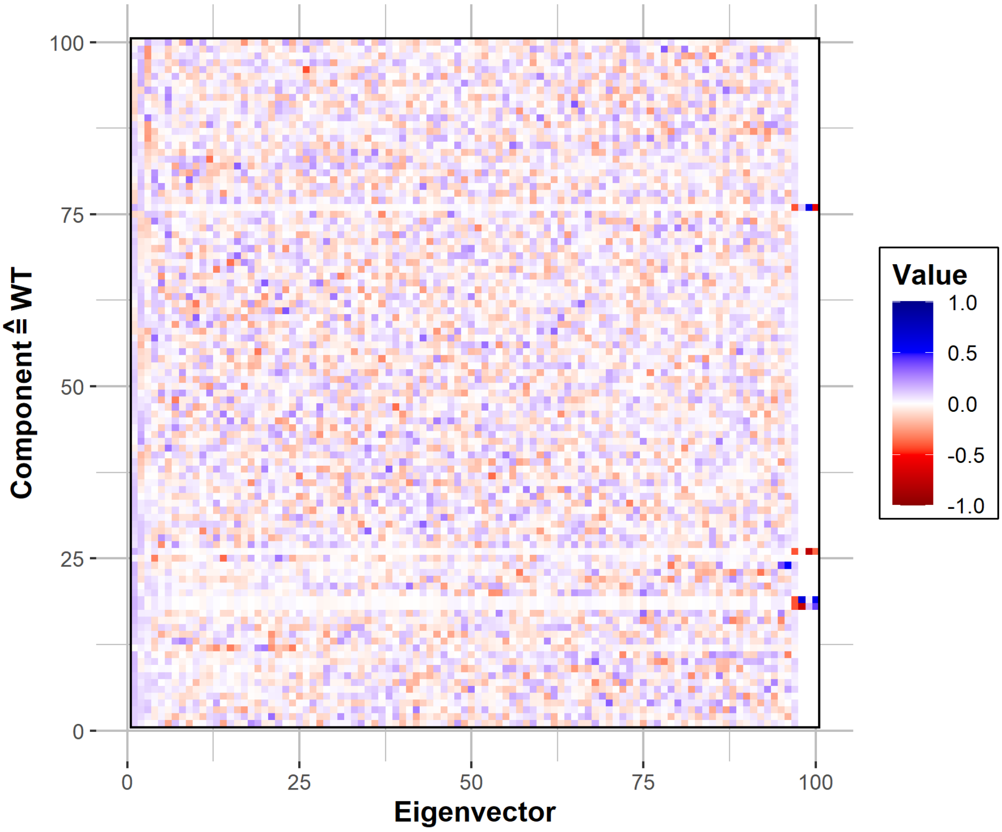


Figure 8.7: Top: Eigenvalue spectrum as histogram counts of the correlation matrix for the active power for half an hour. The inlay shows the large values, while the main plot is zoomed in on small eigenvalues. Bottom: Corresponding eigenvectors of the correlation matrix for the active power. Each column represents an eigenvector. From left to right the corresponding eigenvalue decreases. For numerical values of the entries, see color code.

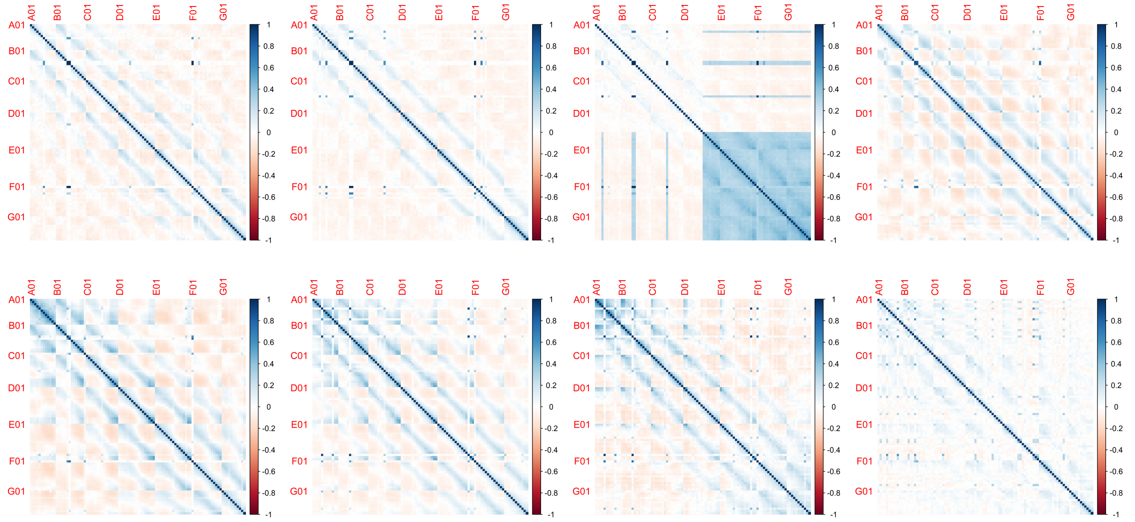


Figure 8.8: Average correlation matrices for the deviations, Ψ, of the individual wind turbines' active power from the mean active power for the different 45° ranges for the wind farm Thanet. The wind directions are from left to right N, NE, E, SE and S, SW, W, NW for the two rows.

period. Both the most and the largest wind speed measurements are available for this direction. A transition between the structures of the south-east and south-west directions can further be seen in the correlation matrix of the south direction. The correlation matrix of the east direction shows strong correlations between the second half of the wind turbines. The respective wind turbines are in a state of failure for a large part of the observation period for that wind direction. The high correlations therefore emerge from our data filling (see Sec. 8.1.2).

The spatial structures are less distinct for the Thanet dataset than for the Riffgat dataset. On the one hand, this may be related to the fact that the distances of the wind turbines for the Thanet wind farm are larger in relation to the rotor diameter of the wind turbines (see Sec. 8.1.1). As a result, the wake effects by the wind turbines are less strong and the wind turbines do not influence each other as much. In particular, the distances between rows are much larger in the wind farm Thanet and are in a range for which the influence of the wind turbines can be considered small [224, 242–244]. On the other hand, the difference between the wind farms may also be due to the different time resolutions of the data and the different lengths of the observation periods. By reducing the temporal resolution of the Thanet data to ten minute averages and calculating the correlation matrices over a time window of half a day we are able to reveal spatial structures more clearly (see App. D Fig. D.3). However, reducing the temporal resolution has again the disadvantage that wind direction and other external circumstances can change significantly within the observation periods, making it impossible for us to draw clear conclusions about their influence.

8.4 Conclusions

We analyzed the operational data of the offshore wind farm Riffgat for the period from 01.03.2014 to 28.02.2015 and the operational data of the offshore wind farm Thanet for February 2017. In both cases, a significant dependence on the spatial structure of the wind farm could be found for the correlation structure of the fluctuations of the active power for the wind turbines. Wind turbines that have spatial proximity to each other show stronger correlations for the electrical power fluctuations with each other. For the Riffgat dataset, a connection between the structures of the correlation matrices and the prevailing wind direction was found. This is surprising since the time periods considered for the correlation matrices were half a day. Over such a long period of time, the wind direction is not constant and can only be approximated to a specific direction. For the high resolution Thanet dataset, this problem was circumvented by considering shorter time periods. It was expected that the structures would be even more pronounced as a function of wind direction for this case. This could not be confirmed in the analysis. Increased correlations due to the spatial proximity of the wind turbines were again found, but a clear correlation with wind direction could only be found for a certain range of wind directions. Using the same time resolution and time periods for the Thanet dataset as previously used for the Riffgat dataset resulted in an enhancement of the structures as a function of wind direction. This is due to the fact that a high temporal resolution leads to stronger noise. Therefore a trade-off between resolution and noise has to be made.

One possible cause for the differences between wind farms is the wind turbine spacing within the wind farms. Especially in relation to the rotor diameters of the wind turbines, the distances between the wind turbines in the wind farm Thanet are significantly larger than in the wind farm Riffgat. For the Riffgat wind farm, the wind turbines within rows have a spacing of 550 m (4.6 rotor diameter). The distance between rows is 600 m (5 rotor diameter). For the Thanet wind farm, the wind turbines within the rows have a spacing of 470 m (5.2 rotor diameter). The spacing between rows is 720 m (8 rotor diameter). Thus, the interactions and wake effects of the wind turbines, especially between rows, are much weaker for the Thanet wind farm than for the Riffgat wind farm. Hence, structures depending on the wind direction are also significantly weaker pronounced. While the influence of wake effects and their complex dynamics are the subject of many studies [24], we have shown that they are intimately connected to the collective behavior. Their influence is measurable without any further input in the cross-correlations of wind farms once the collective behavior is accounted for.

The methods of this chapter are suitable to analyze a large number of existing wind farm layouts with reasonable effort and input-free beyond the measured data. The presented statistical approach complements existing methods to evaluate which layouts minimize turbine interactions. Of course, it cannot replace complex analysis such as fluid dynamics simulations for detailed analysis.

CHAPTER 9

Conclusions

Wind turbines operate in highly variable conditions. Their control system introduces further non-stationarity into the system. We studied this non-stationarity with a focus on mutual correlations between measured observables. Furthermore, we explored some of its consequences for other analysis methods.

We began with the analysis of single wind turbine correlation matrices containing multiple different observables. Applying a matrix distance measure and a bisecting k-means clustering algorithm, previously used in complex systems analysis [42, 43], we successfully distinguished clusters in high frequency wind turbine data using non-overlapping 30 minute correlation matrices. No additional information beyond the data itself was needed for this distinction. Our method confirmed the non-stationary nature of these matrices. We were able to identify the different clusters with operational states induced by the control system of the turbine. While different control states were expected, our analysis provided a tool for their automated detection without prior knowledge. Stability of the states was shown to depend predominantly on wind speed, aligning with expectations of wind turbine behavior. Including the pitch angle as an additional observable into the matrices revealed the standard deviation of wind speed within these intervals as another influential factor. We established the transitional wind speeds separating different states. Thereby, we were able to approximately replicate the cluster distribution using wind speeds averaged over 30 minutes. The approach, initially developed for a single turbine, scaled effectively to multiple turbines without additional effort. Our results underline the significant impact of control dynamics on the correlation patterns in high frequency wind turbine SCADA data. This emphasizes the necessity of considering these variations in operational analyses, such as performance monitoring. The presumption of a static correlation matrix, for instance, when employing principal component analysis for dimensionality reduction on a data set, is not justified by our findings.

The non-stationarity could be taken into account – for example when performing failure analysis on high frequency SCADA data – as a pre-processing procedure. The average wind speed in a time interval could be used to match new or live data to an operational state. Then, with any given method, the comparison to "normal" data could be done within the operational state. This does not require any modifications to established methods, but only their application in subgroups. Given the necessary data, an obvious and interesting extension of the presented analysis would be the transfer onto different turbine types and observables to see how this affects the identified clusters and their dependence on external conditions. Furthermore, the current study is limited in its resolution by the 10-second-resolution of the underlying data. Higher sampling rates would allow the calculation of correlation matrices on shorter epochs. This might help to avoid problems of the current methodology such as the merging of different operational states during one epoch due to changing external conditions. The presented results are also limited to linear correlations, which are shown to be suitable for the analyzed observables, but might not be enough if other observables and turbine types are considered.

Expanding the analysis of operational states, we employed Langevin analysis to study the deterministic and stochastic components of wind turbine control dynamics. We applied the clustering approach to a whole year of wind turbine SCADA data and used the proximity of any given correlation matrix to the observed operational states as a reliable index of the turbine's current control state. Our methodology, inspired by financial market analyses [168, 169], adapted effectively to the wind turbine domain. We performed necessary modifications to accommodate unique phenomena such as control hysteresis, characterized by multiple non-coexisting fixed points. Hereby, we reduced the study of multidimensional system to a one-dimensional time series analysis. The Langevin analysis, including visualizations of drift and potential, provided an intuitive means to comprehend system dynamics. We showed that the complex dynamics of the multidimensional system are captured here in a simplified way. Extensions to larger systems and more observable are possible. For the data at hand, our study provided clarity on the stability of control states and the nature of state transitions. Particularly, it highlighted the complexity of the transition from variable to constant rotation states for the studied turbine. At first, with standard drift estimation, it seemed like the analysis indicated an intermediate state not identified in the clustering. This shows the potential of Langevin analysis to support the determination of the correct cluster number. It may uncover additional states not previously evident in clustering. In our case however, a modified drift estimation method based on peak identification resolved the issue: no intermediate state exists, but two different states might appear for the same wind speed. This is due to control hysteresis. As these states never coexist at the same time, standard drift estimation averages the two fixed points into one seemingly intermediate fixed point. The new estimation method resolves the two different fixed points as a bistable region. This method was specifically tailored to our use case, where an effective trimming of the underlying data revealed the actual fixed points. The second transition from a variable power production to constant power production happens smoothly. From a physics point of view they resemble first and second order phase transitions.

Our results help to understand and evaluate turbine control – especially its dynamics – from measured data, i.e. in real life operation. Furthermore, the extension of static clustering helps to identify wind speed regimes, in which only one cluster is dominant, and (transition) regimes where this might not be the case. This further increases the utility of automated state separation as a pre-processing procedure for any analysis. Above and beyond such applications, there are many interesting prospects to study here. First, shorter epochs for the calculation of the correlation matrix are desirable requiring data of higher resolution. This might help to resolve changing conditions during an epoch. Thereby, the resolution of the Langevin estimation could be increased. Given the required data, transfer to different wind turbines models is possible and might facilitate comparison of systems in real operation. If data is available, it might also be interesting to study the dynamics under different environmental conditions such as onshore and offshore. Lastly, our new peak-based estimation of the drift is a stronger deviation from the theoretical description in Eq. (5.4) than the standard averaging. Further research on its general applicability, maybe of a more theoretical nature, would be interesting.

Building on these findings, we improved the analysis of wind turbine dynamics by integrating operational states into an existing framework: Langevin power curve analysis. Here, the drift and diffusion of a turbine's active power output based on the current wind speed were analyzed. We demonstrated that results are enhanced by considering the identified operational states. This was achieved by employing the suggested pre-processing procedure to segment the data into multiple subgroups, on which our analysis was performed. For the drift we established that the system's dynamics are not uniform but rather distinct across various states. This is corroborated by identifying different stable fixed point per operational state for similar wind speeds, effectively confirming hysteresis in the controller dynamics. More detailed analysis of this effect is desirable for the future. We also saw variations per operational state in the stochastic aspect of the system. These results clearly highlight the necessity of incorporating operational states into the behavioral analysis of wind turbines. While we limited our analysis to one existing analysis technique, a great opportunity for future work is to analyze the effects of non-stationarity on other established methods in the field. Additionally, future investigations could explore the sensitivity of results to different kernel functions and bandwidth when estimating drift and diffusion. Another interesting topic is the dependency of the behavior of other observables in relation to the identified operational states. Such investigations might contribute to a more comprehensive understanding of wind turbine behavior and aid further optimizations.

An important topic in the behavior monitoring field is failure analysis. Here, non-stationarity has already been shown to influence failure analysis for wind turbines [164, 165]. Therefore, it would be desirable to test the benefit of our method for this application in the future. One way to do this, would be the already described pre-processing for an existing failure detection method. Handling of the identified transition regimes and abnormal operating conditions such as curtailment should then be studied as well. Additionally to using the proposed pre-processing for failure analysis methods, we see also a more direct application: Predicting an expected correlation matrix based on a fitted criterion and comparing it to the current one. Subsequently, deviations might point towards operational problems. For this, labeled data with a clear distinction between healthy and faulty periods is indispensable. The currently studied dataset did not contain this information in sufficient quality. Therefore, we have, for now, shown in a theoretical simulation of an arbitrary correlated system that change detection with PCA benefits strongly from accounting for non-stationarity. Our study advanced the understanding of PCA in the context of non-stationary, correlated systems with time series data. We found that incorporating knowledge of a system's potential to occupy multiple "normal" states significantly enhances the sensitivity for change detection. This improvement is not uniformly distributed across all principal components but is most notable in the minor components, which are also the ones most sensitive to changes in the first place. We studied changes to the correlation structure, the mean values and the standard deviations of the time series.

We showed that this technique of failure detection is transferable to real systems with a traffic example. Accounting for the non-stationarity in traffic flow during typical work days increased the ability to detect weekends or holidays with PCA. Our analysis clearly shows that if state-wise non-stationarity is present, it is important to consider it during change detection. The presented analysis is limited to PCA as a change detection method and the simulation results are averaged over many Monte Carlo runs. Therefore, individual systems may exist, where the effect is less strong or stronger. While this is not proven, other change detection methods might also profit from knowledge about non-stationarity. The underlying effect is always the same: It is hard to detect small changes, if the system itself already behaves non-stationary in an unknown way. This underlines the potential of incorporating our results for high frequency wind turbine SCADA data into failure detection systems.

Lastly, we studied the cross-correlations between the active power measurements of multiple turbines. In our analysis of operational data from the RIFFGAT and THANET offshore wind farms, we observed a distinct connection between the correlations of power fluctuations at different turbines and their spatial structure. Proximity within these farms corresponds to stronger power fluctuation correlations. The half-day correlation matrices for RIFFGAT revealed a dependence on prevailing wind directions, a notable finding given the variability of wind. The THANET wind farm's higher resolution data, examined over shorter intervals, only revealed a wind direction dependent correlation within a specific range of wind directions. Examining it on half-day intervals also enhanced the found structures. The deviations between the two wind farms could be attributed to their respective turbine spacing. RIFFGAT'S turbines are closer (in units of the rotor diameter), with inter-row distances allowing for stronger interaction and wake effects, which are less pronounced in the more spaciously arranged THANET wind farm. Our statistical method, while not a substitute for detailed fluid dynamics simulations, offers an efficient approach to evaluate turbine interactions within specific wind farm layouts. It is only capable of capturing the aggregated interactions between turbines, not distinguish individual effects. Further research is necessary into how collective dynamics in wind farms depend on the observation period length and changes of external factors therein. The statistical analysis might then become a complementary tool for optimizing wind farm layouts and positioning.

APPENDIX A

Standard k-means Clustering

The standard k-means algorithm sorts every object $O_i,\ i = 1, \ldots, N$ from a set of N objects into k subsets $\{z_1, \ldots, z_l, \ldots, z_k\}$. Every subset is called a cluster. Generally, the optimal number k has to be determined by a separate method [177, 178].

The input for the algorithm are the objects $O_i,\ i = 1, \ldots, N$ and a distance measure $d(O_i, O_j) \geq 0,\ i, j = 1, \ldots, N$ as well as a method to compute the centroid of any cluster z_l. Note, that the distance measure must also be defined for the centroids. Then the algorithm works as follows:

1. Select k objects as starting cluster centroids.
2. Assign every object to the nearest cluster based on the distance from the object to the cluster centroid.
3. Calculate the new cluster centroids.
4. Repeat steps 2 and 3 until no allocation changes occur.

In Ch. 4 the objects are the correlation matrices of the subset we wish to split into two in the hierarchical approach, therefore $k = 2$ always. The cluster centroids are calculated according to Eq. (4.4) and the distance is calculated by Eq. (4.2).

APPENDIX B

Clustering Results with Spearman's Rank Correlations

The results presented in this study are based on the Pearson correlation coefficient that has been proven useful to establish structural features in complex systems [42, 43, 51, 230]. However, some of the observables used in our analysis, such as wind speed and active power have a well known non-linear dependency. We have therefore tested the robustness of our results, when applying the non-linear measure of Spearman's rank correlation coefficient.

To calculate the time series of Spearman correlation matrices, we rank the individual time series $X_k^{(l)}(t)$, $k = 1, \ldots, K$, $l = 1, \ldots, L$, $t = \tau, \ldots, \tau + T - 1$ for signal k of turbine l for one epoch. Then we proceed by calculating the Pearson correlation matrices for the ranked time series. The following clustering procedure and analysis is carried out in exactly the same way as for the results in Sec. 4.3. We present here results for the case of five observables, which can be directly compared to the results with Pearson correlation coefficients in Sec. 4.3.1.

The silhouette coefficients shown in Fig. B.1 and Tab. B.1 indicate good grouping. They show on average marginally larger values than for the Pearson correlation. Comparing the resulting cluster centers for Spearman correlations in Fig. B.2 with those for simple Pearson correlations in Fig. 4.2, it is obvious that the differences for the structural features revealed in this analysis are minimal. The number of elements changes slightly for some clusters, but the overall result and interpretation are the same for both correlation measures.

As expected from the similarity of the results, also the plots of the cluster allocation over time in Fig. B.3 and over wind speed in Fig. B.4 are very similar to their counterparts in Sec. 4.3.1.

We conclude that for the analysis carried out in this study, the simple Pearson correlation measure is sufficient. The structural differences in the correlation matrices and thereby also the structural differences in the eigenvectors, i.e. principal components, are well captured in the linear correlations.

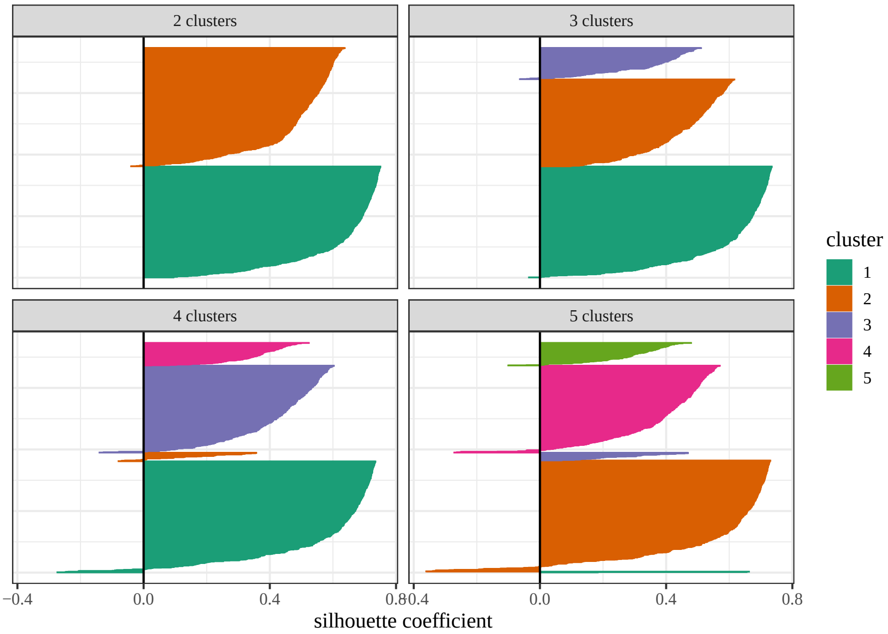


Figure B.1: Silhouette plots for clustering solutions with 2-5 clusters for Spearman's rank correlation matrices. Each clustered element (matrix) is represented by a horizontal line the length of which is the silhouette coefficient for that element. Different clusters are color coded.

Table B.1: Minimum, first quartile, median, mean, third quartile and maximum of silhouette coefficients for the clustering solutions with 2-5 clusters for Spearman ranked correlation matrices of WT1.

clusters	min	1st Qu.	median	mean	3rd Qu.	max
2	-0.040	0.479	0.579	0.550	0.677	0.751
3	-0.064	0.406	0.533	0.507	0.656	0.735
4	-0.274	0.365	0.510	0.476	0.649	0.735
5	-0.361	0.325	0.483	0.450	0.641	0.730

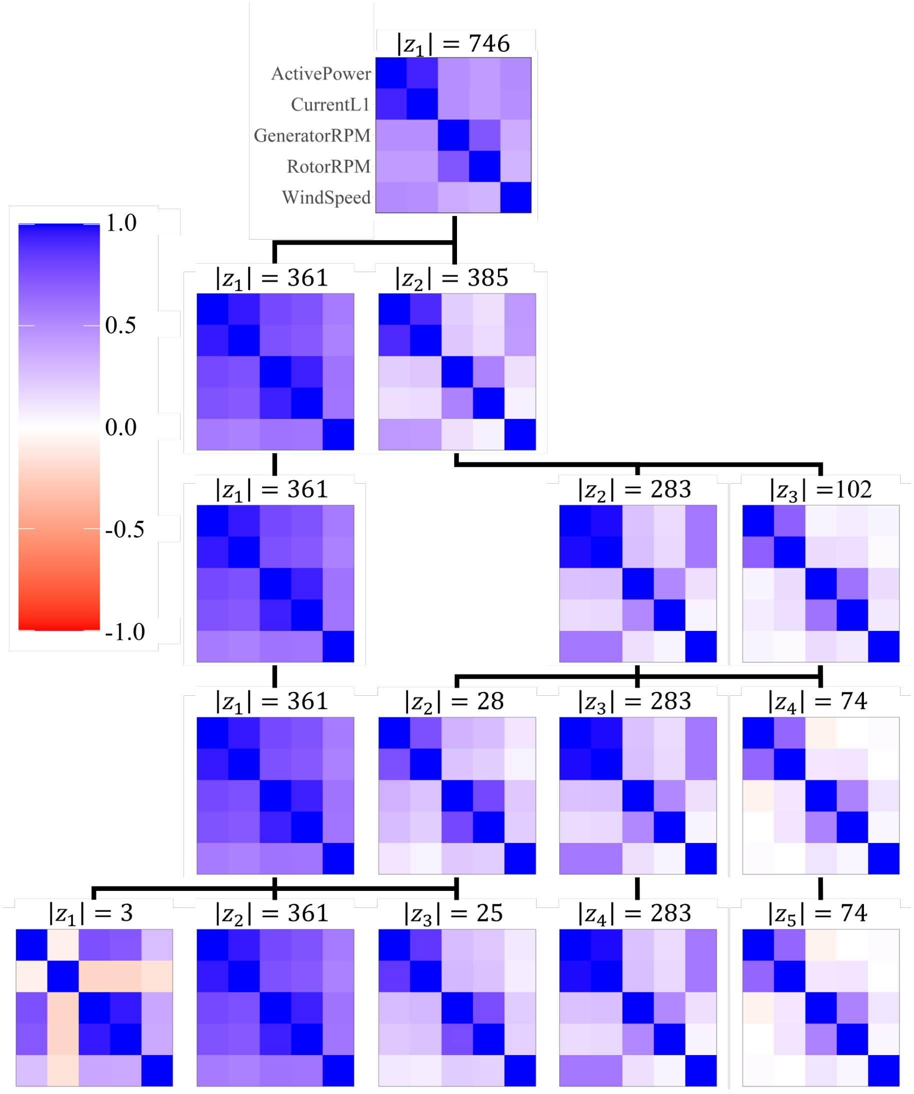


Figure B.2: Spearman's rank correlation matrix cluster centroids as calculated in Eq. (4.4) for WT1 for different numbers of clusters. The color indicates the value of the correlation coefficient. Black lines connect child and parent clusters of the hierarchical algorithm and the number of cluster elements is given as $|z_s|$. Each cluster solution is ordered from low wind speeds (left) to high wind speeds (right) according to the average wind speed in a cluster.

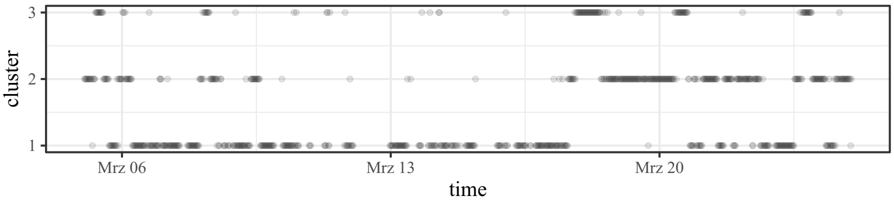


Figure B.3: Cluster identifier s over time for WT1 and Spearman's rank correlations. Each dot represents a 30 minute epoch.

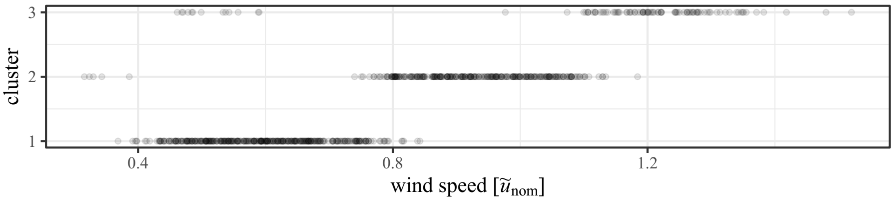


Figure B.4: Cluster identifier s over wind speed for WT1 and Spearman's rank correlations. Each dot represents a 30 minute epoch.

APPENDIX C

Contour Lines for Operational States

To represent the distribution of operational states visually, we utilize kernel density estimation given by

$$\rho(u,P,S)=\sum_{t=1}^{T_{\text{end}}}\kappa_{\text{G,G,B}}\left(\frac{P(t)-P}{h_P},\frac{u(t)-u}{u_h},S(t)-S\right) \tag{C.1}$$

with Gaussian and Boolean kernels $\kappa_{\text{G}}(y)$ and $\kappa_{\text{b}}(y)$ given by

$$\kappa_{\text{G}}(y)=e^{-\frac{1}{2}y^2} \tag{C.2}$$

and

$$\kappa_{\text{B}}=\begin{cases}1 & ,\ y=0\\ 0 & ,\ \text{else}\end{cases} \tag{C.3}$$

Contour lines are then generated using the formula:

$$f(u,P,S)=\begin{cases}1 & ,\ \rho(u,P,S)\geq\rho_0\\ 0 & ,\ \text{else}\end{cases} \tag{C.4}$$

Here, ρ_0 is a predefined threshold set to an arbitrary value of 20.

APPENDIX D

Further Wind Farm Correlation Matrices

D.1 Further Correlation Matrices Thanet

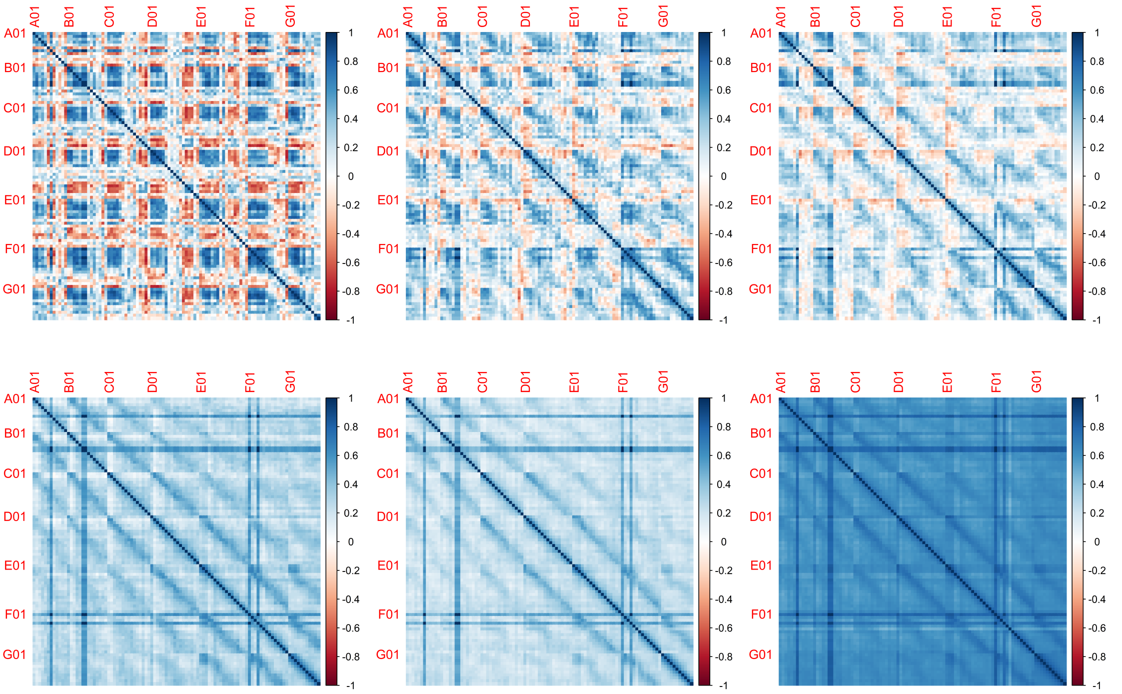


Figure D.1: Correlation Matrices for the active power over ten minutes (top left), half an hour (top middle), an hour (top right) six hours (bottom left), half a day (bottom middle) and a day (bottom right) for the wind farm THANET, February 8.

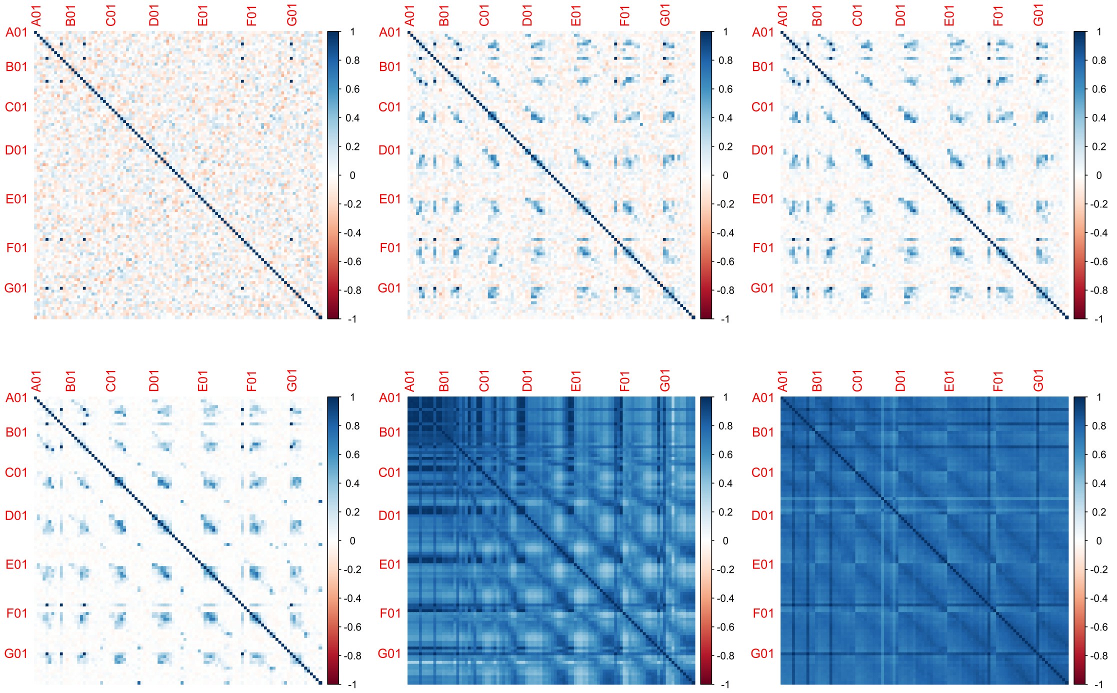


Figure D.2: Correlation Matrices for the active power over ten minutes (top left), half an hour (top middle), an hour (top right) six hours (bottom left), half a day (bottom middle) and a day (bottom right) for the wind farm THANET, February 27.

D.2 Correlation matrices Thanet for Half a Day

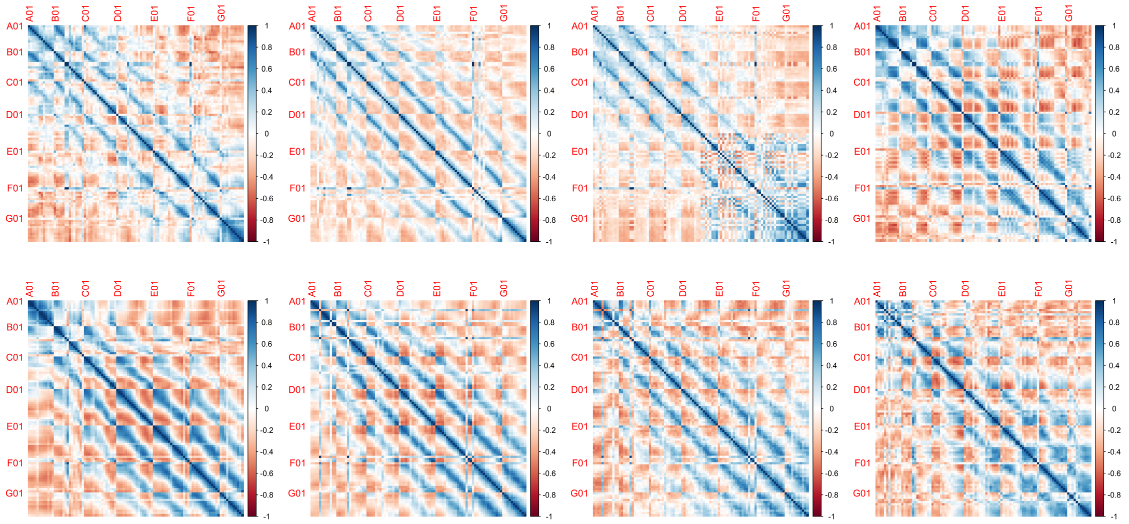


Figure D.3: Average correlation matrices for the active power without the contribution of the first eigenvalue for the different 45° ranges for the wind farm THANET over half a day. The wind directions are from left to right N, NE, E, SE and S, SW, W, NW for the two rows.

Acronyms

AC	Alternating Current.
CFD	Computational Fluid Dynamics.
CMS	Condition Monitoring System.
DC	Direct Current.
DFIG	Doubly-Fed Induction Generator.
FRP	Fiber-Reinforced Plastic.
HAWT	Horizontal Axis Wind Turbine.
LCOE	Levelized Cost Of Energy.
O&M	Operation and Maintenance.
PCA	Principal Component Analysis.
RDS-PP	Reference Designation System for Power Plants.
RMT	Random Matrix Theory.
RWS	Rescaled wind speed.
SCADA	Supervisory Control and Data Acquisition.
VAWT	Vertical Axis Wind Turbine.
WiSAbigdata	Wind farm virtual Site Assistant for O&M decision support — advanced methods for big data analysis.
ZEUS	Zustands-Ereignis-Ursachen-Schlüssel – Condition-Event-Cause-Key.

List of Figures

List of Tables

Bibliography

[1] H. M. Bette, E. Jungblut, and T. Guhr. *Nonstationarity in correlation matrices for wind turbine SCADA-data.* Wind Energy 26.8 (2023), 826–849. DOI: `10.1002/we.2843`.

[2] H. M. Bette et al. *Dynamics of wind turbine operational states* (2023). arXiv: `2310.06098 [physics.flu-dyn]`.

[3] C. Philipp et al. *Extension of the Langevin power curve analysis by separation per operational state* (2023). arXiv: `2305.15512 [physics.data-an]`.

[4] H. M. Bette, M. Schreckenberg, and T. Guhr. *Sensitivity of principal components to system changes in the presence of non-stationarity.* Journal of Statistical Mechanics: Theory and Experiment 2023.10 (2023), 103402. DOI: `10.1088/1742-5468/ad0033`.

[5] E. Jungblut, H. M. Bette, and T. Guhr. *Spatial Structures of Wind Farms: Correlation Analysis of the Generated Electrical Power* (2022). arXiv: `2210.12969 [stat.AP]`.

[6] R. Knecktys et al. *Risk Theory and Pricing of "Pay-for-Performance" Business Models* (2022). arXiv: `2212.09585 [q-fin.MF]`.

[7] R. Gasch and J. Twele, eds. *Windkraftanlagen: Grundlagen, Entwurf, Planung und Betrieb.* Vieweg: Springer, 2016.

[8] E. Hau. *Windkraftanlagen: Grundlagen, Technik, Einsatz, Wirtschaftlichkeit.* Berlin, Heidelberg: Springer, 2016.

[9] *Climate Change 2023: Synthesis Report.* Tech. rep. Intergovernmental Panel on Climate Change, 2023.

[10] *Statistical Review Of World Energy.* Tech. rep. Energy Institute, 2023.

[11] *World Energy Outlook 2022.* Tech. rep. International Energy Agency, 2022.

[12] *Renewables 2023 – Global Status Report.* Tech. rep. Renewable Energy Policy Network for the 21st Century, 2023.

[13] G. A. M. van Kuik et al. *Long-term research challenges in wind energy – a research agenda by the European Academy of Wind Energy.* Wind Energy Science 1.1 (2016), 1–39. DOI: `10.5194/wes-1-1-2016`.

[14] P. Veers et al. *Grand challenges in the science of wind energy.* Science 366.6464 (2019), eaau2027. DOI: `10.1126/science.aau2027`.

[15] A. Bianchini et al. *Current status and grand challenges for small wind turbine technology.* Wind Energy Science 7.5 (2022), 2003–2037. DOI: `10.5194/wes-7-2003-2022`.

[16] T. Burton et al. *Wind Energy Handbook.* Chichester: John Wiley & Sons, Ltd, 2021.

[17] W. Musial et al. *Offshore Wind Market Report: 2023 Edition.* Tech. rep. U.S. Department of Energy, 2023.

[18] J. Maldonado-Correa et al. *Using SCADA data for wind turbine condition monitoring: A systematic literature review.* Energies 13.12 (2020), 3132. DOI: `10.3390/en13123132`.

[19] L. Thomas and M. Ramachandra. *Advanced materials for wind turbine blade - A Review.* Materials Today: Proceedings 5.1, Part 3 (2018), 2635–2640. DOI: `doi.org/10.1016/j.matpr.2018.01.043`.

[20] V. Igwemezie, A. Mehmanparast, and A. Kolios. *Current trend in offshore wind energy sector and material requirements for fatigue resistance improvement in large wind turbine support structures – A review.* Renewable and Sustainable Energy Reviews 101 (2019), 181–196. DOI: `10.1016/j.rser.2018.11.002`.

[21] Y. Zhao et al. *Fault prediction and diagnosis of wind turbine generators using SCADA data.* Energies 10.8 (2017), 1210. DOI: `10.3390/en10081210`.

[22] J. Meyers et al. *Wind farm flow control: prospects and challenges.* Wind Energy Science 7.6 (2022), 2271–2306. DOI: `10.5194/wes-7-2271-2022`.

[23] U. Ritschel and M. Beyer. “Modern Wind Turbines”. *Designing Wind Turbines: Engineering and Manufacturing Process in the Industrial Context.* Cham: Springer, 2022, 13–41.

[24] F. Porté-Agel, M. Bastankhah, and S. Shamsoddin. *Wind-Turbine and Wind-Farm Flows: A Review.* Boundary-Layer Meteorology 174 (2020), 1–59. DOI: `10.1007/s10546-019-00473-0`.

[25] D. L. Stein and C. M. Newman. *Spin Glasses and Complexity.* Princeton: Princeton University Press, 2013.

[26] H. E. Stanley. *Phase transitions and critical phenomena.* Oxford: Clarendon Press, 1971.

[27] G. Parisi. *Order Parameter for Spin-Glasses.* Physical Review Letters 50.24 (1983), 1946–1948. DOI: `10.1103/PhysRevLett.50.1946`.

[28] U. Schollwöck et al. *Quantum Magnetism.* Berlin, Heidelberg: Springer, 2004.

[29] J. Kwapień and S. Drożdż. *Physical approach to complex systems.* Physics Reports 515.3 (2012), 115–226. DOI: `10.1016/j.physrep.2012.01.007`.

[30] A. Cavagna, I. Giardina, and T. S. Grigera. *The physics of flocking: Correlation as a compass from experiments to theory.* Physics Reports 728 (2018), 1–62. DOI: `10.1016/j.physrep.2017.11.003`.

[31] K. Kaneko. *Life: An Introduction to Complex Systems Biology.* Berlin, Heidelberg: Springer, 2006.

[32] D. S. Fisher. *Course 11 Evolutionary dynamics.* Complex Systems. Ed. by J.-P. Bouchaud, M. Mézard, and J. Dalibard. Vol. 85. Les Houches. Elsevier, 2007, 395–446. DOI: `10.1016/S0924-8099(07)80018-7`.

[33] D. L. Turcotte and B. D. Malamud. *Earthquakes as a Complex System.* International Handbook of Earthquake and Engineering Seismology, Part A. Ed. by W. H. Lee et al. Vol. 81. International Geophysics. Academic Press, 2002, 209–227. DOI: `10.1016/S0074-6142(02)80217-0`.

[34] A. Chattopadhyay, P. Hassanzadeh, and S. Pasha. *Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data.* Scientific Reports 10.1 (2020), 1317. DOI: `10.1038/s41598-020-57897-9`.

[35] M. Fathi et al. *Big Data Analytics in Weather Forecasting: A Systematic Review.* Archives of Computational Methods in Engineering 29.2 (2022), 1247–1275. DOI: `10.1007/s11831-021-09616-4`.

[36] K. Kashinath et al. *Physics-informed machine learning: case studies for weather and climate modelling.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 379.2194 (2021), 20200093. DOI: `10.1098/rsta.2020.0093`.

[37] K. Hasselmann. *Stochastic climate models Part I. Theory.* Tellus 28.6 (1976), 473–485. DOI: `10.1111/j.2153-3490.1976.tb00696.x`.

[38] K. Hasselmann. *Multi-pattern fingerprint method for detection and attribution of climate change.* Climate Dynamics 13.9 (1997), 601–611. DOI: `10.1007/s003820050185`.

[39] N. Boers et al. *Complex networks reveal global pattern of extreme-rainfall teleconnections.* Nature 566.7744 (2019), 373–377. DOI: `10.1038/s41586-018-0872-x`.

[40] J. Ludescher et al. *Network-based forecasting of climate phenomena.* Proceedings of the National Academy of Sciences 118.47 (2021), e1922872118. DOI: `10.1073/pnas.1922872118`.

[41] 2021.

[42] M. C. Münnix et al. *Identifying states of a financial market.* Scientific Reports 2 (2012), 644. DOI: `10.1038/srep00644`.

[43] A. J. Heckens, S. M. Krause, and T. Guhr. *Uncovering the dynamics of correlation structures relative to the collective market motion.* Journal of Statistical Mechanics: Theory and Experiment 2020.10 (2020), 103402. DOI: `10.1088/1742-5468/abb6e2`.

[44] J.-P. Bouchaud and M. Potters. "Financial applications of random matrix theory: a short review". *The Oxford Handbook of Random Matrix Theory.* Oxford: Oxford University Press, 2015.

[45] L. Zhao et al. *Stock market as temporal network.* Physica A: Statistical Mechanics and its Applications 506 (2018), 1104–1112. DOI: `10.1016/j.physa.2018.05.039`.

[46] T. A. Schmitt et al. *Credit risk and the instability of the financial system: An ensemble approach.* Europhysics Letters 105.3 (2014), 38004. DOI: `10.1209/0295-5075/105/38004`.

[47] M. C. Münnix, R. Schäfer, and T. Guhr. *A Random Matrix Approach to Credit Risk.* PLOS ONE 9.5 (2014), 1–9. DOI: `10.1371/journal.pone.0098030`.

[48] R. K. Sawyer. *Social emergence: Societies as complex systems.* Cambridge: Cambridge University Press, 2005.

[49] R. Condorelli. *Complex Systems Theory: Some Considerations for Sociology.* Open Journal of Applied Sciences 6.7 (2016), 422–448. DOI: `10.4236/ojapps.2016.67044`.

[50] N. Crossley. *Small-World Networks, Complex Systems and Sociology.* Sociology 42.2 (2008), 261–277. DOI: `10.1177/0038038507087353`.

[51] S. Wang et al. *Quasi-stationary states in temporal correlations for traffic systems: Cologne orbital motorway as an example.* Journal of Statistical Mechanics: Theory and Experiment 2020.10 (2020), 103404. DOI: `10.1088/1742-5468/ABBCD3`.

[52] A. Bellouquid, E. de Angelis, and L. Fermo. *Toward the Modeling of Vehicular Traffic as a Complex System: A Kinetic Theory Approach.* Mathematical Models and Methods in Applied Sciences 22.supp01 (2012), 1140003. DOI: `10.1142/S0218202511400033`.

[53] S. Albeverio et al., eds. *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach.* Heidelberg: Physica, 2007.

[54] T. McPhearson et al. *Advancing understanding of the complex nature of urban systems.* Ecological Indicators 70 (2016), 566–573. DOI: `10.1016/j.ecolind.2016.03.054`.

[55] J. T. Matamalas, S. Gómez, and A. Arenas. *Abrupt phase transition of epidemic spreading in simplicial complexes.* Physical Review Research 2.1 (2020), 012049. DOI: `10.1103/PhysRevResearch.2.012049`.

[56] H. N. Kouser, R. Barnard-Mayers, and E. Murray. *Complex systems models for causal inference in social epidemiology.* Journal of Epidemiology & Community Health 75.7 (2021), 702–708. DOI: `10.1136/jech-2019-213052`.

[57] L. Cameron and D. Larsen-Freeman. *Complex Systems and Applied Linguistics.* International Journal of Applied Linguistics 17.2 (2007), 226–240. DOI: `10.1111/j.1473-4192.2007.00148.x`.

[58] J. S. Kelso. *Dynamic patterns: The self-organization of brain and behavior.* Cambridge: The MIT Press, 1997.

[59] J. Janarek et al. *Investigating structural and functional aspects of the brain's criticality in stroke.* Scientific Reports 13.1 (2023), 12341. DOI: `10.1038/s41598-023-39467-x`.

[60] T. Padmanabhan and V. Padmanabhan. "The Thirst for Power". *The Dawn of Science: Glimpses from History for the Curious Mind.* Cham: Springer, 2019, 255–263.

[61] M. Sathyajith. *Wind Energy: Fundamentals, Resource Analysis and Economics.* Berlin, Heidelberg: Springer, 2006.

[62] E. W. Golding. *The Generation Of Electricity By Wind Power.* New York: E. u. F. N. Spon LTD, 1977.

[63] A. G. Drachmann. "Heron's Windmill". *Renewable Energy: Four Volume Set.* Ed. by B. Sorensen. London: Routledge, 64–69.

[64] A. Y. al-Hassan and D. R. Hill. *Islamic Technology: An Illustrated History.* Cambridge: Cambridge University Press, 1986, 52–54.

[65] A. Lucas. *Wind, Water, Work: Ancient And Medieval Milling Technology.* Technology and change in history, v. 8. Leiden, Boston: Brill, 2006, 65.

[66] L. W. Jr. *Medieval Technology and Social Change.* Oxford: Oxford University Press, 1968, 87.

[67] F.-W. Wellmer and J. Gottschalk. *Leibniz' Scheitern im Oberharzer Silberbergbau – neu betrachtet, insbesondere unter klimatischen Gesichtspunkten.* Studia Leibnitiana 42.2 (2010), 186–207.

[68] H. P. Münzenmayer. *Leibniz' Inventum Memorabile Die Konzeption einer Drehzahlregelung vom März 1686.* Studia Leibnitiana 8.1 (1976), 113–119.

[69] L. Euler. *Recherches plus exactes sur l'effet des moulins à vent.* Mémoires de l'académie des sciences de Berlin 12 (1758), 165–234.

[70] C. S. Gillmor. "Torsion Studies". *Coulomb and the Evolution of Physics and Engineering in Eighteenth-Century France.* Princeton: Princeton University Press, 1971, 139–174.

[71] E. Fröde and W. Fröde. *Windmühlen.* Köln: Du Mont Verlag, 1981.

[72] G. A. van Kuik. *The Lanchester–Betz–Joukowsky limit.* Wind Energy 10.3 (2007), 289–291. DOI: `10.1002/we.218`.

[73] T. Baker. *A Field Guide to American Windmills.* Oklahoma: University of Oklahoma Press, 1985.

[74] E. R. Leonhardt. *Die Internationale Elektrische Ausstellung Wien 1883.* Freiberg: Craz und Gerlach, 1884.

[75] T. J. Price. *James Blyth — Britain's First Modern Wind Power Pioneer.* Wind Engineering 29.3 (2005), 191–200. DOI: `10.1260/030952405774354921`.

[76] *Mr. Brush's Windmill Dynamo.* Scientific American 63.25 (1890), 389–389.

[77] P. L. Cour. *Forsøgsmøllen.* Kopenhagen: Det Nordiskke Forlag, 1900.

[78] P. C. Putnam. *Power from the Wind.* New York: Van Nostrand Rheinhold, 1948.

[79] S. Windkraft eV. *Betriebserfahrungen mit einer Windkraftanlage von 100 kW.* DK 621.311.4 (1963).

[80] E. Hau, J. Langenbrinck, and W. Palz. *WEGA Large Wind Turbines.* Berlin, Heidelberg: Springer, 1993.

[81] E. Commission. *WEGA II Large Wind Turbine Scientific Evaluation Project.* 2000.

[82] M. Hutchinson and F. Zhao. *Global Wind Report 2023.* Tech. rep. Global Wind Energy Council, 2023.

[83] A. Betz. *Einführung in die Theorie der Strömungsmaschinen.* Karlsruhe: Braun, 1959.

[84] E. Hau. *Wind Turbines: Fundamentals, Technologies, Application, Econcomics.* Berlin, Heidelberg: Springer, 2006.

[85] G. Giebel and C. B. Hasager. "An Overview of Offshore Wind Farm Design". *MARE-WINT: New Materials and Reliability in Offshore Wind Turbine Technology.* Ed. by W. Ostachowicz et al. Cham: Springer, 2016, 337–346.

[86] L. Parada et al. *Wind farm layout optimization using a Gaussian-based wake model.* Renewable Energy 107 (2017), 531–541. DOI: `https://doi.org/10.1016/j.renene.2017.02.017`.

[87] L. E. Andersson et al. *Wind farm control - Part I: A review on control system concepts and structures.* IET Renewable Power Generation 15.10 (2021), 2085–2108. DOI: `https://doi.org/10.1049/rpg2.12160`.

[88] S. Watson et al. *Future emerging technologies in the wind power sector: A European perspective.* Renewable and Sustainable Energy Reviews 113 (2019), 109270. DOI: `10.1016/j.rser.2019.109270`.

[89] *Renewables 2022 – Global Status Report.* Tech. rep. Renewable Energy Policy Network for the 21st Century, 2022.

[90] D. Franke et al. *BGR Energiestude: Daten und Entwicklungen der deutschen und globalen Energieversorgung.* Tech. rep. Bundesanstalt für Geowissenschaften und Rohstoffe, 2022.

[91] *Net Zero by 2050.* Tech. rep. International Energy Agency, 2021.

[92] M. Hannan et al. *Wind Energy Conversions, Controls, and Applications: A Review for Sustainable Technologies and Directions.* Sustainability 15.5 (2023), 3986.

[93] C. Kost et al. *Levelized Cost Of Electricity: Renewable Energy Technologies.* Tech. rep. Fraunhofer-Institut für Solare Energiesysteme ISE, 2021.

[94] P. Tavner. *Offshore wind turbines: reliability, availability and maintenance.* London: Institution of Engineering and Technology, 2012. DOI: `10.1049/PBRN013E`.

[95] J. Feuchtwang and D. Infield. *Offshore wind turbine maintenance access: a closed-form probabilistic method for calculating delays caused by sea-state.* Wind Energy 16.7 (2013), 1049–1066. DOI: `https://doi.org/10.1002/we.1539`.

[96] E. J. N. Menezes, A. M. Araújo, and N. S. B. da Silva. *A review on wind turbine control and its associated methods.* Journal of Cleaner Production 174 (2018), 945–953. DOI: `10.1016/J.JCLEPRO.2017.10.297`.

[97] C. D. Dao et al. *Integrated condition-based maintenance modelling and optimisation for offshore wind turbines.* Wind Energy January (2021), 1–19. DOI: `10.1002/we.2625`.

[98] M. Ulmer et al. *Early Fault Detection Based on Wind Turbine SCADA Data Using Convolutional Neural Networks. PHM Society European Conference.* Vol. 5. 1. 2020, 9. DOI: `10.36001/phme.2020.v5i1.1217`.

[99] H. Yan et al. *Fault Diagnosis of Wind Turbine Based on PCA and GSA-SVM.* 2019 Prognostics and System Health Management Conference (PHM-Paris) (2019), 13–17. DOI: `10.1109/PHM-Paris.2019.00010`.

[100] S. Martin-del-Campo and K. Al-Kahwati. *Unsupervised Ranking of Outliers in Wind Turbines via Isolation Forest with Dictionary Learning. PHM Society European Conference.* Vol. 5(1). 2020, 9. DOI: `10.36001/phme.2020.v5i1.1164`.

[101] W. Chen et al. *Diagnosis of wind turbine faults with transfer learning algorithms.* Renewable Energy 163 (2021), 2053–2067. DOI: `10.1016/j.renene.2020.10.121`.

[102] R. K. Pandit and D. Infield. *SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes.* IET Renewable Power Generation 12.11 (2018), 1249–1255. DOI: `10.1049/iet-rpg.2018.0156`.

[103] A. Meyer and B. Brodbeck. *Data-driven Performance Fault Detection in Commercial Wind Turbines.* PHM Society European Conference 5.1 (2020), 7. DOI: `10.36001/phme.2020.v5i1.1276`.

[104] C. Dao, B. Kazemtabrizi, and C. Crabtree. *Wind turbine reliability data review and impacts on levelised cost of energy.* Wind Energy 22.12 (2019), 1848–1871. DOI: `10.1002/we.2404`.

[105] L. Colone et al. *Assessing the Utility of Early Warning Systems for Detecting Failures in Major Wind Turbine Components.* Journal of Physics: Conference Series 1037.3 (2018). DOI: `10.1088/1742-6596/1037/3/032005`.

[106] Y. Li et al. *A review of condition-based maintenance: Its prognostic and operational aspects.* Frontiers of Engineering Management 7.3 (2020), 323–334. DOI: `10.1007/s42524-020-0121-5`.

[107] L. Trampe and S. Opitz. *Vom Labor in die Anwendung - Strukturüberwachung für die Windenergie.* Tech. rep. Deutsches Zentrum für Luft- und Raumfahrt e. V., 2022.

[108] A. Stetco et al. *Machine learning methods for wind turbine condition monitoring: A review.* Renewable Energy 133 (2019), 620–635. DOI: `10.1016/j.renene.2018.10.047`.

[109] K. Sayed and H. Gabbar. *SCADA and smart energy grid control automation.* Smart Energy Grid Engineering. Ed. by H. A. Gabbar. Academic Press, 2017, 481–514. DOI: `10.1016/B978-0-12-805343-0.00018-8`.

[110] J. Tautz-Weinert and S. J. Watson. *Using SCADA data for wind turbine condition monitoring - A review.* IET Renewable Power Generation 11.4 (2017), 382–394. DOI: `10.1049/iet-rpg.2016.0248`.

[111] X. Jin, Z. Xu, and W. Qiao. *Condition monitoring of wind turbine generators using SCADA data analysis.* IEEE Transactions on Sustainable Energy 12.1 (2021), 202–210. DOI: `10.1109/TSTE.2020.2989220`.

[112] S. Kolumbán, S. Kapodistria, and N. Nooraee. “Short and long-term wind turbine power output prediction”. 2017. arXiv: `1707.06497v1`.

[113] J. Gottschall and J. Peinke. *How to improve the estimation of power curves for wind turbines.* Environmental Research Letters 3.1 (2008), 015005. DOI: `10.1088/1748-9326/3/1/015005`.

[114] U. Aziz et al. *Critical comparison of power-based wind turbine fault-detection methods using a realistic framework for SCADA data simulation.* Renewable and Sustainable Energy Reviews 144 (2021), 110961. DOI: `10.1016/J.RSER.2021.110961`.

[115] G. Helbing and M. Ritter. *Deep Learning for fault detection in wind turbines.* Renewable and Sustainable Energy Reviews 98 (2018), 189–198. DOI: `10.1016/j.rser.2018.09.012`.

[116] F. Pozo, Y. Vidal, and Ó. Salgado. *Wind turbine condition monitoring strategy through multiway PCA and multivariate inference.* Energies 11.4 (2018), 749. DOI: `10.3390/en11040749`.

[117] Y. Vidal, F. Pozo, and C. Tutivén. *Wind turbine multi-fault detection and classification based on SCADA data.* Energies 11.11 (2018), 3018. DOI: `10.3390/en11113018`.

[118] A. Pliego Marugán and F. P. García Márquez. *Advanced analytics for detection and diagnosis of false alarms and faults: A real case study.* Wind Energy 22.11 (2019), 1622–1635. DOI: `https://doi.org/10.1002/we.2393`.

[119] M. Beretta et al. *Quantification of the Information Loss Resulting from Temporal Aggregation of Wind Turbine Operating Data.* Applied Sciences 11.17 (2021). DOI: `10.3390/app11178065`.

[120] P. F. Odgaard and K. E. Johnson. *Wind turbine fault detection and fault tolerant control - An enhanced benchmark challenge. 2013 American Control Conference*. IEEE, 2013, 4447–4452. DOI: `10.1109/acc.2013.6580525`.

[121] *Technische Richtlinie für Erzeugungseinheiten: Teil 7: Instandhaltung von Kraftwerken für Erneuerbare Energien: Rubrik D2: Zustands-Ereignis-Ursachen-Schlüssel für Erzeugungseinheiten (ZEUS)*. Tech. rep. Fördergesellschaft Windenergie und andere Erneuerbare Energien e. V., 2013.

[122] *VGB-Standard RDS-PP: Application Guideline Part 32: Wind Power Plants*. Tech. rep. VGB Powertech e. V., 2014.

[123] J. D. Hamilton. *Time Series Analysis*. Princeton: Princeton University Press, 1994. DOI: `10.1515/9780691218632`.

[124] G. Marti et al. *A review of two decades of correlations, hierarchies, networks and clustering in financial markets*. Progress in Information Geometry. Signals and Communication Technology. Ed. by F. Nielsen. Cham: Springer, 2021, 245–274.

[125] J. Wishart. *The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population*. Biometrika 20A.1/2 (1928), 32–52. DOI: `10.2307/2331939`.

[126] E. P. Wigner. *Characteristic Vectors of Bordered Matrices With Infinite Dimensions*. Annals of Mathematics 62.3 (1955), 548–564. DOI: `10.2307/1970079`.

[127] T. Guhr, A. Müller–Groeling, and H. A. Weidenmüller. *Random-matrix theories in quantum physics: common concepts*. Physics Reports 299.4 (1998), 189–425. DOI: `10.1016/S0370-1573(97)00088-4`.

[128] B. Collins and I. Nechita. *Random matrix techniques in quantum information theory*. Journal of Mathematical Physics 57.1 (2015), 015215. DOI: `10.1063/1.4936880`.

[129] D. Paul and A. Aue. *Random matrix theory in statistics: A review*. Journal of Statistical Planning and Inference 150 (2014), 1–29. DOI: `10.1016/j.jspi.2013.09.005`.

[130] L. Laloux et al. *Random Matrix Theory and Financial Correlations*. International Journal of Theoretical and Applied Finance 03.03 (2000), 391–397. DOI: `10.1142/S0219024900000255`.

[131] V. Plerou et al. *A random matrix theory approach to financial cross-correlations*. Physica A: Statistical Mechanics and its Applications 287.3 (2000), 374–382. DOI: `10.1016/S0378-4371(00)00376-9`.

[132] Vinayak and T. H. Seligman. *Time series, correlation matrices and random matrix models*. AIP Conference Proceedings 1575.1 (2014), 196–217. DOI: `10.1063/1.4861704`.

[133] P. Gopikrishnan et al. *Quantifying and interpreting collective behavior in financial markets*. Physical Review E 64.3 (2001), 035106. DOI: `10.1103/PhysRevE.64.035106`.

[134] D. Chetalova, R. Schäfer, and T. Guhr. *Zooming into market states.* Journal of Statistical Mechanics: Theory and Experiment 2015.1 (2015), P01029. DOI: `10.1088/1742-5468/2015/01/P01029`.

[135] T. Guhr and B. Kälber. *A new method to estimate the noise in financial correlation matrices.* Journal of Physics A: Mathematical and General 36.12 SPEC. ISS. (2003), 3009–3032. DOI: `10.1088/0305-4470/36/12/310`. arXiv: `0206577 [cond-mat]`.

[136] K. Pearson. *LIII. On lines and planes of closest fit to systems of points in space.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2.11 (1901), 559–572. DOI: `10.1080/14786440109462720`.

[137] H. Hotelling. *Analysis of a complex of statistical variables into principal components.* Journal of Educational Psychology 24.6 (), 417–441. DOI: `10.1037/h0071325`.

[138] H. Hotelling. *Relations Between Two Sets of Variates.* Biometrika 28.3/4 (1936), 321–377.

[139] M. Greenacre et al. *Principal component analysis.* Nature Reviews Methods Primers 2.1 (2022), 100. DOI: `10.1038/s43586-022-00184-w`.

[140] D. J. Denis. "Principal components analysis". *Applied Univariate, Bivariate, and Multivariate Statistics.* Hoboken: John Wiley & Sons, Ltd., 2021, 423–448.

[141] N. Vaswani and P. Narayanamurthy. *Static and Dynamic Robust PCA and Matrix Completion: A Review.* Proceedings of the IEEE 106.8 (2018), 1359–1379. DOI: `10.1109/JPROC.2018.2844126`.

[142] M. Ahmadi et al. *Detection of brain lesion location in MRI images using convolutional neural network and robust PCA.* International Journal of Neuroscience 0.0 (2021), 1–12. DOI: `10.1080/00207454.2021.1883602`.

[143] S. A. Rahim and G. Manson. *Kernel Principal Component Analysis for Structural Health Monitoring and Damage Detection of an Engineering Structure Under Operational Loading Variations.* Journal of Failure Analysis and Prevention 21.6 (2021), 1981–1990. DOI: `10.1007/S11668-021-01260-1/FIGURES/8`.

[144] W. Bounoua and A. Bakdi. *Fault detection and diagnosis of nonlinear dynamical processes through correlation dimension and fractal analysis based dynamic kernel PCA.* Chemical Engineering Science 229 (2021), 116099. DOI: `10.1016/J.CES.2020.116099`.

[145] H. Haimi et al. *Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant.* Engineering Applications of Artificial Intelligence 52 (2016), 65–80. DOI: `10.1016/J.ENGAPPAI.2016.02.003`.

[146] L. M. Elshenawy and T. A. Mahmoud. *Fault diagnosis of time-varying processes using modified reconstruction-based contributions.* Journal of Process Control 70 (2018), 12–23. DOI: `10.1016/J.JPROCONT.2018.07.017`.

[147] P. Kazemi et al. *Fault detection and diagnosis in water resource recovery facilities using incremental PCA*. Water Science and Technology 82.12 (2020), 2711–2724. DOI: `10.2166/WST.2020.368`.

[148] B. D. Ketelaere, M. Hubert, and E. Schmitt. *Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent, High-Dimensional Data*. Journal of Quality Technology 47.4 (2015), 318–335. DOI: `10.1080/00224065.2015.11918137`.

[149] T. Rato et al. *A systematic comparison of PCA-based Statistical Process Monitoring methods for high-dimensional, time-dependent Processes*. AIChE Journal 62.5 (2016), 1478–1493. DOI: `10.1002/AIC.15062`.

[150] A. Bakdi and A. Kouadri. *A new adaptive PCA based thresholding scheme for fault detection in complex systems*. Chemometrics and Intelligent Laboratory Systems 162 (2017), 83–93. DOI: `10.1016/j.chemolab.2017.01.013`.

[151] A. Sánchez-Fernández et al. *Fault detection based on time series modeling and multivariate statistical process control*. Chemometrics and Intelligent Laboratory Systems 182 (2018), 57–69. DOI: `10.1016/j.chemolab.2018.08.003`.

[152] F. Harrou et al. *Improved principal component analysis for anomaly detection: Application to an emergency department*. Computers & Industrial Engineering 88 (2015), 63–77. DOI: `10.1016/J.CIE.2015.06.020`.

[153] X. Li et al. *Fault Data Detection of Traffic Detector Based on Wavelet Packet in the Residual Subspace Associated with PCA*. Applied Sciences 9 (2019), 3491. DOI: `10.3390/APP9173491`.

[154] M. A. F. Pimentel et al. *A review of novelty detection*. Signal Processing 99 (2014), 215–249. DOI: `10.1016/j.sigpro.2013.12.026`.

[155] H. Dutta et al. *Distributed top-K outlier detection from astronomy catalogs using the DEMAC system*. Society for Industrial and Applied Mathematics Publications, 2007, 473–478. DOI: `10.1137/1.9781611972771.47`.

[156] L. I. Kuncheva and W. J. Faithfull. *PCA feature extraction for change detection in multidimensional unlabeled data*. IEEE Transactions on Neural Networks and Learning Systems 25.1 (2014), 69–80. DOI: `10.1109/TNNLS.2013.2248094`.

[157] M.-L. Shyu et al. *A Novel Anomaly Detection Scheme Based on Principal Component Classifier*. University of Miami, 2003.

[158] M. Tveten. *Which principal components are most sensitive in the change detection problem?* Stat 8.1 (2019), e252. DOI: `10.1002/sta4.252`.

[159] L. Shang et al. *Stable recursive canonical variate state space modeling for time-varying processes*. Control Engineering Practice 36 (2015), 113–119. DOI: `10.1016/J.CONENGPRAC.2014.12.006`.

[160] G. Canonaco, M. Restelli, and M. Roveri. *Model-Free Non-Stationarity Detection and Adaptation in Reinforcement Learning*. Froniers in Artificial Intelligence and Applications 325 (2020), 1047–1054. DOI: `10.3233/FAIA200200`.

[161] C. P. B. de Mesquita et al. *Taking climate change into account: Non-stationarity in climate drivers of ecological response.* Journal of Ecology 109.3 (2021), 1491–1500. DOI: `10.1111/1365-2745.13572`.

[162] G. Chen et al. *An integrated framework for statistical change detection in running status of industrial machinery under transient conditions.* ISA Transactions 94 (2019), 294–306. DOI: `10.1016/j.isatra.2019.03.026`.

[163] J. Huang et al. *Fault detection for chemical processes based on non-stationarity sensitive cointegration analysis.* ISA Transactions 129 (2022), 321–333. DOI: `10.1016/J.ISATRA.2022.02.010`.

[164] R. Zimroz et al. *Diagnostics of bearings in presence of strong operating conditions non-stationarity - A procedure of load-dependent features processing with application to wind turbine bearings.* Mechanical Systems and Signal Processing 46.1 (2014), 16–27. DOI: `10.1016/j.ymssp.2013.09.010`.

[165] T. Bull, M. D. Ulriksen, and D. Tcherniak. *The effect of environmental and operational variabilities on damage detection in wind turbine blades.* 2018.

[166] L. D. Avendano-Valencia et al. *Operational regime clustering for the construction of PCE-based surrogates of operational wind turbines.* 2017, 961–964.

[167] G. Buccheri, S. Marmi, and R. N. Mantegna. *Evolution of correlation structure of industrial indices of US equity markets.* Phys. Rev. E 88.1 (2013), 012806. DOI: `10.1103/PhysRevE.88.012806`.

[168] P. Rinn et al. *Dynamics of quasi-stationary systems: Finance as an example.* Europhysics Letters 110.6 (2015), 68003. DOI: `10.1209/0295-5075/110/68003`. arXiv: `1502.07522`.

[169] Y. Stepanov et al. *Stability and hierarchy of quasi-stationary states: Financial markets as an example.* Journal of Statistical Mechanics: Theory and Experiment 2015.8 (2015), P08011. DOI: `10.1088/1742-5468/2015/08/P08011`. arXiv: `1503.00556`.

[170] T. Braun et al. *Correlated power time series of individual wind turbines: A data driven model approach.* Journal of Renewable and Sustainable Energy 12.2 (2020), 023301. DOI: `10.1063/1.5139039`.

[171] M. E. J. Newman. *Communities, modules and large-scale structure in networks.* Nature Phys 8 (2012), 25–31. DOI: `10.1038/NPHYS2162`.

[172] S. P. Patankar et al. *Path-dependent connectivity, not modularity, consistently predicts controllability of structural brain networks.* Network Neuroscience 4.4 (2020), 1091–1121. DOI: `10.1162/netn\_a\_00157`.

[173] H. Toju et al. *Network modules and hubs in plant-root fungal biomes.* Journal of The Royal Society Interface 13.116 (2016), 20151097. DOI: `10.1098/rsif.2015.1097`.

[174] M. Steinbach, G. Karypis, and V. Kumar. *A Comparison of Document Clustering Techniques. KDD Workshop on Text Mining.* 2000.

[175] P.-N. Tan et al. *Introduction to Data Mining (2nd Edition)*. New York: Pearson Education, 2019.

[176] S. P. Lloyd. *Least Squares Quantization in PCM*. IEEE Transactions on Information Theory 28.2 (1982), 129–137. DOI: `10.1109/TIT.1982.1056489`.

[177] A. K. Jain. *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters 31.8 (2010), 651–666. DOI: `10.1016/j.patrec.2009.09.011`.

[178] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data*. Hoboken: John Wiley & Sons, Inc., 1990.

[179] P. J. Rousseeuw. *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics 20.C (1987), 53–65. DOI: `10.1016/0377-0427(87)90125-7`.

[180] R. J. Schütt. "Control of Wind Energy Systems". *Understanding Wind Power Technology: Theory, Deployment and Optimisation*. Ed. by A. SChaffarczyk. Chichester: Wiley, 2014, 340–368.

[181] C. Behnken, M. Wächter, and J. Peinke. *Multipoint reconstruction of wind speeds*. Wind Energy Science 5.3 (2020), 1211–1223. DOI: `10.5194/WES-5-1211-2020`.

[182] J. Weber et al. *Wind Power Persistence Characterized by Superstatistics*. Scientific Reports 9.1 (2019), 19971. DOI: `10.1038/s41598-019-56286-1`. arXiv: `1810.06391`.

[183] R. D. Maesschalck, D. Jouan-Rimbaud, and D. L. Massart. *The Mahalanobis distance*. Chemometrics and Intelligent Laboratory Systems 50.1 (2000), 1–18. DOI: `10.1016/s0169-7439(99)00047-7`.

[184] R. Ruiz de la Hermosa González-Carrato. *Wind farm monitoring using Mahalanobis distance and fuzzy clustering*. Renewable Energy 123 (2018), 526–540. DOI: `10.1016/j.renene.2018.02.097`.

[185] K. Sayed, A. G. Abo-Khalil, and A. M. Eltamaly. *Wind Power Plants Control Systems Based on SCADA System*. Green Energy and Technology (2021), 109–151. DOI: `10.1007/978-3-030-64336-2_6/FIGURES/21`.

[186] E. J. N. Menezes, A. M. Araújo, and N. S. B. da Silva. *A review on wind turbine control and its associated methods*. Journal of Cleaner Production 174 (2018), 945–953. DOI: `10.1016/J.JCLEPRO.2017.10.297`.

[187] L. Pustina, F. Biral, and J. Serafini. *A novel Economic Nonlinear Model Predictive Controller for power maximisation on wind turbines*. Renewable and Sustainable Energy Reviews 170 (2022), 112964. DOI: `10.1016/J.RSER.2022.112964`.

[188] B. Fernandez-Gauna et al. *Actor-critic continuous state reinforcement learning for wind-turbine control robust optimization*. Information Sciences 591 (2022), 365–380. DOI: `10.1016/J.INS.2022.01.047`.

[189] F. Pöschke et al. *Model-based wind turbine control design with power tracking capability: A wind-tunnel validation*. Control Engineering Practice 120 (2022), 105014. DOI: `10.1016/J.CONENGPRAC.2021.105014`.

[190] J. López-Queija et al. *Review of control technologies for floating offshore wind turbines.* Renewable and Sustainable Energy Reviews 167 (2022), 112787. DOI: `10.1016/J.RSER.2022.112787`.

[191] M. Wächter et al. *Power performance of wind energy converters characterized as stochastic process: applications of the Langevin power curve.* Wind Energy 14.6 (2011), 711–717. DOI: `10.1002/we.453`.

[192] H. Risken. *The Fokker-Planck equation.* Berlin, Heidelberg: Springer, 1996.

[193] M. R. R. Tabar. *Analysis and data-based reconstruction of complex nonlinear dynamical systems.* Cham: Springer, 2019.

[194] E. A. Nadaraya. *On Estimating Regression.* Theory of Probability & Its Applications 9.1 (1964), 141–142. DOI: `10.1137/1109020`.

[195] G. S. Watson. *Smooth Regression Analysis.* Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 26.4 (1964), 359–372.

[196] V. Epanechnikov. *Nonparametric estimation of a multidimensional probability density.* Teoriya veroyatnostei i ee primeneniya 14.1 (1969), 156–161.

[197] D. Wied and R. Weißbach. *Consistency of the kernel density estimator: a survey.* Statistical Papers 53 (2012), 1–21.

[198] B. W. Silverman. *Density estimation for statistics and data analysis.* London, New York: Chapman & Hall, 1986.

[199] M. Siefert et al. *On a quantitative method to analyze dynamical and measurement noise.* Europhysics Letters 61.4 (2003), 466. DOI: `10.1209/epl/i2003-00152-9`.

[200] R. Tabar. *Analysis and data-based reconstruction of complex nonlinear dynamical systems.* Vol. 730. Cham: Springer, 2019.

[201] T. A. Mücke et al. *Langevin power curve analysis for numerical wind energy converter models with new insights on high frequency power performance.* Wind Energy 18.11 (2015), 1953–1971.

[202] P. Milan et al. *Applications of the Langevin power curve. Proceedings of the EWEC.* 2010.

[203] P. P. Lin et al. *Discontinuous Jump Behavior of the Energy Conversion in Wind Energy Systems.* PRX Energy 2.3 (2023), 033009. DOI: `10.1103/PRXEnergy.2.033009`.

[204] E. A. Nadaraya. *On estimating regression.* Theory of Probability & Its Applications 9.1 (1964), 141–142.

[205] *61400-12-1 Wind turbines–Part 12-1: Power performance measurements of electricity producing wind turbines.* Tech. rep. IEC, 2005.

[206] H. R. Kerner et al. *Comparison of novelty detection methods for multispectral images in rover-based planetary exploration missions.* Data Mining and Knowledge Discovery 34.6 (2020), 1642–1675. DOI: `10.1007/s10618-020-00697-6`.

[207] K. B. Newhart et al. *Data-driven performance analyses of wastewater treatment plants: A review.* Water Research 157 (2019), 498–513. DOI: `10.1016/J.WATRES.2019.03.030`.

[208] A. Maleki, I. Sadeghkhani, and B. Fani. *Statistical sensorless short-circuit fault detection algorithm for photovoltaic arrays.* Journal of Renewable and Sustainable Energy 11.5 (2019), 053501. DOI: `10.1063/1.5119055`.

[209] S. Wang et al. *Collective behavior in the North Rhine-Westphalia motorway network.* Journal of Statistical Mechanics: Theory and Experiment 2021.12 (2021), 123401. DOI: `10.1088/1742-5468/ac3662`.

[210] E. Hellinger. *Neue Begründung der Theorie quadratischer Formen und unendlichvielen Veränderlichen.* Journal für die reine und angewandte Mathematik 136 (1909), 210–271. DOI: `10.1515%2Fcrll.1909.136.210`.

[211] H. Joe. *Generating random correlation matrices based on partial correlations.* Journal of Multivariate Analysis 97.10 (2006), 2177–2189. DOI: `10.1016/J.JMVA.2005.05.010`.

[212] N. J. Higham. *Computing the nearest correlation matrix—a problem from finance.* IMA Journal of Numerical Analysis 22.3 (2002), 329–343. DOI: `10.1093/IMANUM/22.3.329`.

[213] J. H. J. Ward. *Hierarchical Grouping to Optimize an Objective Function.* Journal of the American Statistical Association 58.301 (1963), 236–244. DOI: `10.1080/01621459.1963.10500845`.

[214] F. Murtagh and P. Legendre. *Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?* Journal of Classification 31 (2014), 274–295. DOI: `10.1007/s00357-014-9161-z`.

[215] P. Legendre and L. Legendre. *Numerical Ecology.* Amsterdam: Elsevier, 2012.

[216] M. Nawaz et al. *Improved process monitoring using the CUSUM and EWMA-based multiscale PCA fault detection framework.* Chinese Journal of Chemical Engineering 29 (2021), 253–265. DOI: `https://doi.org/10.1016/j.cjche.2020.08.035`.

[217] M. A. de Carvalho Michalski and G. F. Martha de Souza. *Comparing PCA-based fault detection methods for dynamic processes with correlated and Non-Gaussian variables.* Expert Systems with Applications 207 (2022), 117989. DOI: `https://doi.org/10.1016/j.eswa.2022.117989`.

[218] H. Zhang et al. *Sensor fault detection and diagnosis for a water source heat pump air-conditioning system based on PCA and preprocessed by combined clustering.* Applied Thermal Engineering 160 (2019), 114098. DOI: `https://doi.org/10.1016/j.applthermaleng.2019.114098`.

[219] N. O. Jensen. *A note on wind generator interaction.* Risø National Laboratory 2411 (1983).

[220] S. Frandsen et al. *Analytical modelling of wind speed deficit in large offshore wind farms.* Wind Energy 9.1-2 (2006), 39–53. DOI: `10.1002/we.189`.

[221] R. J. Barthelmie et al. *Modelling and measurements of wakes in large wind farms.* Journal of Physics: Conference Series 75 (2007), 012049. DOI: `10.1088/1742-6596/75/1/012049`.

[222] X. Gao et al. *Investigation of wind turbine performance coupling wake and topography effects based on LiDAR measurements and SCADA data.* Applied Energy 255 (2019), 113816. DOI: `10.1016/j.apenergy.2019.113816`.

[223] F. Castellani et al. *Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment.* Applied Energy 185 (2017), 1076–1086. DOI: `10.1016/j.apenergy.2015.12.049`.

[224] P. McKay, R. Carriveau, and D. S.-K. Ting. *Wake impacts on downstream wind turbine performance and yaw alignment.* Wind Energy 16.2 (2013), 221–234. DOI: `10.1002/we.544`.

[225] R. J. Barthelmie et al. *Modelling and measurements of power losses and turbulence intensity in wind turbine wakes at Middelgrunden offshore wind farm.* Wind Energy 10.6 (2007), 517–528. DOI: `10.1002/we.238`.

[226] M. Méchali et al. *Wake effects at Horns Rev and their influence on energy production. European wind energy conference and exhibition.* Vol. 1. 2006, 10–20.

[227] M. Ali et al. *Wind Farm Model Aggregation Using Probabilistic Clustering.* IEEE Transactions on Power Systems 28.1 (2013), 309–316. DOI: `10.1109/TPWRS.2012.2204282`.

[228] R. M. Castro and J. F. De Jesus. *An aggregated wind park model. 13th Power System Computation Conference-PSCC.* Vol. 99. 1999.

[229] A. Marinopoulos et al. *Investigating the impact of wake effect on wind farm aggregation. 2011 IEEE Trondheim PowerTech.* 2011, 1–5. DOI: `10.1109/PTC.2011.6019323`.

[230] S. Gartzke et al. *Spatial correlation analysis of traffic flow on parallel motorways in Germany.* Physica A: Statistical Mechanics and its Applications 599 (2022), 127367. DOI: `https://doi.org/10.1016/j.physa.2022.127367`.

[231] EWE AG. *Offshore-Windpark RIFFGAT GmbH & Co. KG.* Accessed: March 02, 2021.

[232] TenneT TSO GmbH. *Riffgat Drehstrom-Netzanbindung vor der Küste Borkums.* Accessed: March 02, 2021.

[233] BBC. *Largest offshore wind farm opens off Thanet in Kent.* Accessed: March 02, 2021.

[234] Vattenfall AB. *Thanet Offshore Wind Farm.* Accessed: January 27, 2021.

[235] M. Schlechtingen and I. Ferreira Santos. *Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection.* Mechanical Systems and Signal Processing 25.5 (2011), 1849–1875. DOI: `10.1016/j.ymssp.2010.12.007`.

[236] M. Schlechtingen, I. F. Santos, and S. Achiche. *Using Data-Mining Approaches for Wind Turbine Power Curve Monitoring: A Comparative Study.* IEEE Transactions on Sustainable Energy 4.3 (2013), 671–679. DOI: `10.1109/TSTE.2013.2241797`.

[237] J. Godwin, P. Matthews, and C. Watson. *Classification and detection of electrical control system faults through SCADA data analysis.* Chemical engineering transactions 33 (2013), 985–990.

[238] L. Zheng, W. Hu, and Y. Min. *Raw Wind Data Preprocessing: A Data-Mining Approach.* IEEE Transactions on Sustainable Energy 6.1 (2015), 11–19. DOI: `10.1109/TSTE.2014.2355837`.

[239] J. Liesen and V. Mehrmann. *Linear Algebra.* Cham: Springer, 2015.

[240] L. Giada and M. Marsili. *Data clustering and noise undressing of correlation matrices.* Phys. Rev. E 63.6 (2001), 061101. DOI: `10.1103/PhysRevE.63.061101`.

[241] C. Borghesi, M. Marsili, and S. Micciché. *Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode.* Phys. Rev. E 76.2 (2007), 026104. DOI: `10.1103/PhysRevE.76.026104`.

[242] J. Annoni et al. *Evaluating wake models for wind farm control. 2014 American Control Conference.* 2014, 2517–2523. DOI: `10.1109/ACC.2014.6858970`.

[243] A. Crespo, J. Hernández, and S. Frandsen. *Survey of modelling methods for wind turbine wakes and wind farms.* Wind Energy 2.1 (1999), 1–24. DOI: `10.1002/(SICI)1099-1824(199901/03)2:1<1::AID-WE16>3.0.CO;2-7`.

[244] P. Builtjes and J. Smit. *Calculation of Wake Effects in Wind Turbine Parks.* Wind Engineering 2.3 (1978), 135–145.

DOI: 10.17185/duepublico/82061
URN: urn:nbn:de:hbz:465-20240718-073840-1