Delft University of Technology
Faculty of EEMCS

Fault Detection for Wind Energy using Multi-Scale Statistics

BY
F. M. Jacobs

to obtain the degree of Master of Science in Applied Mathematics - Stochastics
at the Delft University of Technology,
to be defended publicly on 24 October

Student number:	4879694		
Project duration:	March 1, 2024 – October 24, 2024		
Thesis committee:	Dr. F. Mies,	TU Delft,	daily supervisor
	Prof. Dr. Ir. G. Jongbloed,	TU Delft,	responsible professor
	Dr. Ir. G.F. Nane	TU Delft	associate professor
	Ir. N. Van den Bos,	Vandebron,	external supervisor

TU Delft

Executive Summary

Wind energy has become an increasingly important part of the global energy transition, with wind turbines playing a key role in renewable energy production. However, the variability of wind conditions and the technical complexities of wind turbines present significant challenges for reliable energy output and efficient operations. One of the main tasks in wind energy data analysis is detecting change points, moments in time when the behavior of a system, such as a turbine's power output, changes. These points often signify important events like faults, system upgrades, environmental impacts (such as icing or grid curtailment), or maintenance activities.

This thesis is motivated by the need to improve the detection of such change points in wind energy data, particularly for Vandebron, a Dutch company focused on renewable energy. Vandebron seeks to improve the performance of its machine learning models for wind power prediction by filtering out anomalous data, caused by events such as unplanned maintenance or environmental factors like bird strikes or icing. Current methods for anomaly and change point detection in wind energy are either too simplistic to handle the complex data or too rigid in dealing with non-linear relationships that are inherent in wind energy systems.

To address this, this thesis proposes the use of Narrowest Significance Pursuit (NSP), a method developed by [Fryzlewicz, 2023], originally designed to detect change points in linear models. The NSP method identifies intervals in a time series that are likely to contain change points by controlling the Type I error while maintaining statistical rigor. It does so by calculating a test statistic for intervals in the data and returning those that meet a specified significance threshold. However, wind energy data often exhibits non-linear patterns. For instance, the relationship between wind speed and power output is typically non-linear. To handle such relationships, this thesis extends the NSP method to detect change points in non-linear time series, adapting it to better model the dynamics of wind turbines.

Several extensions of NSP are proposed to accommodate the non-linear nature of wind energy data. First, parametric methods are proposed either with input or model error. These methods require a definition for the dependence between the wind speed and power output, called the power curve, of which two were used in the application. The selection of the function describing the power curve brings the problem of model misspecification. Using a less accurate function immediately causes for a specific model to perform worse. In addition to the parametric methods, two non-parametric methods are proposed as well. The first makes use of the monotone nature of the power curve with isotonic regression. The second models the curve with an S-shaped regression, created by a convex and concave part.

Simulations were used to assess the performance of these NSP extensions, with manually inserted change points representing various anomalous periods of different lengths and magnitudes. The methods tested included logistic and cubic models with both output and input error, as well as isotonic regression. Overall, the isotonic model outperformed the others, achieving 100% coverage of the change points while detecting the largest number of changes and returning the narrowest intervals. The logistic and cubic models with output error also performed well, with the logistic model slightly outperforming the cubic one, though this might be influenced by the fact that the simulations themselves were based on a logistic function. Of all the methods the ones with input error scored the lowest, the cubic even with a coverage of 93.99% which is

below the significance of 95%. This is partly influenced by the construction of the simulations and by the implementations of these functions.

The models were also validated on a real-world dataset provided by Vandebron, which included actual change points, wind speed, and power output. All models were tested on the set as well, after which the isotonic method consistently exhibited the best results. In Figure 1 a segment of the validation set is visible. It displays the significant interval detected by the isotonic model, using the wind speed and power, and the anomalous period, called *Down.Curt.*.

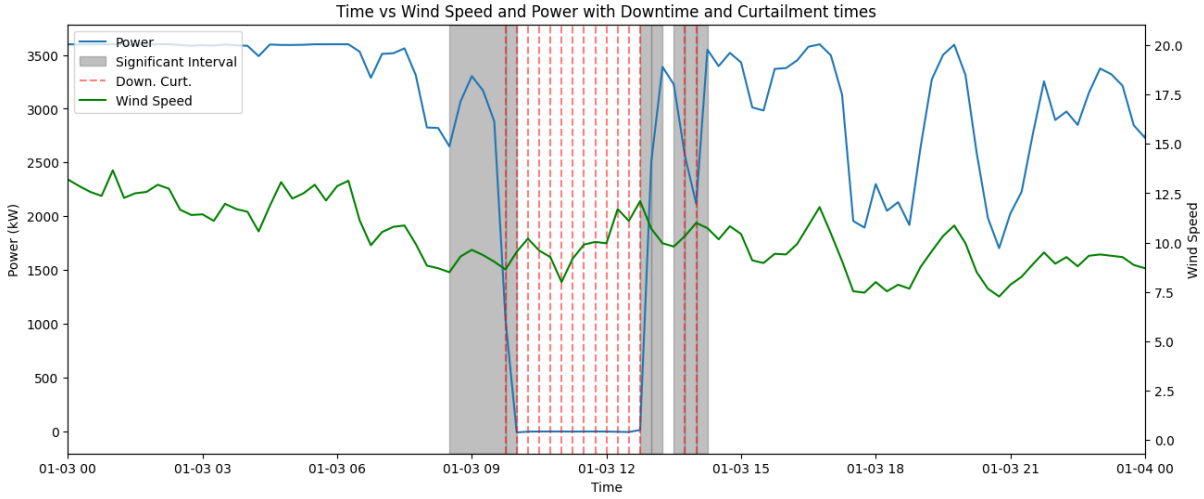


Figure 1: Segment of detected changes in the validation set by the isotonic model.

Limitations of this research lie in the simulations, to solve this the amount of simulations could be increased and they could be constructed based on the isotonic model. Additionally, the problem of model misspecification in the parametric methods remains a challenge, and further research could focus on refining these models or exploring alternatives. Another area for improvement is the inclusion of industry-standard methods in the performance assessment, which would provide a clearer benchmark for evaluating the effectiveness of the NSP extensions. The S-shaped regression method, although described, was not fully tested in the current research, and future studies could explore its potential in more depth. Furthermore the application of the method could be expanded to solar energy, which is also of interest to Vanderbon.

In conclusion, this thesis demonstrates the significant potential of the extended NSP methods for improving change point detection in wind energy data. The isotonic regression model, in particular, shows great promise in capturing both abrupt and gradual shifts in turbine performance, leading to more accurate power forecasts and better maintenance planning. Despite some limitations, the findings provide a solid foundation for future research and development in the field, with opportunities to further refine the models and expand their application to other renewable energy sources.

Contents

1 | Introduction

In this chapter the motivation for the thesis project will be given. After this a few types of change points in the field of wind energy will be explained and the chapter ends with the research questions and thesis outline.

1.1 Motivation for change point detection

The urgency to combat climate change and reduce greenhouse gas emissions has accelerated the global shift from fossil fuels to renewable energy sources. According to [Eurostat, 2024] the EU the share of renewable energy consumed in 2022 was 23%, which is an increase of 1.1% from 2021. But the share of renewable demands a substantial growth to meet the target of 42.5% for 2030. With wind and solar being the two main sources at the moment to scale to meet the target. According to the Global Wind Energy Council (GWEC) 2023 was a record year for the wind industry, with 117GW of new capacity representing a 50% year-on-year increase from 2022, [GWEC, 2024].

Despite this rapid growth, the wind industry still faces several challenges that needs to be addressed to reach its full potential. One of the main challenges is mitigating the operations and maintenance (O&M) costs, which according to [Costa et al., 2021] account for 20% to 25% of total costs. Additionally, the variability of wind creates significant uncertainties in power output, complicating efforts to provide accurate and reliable power forecasts. These techniques enable the identification of performance deviation, faults, or structural changes in turbine behaviour. This can in turn be used to enhance forecasting methods, optimize maintenance schedules and reduce operational costs.

For Vandebron, a pioneering energy company committed to accelerating the energy transition by connecting renewable energy producers to the consumers, the effective use of machine learning models is crucial to optimizing their operations. One of the requirements for these models is high-quality input data. However, data collected from wind turbines often contain anomalies due to maintenance, equipment malfunctions or specific regulation regarding for example birds and bats or turbine shadow. As Vandebron continuous to expand their portfolio of wind energy assets, developing effective tools to identify faulty data from their wind assets is important to ensure their models are trained on accurate and reliable data, leading to better predictions.

Change point detection plays a significant role in this context by enabling the identification of data points where structural changes or deviations occur. In practise these changes mean periods of, for example maintenance, this is anomalous data and should not be used in training machine learning models. Excluding these periods can improve the accuracy of predictive models used in power forecasting and condition monitoring. Various methods for change point detection have been developed, ranging from traditional statistical methods to algorithms specific to wind energy. Each of these with its own strengths and limitations when applied to wind energy.

These techniques enable the identification of performance deviation, faults, or structural changes in turbine behaviour. This can in turn be used to enhance forecasting methods, optimize maintenance schedules and reduce operational costs.

1.2 Types of change points

The faulty data in wind turbine operations can be caused by various events, many of which involve the sudden shutdown of turbines. One major cause of such data issues is unplanned maintenance, where a wind turbine is unexpectedly shut down due to component failures or faults. This often occurs while production is predicted, leading to discrepancies in data and resulting in significant operational costs, as calculated by [Carroll et al., 2016].

Another common cause of data irregularities is curtailment, which involves the intentional reduction of power output due to external factors such as grid limitations or market conditions. This deliberate intervention, often initiated by grid operators, can result in unexpected drops in power production data [Bird et al., 2016]. Luckily, data on curtailment is usually available to the energy companies, thus not always necessary to detect. A related but more sustained impact on turbine data can occur due to a limit imposed on the grid inverter, which restricts the amount of power a turbine can deliver to the grid. This limit can be dynamically adjusted by the grid operator based on grid stability needs, as thoroughly explained by [Chinchilla et al., 2006].

In addition to these operational and market-driven events, there are also natural causes for turbine shutdowns that lead to faulty data. For example, bird collisions pose a significant threat for the birds themselves as well as the turbines, as large numbers of birds may fly into the turbine blades. This results in radar-assisted shutdowns to prevent these collisions, [Tomé et al., 2017]. Similarly, turbines may also be turned off temporarily to mitigate shadow flickering, a phenomenon where the moving turbine blades cast intermittent shadows that can be annoying to nearby residents [Haac et al., 2022].

Lastly, icing is a natural cause of turbine shutdowns, particularly in colder climates or during winter months. Ice accumulation on the turbine blades affects their aerodynamic properties, reducing power output and potentially causing damage if not addressed promptly. This is particularly a problem in regions with harsh winter conditions, as noted by [Barber et al., 2011].

By identifying this anomalous data with diverse causes, Vandebron can implement robust change point detection methods to ensure that their machine learning models remain accurate and reliable.

1.3 Research questions and thesis structure

The most important research questions that will be answered in this thesis are the following:

1. How to apply data segmentation methods to the wind energy power data to detect (partial) downtime?

2. Can the rigorous method be used to improve upon industry standards?

This thesis will start with literature review on different methods for change point detection, followed by a description of the power curve in Chapter 3, an important concept in this field. In Chapter 2 a description is given of current methods and their advantages and limitations. This is followed by the a description of the Narrowest Significance Pursuit method proposed by [Fryzlewicz, 2023] in Chapter 4 and five proposed adaptations in Chapter 5. Next, in Chapter 6 it is explained how these adaptations can be applied to wind energy and the results of the implementation are discussed in Chapter 7. The thesis ends with the conclusion, where the research questions will be answered and a discussion on the implications, limitations and recommendations following this project. Included with the thesis are two csv files with datasets called *clean_power_data* and *missed_production_data*, the usage of both will be provided at appropriate moments.

2 | Industry Methods for Change Point Detection and Outlier Filtering

With the rapid expansion of wind energy as an important component of the global renewable energy mix, the number of wind turbines installed worldwide has increased substantially. This growth, however, has brought new challenges that need to be addressed to ensure the sustainability and efficiency of wind energy production. One of the primary challenges is the escalating cost of operations and maintenance (O&M), which can, according to [Costa et al., 2021], constitute up to 25% of the total cost of wind energy production over a turbine's lifetime. Additionally, the highly unpredictable nature of wind introduces significant variability in energy output, making accurate power forecasting a complex yet essential task. High expectations for the reliability and stability of wind power have intensified the demand for precise and reliable power predictions.

To address these challenges, extensive research has been directed towards optimizing O&M strategies and improving the accuracy of wind power forecasts. Anomaly detection and change point detection have emerged as crucial tools in this context, enabling early identification of performance deviations, faults, or structural changes in turbine behavior. Detecting such changes can help mitigate potential failures, optimize maintenance schedules, enhance energy production efficiency, and ultimately reduce operational costs.

Given the importance of these techniques, various methods for change point detection have been developed and adapted specifically for wind energy applications. These methods range from traditional statistical approaches to advanced machine learning algorithms, each with its own strengths and limitations. Understanding the different methods available, and how they can be effectively applied in the wind energy sector, is essential for advancing turbine reliability and improving the overall economics of wind power.

Alongside change point detection, outlier filtering plays a crucial role in analyzing wind turbine data to ensure accurate modeling of the relationship between wind speed and power output. Outliers can distort this relationship, leading to inaccuracies in power curve models and false alarms or missed change points detection methods.

This chapter provides a comprehensive review of several methods used for change point detection and anomaly filtering, both in general and with specific reference to the wind energy industry. It explores a wide range of approaches, highlighting their core principles, advantages, and practical challenges.

2.1 Methods for anomaly detection

In this section we will outline some methods for anomaly detection in the power output of wind turbines. These methods can form the basis of change point detection methods.

2.1.1 Filtering with logistic functions based on quantile regression

The first method is a parametric method specifically designed for wind power, described by [Jing et al., 2021]. They propose a combination of quantile regression (QR) with logistic func-

tions (LF) for wind turbine power curve modeling. It is first discussed how to use the logistic function to capture the uncertainty of wind power output, by estimating different quantiles (e.g., 5%, 50%, 95%) of the power curve. Next, an adaptive outlier filtering technique is developed base on the QRLF combination. This technique removes sparse and stacked outliers in the power data, using the symmetrical properties of the wind power distribution.

The core principle behind this outlier filtering is that when a wind turbine functions normally, the distribution of power output at a given wind speed is approximately symmetrical around the median. This implies that the distances between the median quantile (PC_{q50}) and the lower (PC_{q5}) and upper (PC_{q95}) quantiles should be roughly the same. Any significant deviation from this symmetry suggests the presence of outliers, or change points. The threshold in the article is chosen by the cross validation of multiple wind turbines, however no further proofs as to the performance of the method is given. Once outliers are detected they are removed from the data set and the quantiles recalculated to reflect the remaining data. This process repeats itself iteratively until the threshold is no longer exceeded.

This method does come with its limitations. Most importantly, the accuracy of QRLF is influenced by the initial parameter values chosen for the optimization of the logistic function. Sub optimal initial settings can lead to converging to local optima, in the optimization routine of the algorithm, which affects the overall fitting performance. The authors suggest enhancing the reliability by repeating the optimization step multiple times. While this method does filter outliers from the data, its main focus is to model the wind turbine power curve.

However, the algorithm in Section 2.2.1 improves on this method by using it in combination with Bayesian change point detection, providing a probabilistic framework for identifying abrupt changes in wind turbine data.

2.1.2 Isolation forest and deep learning neural networks

A fundamentally different approach to anomaly detection was proposed for the first time by [Liu et al., 2008], called the isolation forest. This type of algorithm uses binary trees and since its publication in 2008 the algorithm has often been applied and enhanced. The isolation forest, or iForest, is described in an overview of anomaly detection methods by [Bilendo et al., 2022]. It is a tree-based method to detect anomalies by isolating outliers rather then defining normal data. It does this by randomly selecting features and splitting values to isolate data points. The idea is that anomalies are isolated in fewer steps because they are less frequent and distinct from normal data. This method is thus only useful to filter clear outliers, since if it is a pattern change or these outliers occur often, the isolation is not easily done.

The method is successfully applied by [Lin et al., 2020] to clean data before using it to improve deep learning neural networks for power prediction.

2.2 Methods for change point detection

The techniques described in this section are specifically developed as change point detection methods. Starting with a Bayesian extension to one of the anomaly detection method. Followed by a combination of two statistical tools; cumulative sum and marked empirical processes. Lastly, a kernel-based machine learning technique is outlined.

2.2.1 Bayesian change point-quartile combined algorithm

The Bayesian change point-quartile combined algorithm, a method proposed by [Wang et al., 2023], is designed for cleaning abnormal data in wind turbine power curves. It uses a Bayesian approach for change point detection, after which it uses the data filtering method of [Jing et al., 2021].

The first step is identifying the boundary between normal and abnormal data within the power curve. In order to do so, they calculate a difference sequence, which is the difference

between the power value and the maximum power observed within a specific wind speed range. The core assumption here is that this power difference sequence can be represented as two uniform distributions, one for normal data and the other for the abnormal data.

The Bayesian approach defines prior distributions for the parameters of the two uniform distributions, after which the parameters are determined based on Jeffreys criteria. The prior distribution for the change point is also uniform, since there is no knowledge about its location.

Using Bayes' theorem, the method constructs a posterior distribution to estimate the change point. The change point is identified as the time at which the posterior probability reaches its maximum.

Advantages of this method are the effectiveness and efficiency. Which means the combined algorithm is shown to be effective in identifying and removing various types of outliers. Aside from this the Bayesian method is computationally efficient since it doesn't need extensive iterative calculations. However, it also has its limitations. For one, the assumption that the power difference sequence follows two uniform distributions may not always be the case. In addition to this, the performance may be sensitive to the choice of prior distributions and parameters. Lastly, the method focuses on detecting a single change point, whereas it often happens that more change points occur in one dataset.

2.2.2 CUSUM combined with marked empirical processes

[Mohr and Neumeyer, 2020] developed a nonparametric procedure to detect change points in time series data. The model considers a weakly dependent time series with multivariate covariates and univariate observations. The main objective is to test whether the nonparametric conditional mean remains stable over time or if it experiences changes. In order to do so, two statistical tools are combined: the CUSUM (cumulative sum) method and marked empirical processes. The CUSUM method is a classic technique for change point detection, also in general considered by [Honda, 1997] and [Su and Xiao, 2008], and for processes in wind turbines in particular by [Xu et al., 2020], [Dao, 2021] and [Latiffianti et al., 2022]. Marked empirical processes have been suggested by [Stute, 1997] in order to assess the goodness-of-fit in nonparametric regression models with i.i.d. data.

At the core of the methodology of Mohr and Neumeyer lies the CUSUM test statistic, which is constructed based on the residuals obtained from a nonparametric estimate of the regression function. The CUSUM statistic accumulates the deviations between the observed values and their corresponding predicted values, calculated as:

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (Y_i - \hat{m}(X_i)),$$

where Y_i are the observed values, $\hat{m}(X_i)$ is the estimated conditional mean from the nonparametric regression, n is the total number of observations, and t is a time index ranging from 0 to 1. This statistic is particularly sensitive to shifts in the mean, making it suitable for detecting both gradual and abrupt changes in the data. To evaluate whether these deviations are statistically significant, the authors define a marked empirical process that monitors the distribution of the residuals across different covariate values and over time. Under the null hypothesis (no change points), this process converges weakly to a centered Gaussian process, which allows for the derivation of asymptotic distributions for commonly used tests, such as the Kolmogorov-Smirnov and Cramér-von-Mises tests.

In practical applications, if the calculated CUSUM statistic exceeds critical values derived from these tests, it suggests a change in the mean, leading to the rejection of the null hypothesis. However, the authors note that the standard asymptotic results depend on the strict stationarity of the residuals, which may not always hold in real-world datasets characterized by inherent variability. To enhance the method's applicability, particularly in cases of multivariate covariates

or non-stationary variance, a wild bootstrap approach is recommended for estimating critical values. This is crucial for maintaining robust inference in situations where classical assumptions may not hold, thus ensuring the method remains effective even in complex operational datasets, such as those encountered in wind energy systems.

Another advantage of this method, aside from the possibility of wild bootstrapping, is the nonparametric nature, thus not relying on strong assumptions about the functional form of the regression function. In addition to that, the fact that the approach accommodates weakly dependent time series data makes it suitable for data on wind energy production. Lastly, the test is shown to be consistent against fixed alternatives of one change point, which ensure its ability to detect changes as the sample size increases.

The procedure, however, also has a limitation, the method primarily focuses on detecting a single change point. Hence, for out aim to be able to detect multiple changes within a time period, this is not preferred.

The CUSUM method has directly been used in wind turbine processes but with different approaches. [Xu et al., 2020] propose an adaptive fault detection scheme combining Random Forest (RF) and CUSUM to monitor wind turbines. The method leverages RF to model the non-linear relationship between multiple supervisory control and data acquisition (SCADA) variables, generating residuals that are then monitored using an adaptive CUSUM chart. This allows for the detection of faults while minimizing false alarms, even in noisy operational environments. Although the method is focused on detecting anomalies in variables such as temperature and generator efficiency, it includes power output as one of the key SCADA parameters being monitored. In comparison, [Latiffianti et al., 2022] also use a CUSUM-based method but focus on detecting gearbox failures using Minimum Spanning Tree (MST)-based anomaly scores, which capture system-wide operational anomalies. While both approaches aim to reduce false alarms and improve fault detection, Latiffianti's method emphasizes detecting gradual system degradation, whereas Xu's scheme focuses more on adaptive monitoring of abrupt changes. [Dao, 2021], on the other hand, applies CUSUM to detect structural changes in regression models based on temperature data, with a less direct focus on power output. Xu's method, with its integration of RF, provides more flexibility in capturing non-linear patterns, making it highly robust in diverse fault detection scenarios compared to the more specialized methods of Latiffianti and Dao.

2.2.3 Kernel-Based change point detection

Besides statistical methods, there are also methods based on machine learning to detect change points in data. These methods are broadly used for classifying the state of a turbine, or one of its components, as either healthy or faulty. This is done by using available SCADA data of a wind turbine. The following kernel-based method uses the broad range of data to determine the state of a wind turbine.

In the field of machine learning, many methods developed for monitoring the state of wind turbines operate within a semi-supervised framework, commonly referred to as normal behavior modeling (NBM). These methods leverage SCADA data to differentiate between healthy and faulty states of turbine components. One of the first to describe this method were [Schlechtingen and Santos, 2011] after which more research in this field followed. A comparative study on these and their challenges was done by [Leahy et al., 2019]. Most issues identified were in relation to data preprocessing and [Letzgus, 2020] addresses these to improve on the method. To overcome these problems he executes three steps: removal of non-operational periods, normalization with operational state and ambient conditions, and resampling with reduced temporal resolution. After this, the method uses a Laplacian Kernel, also explained by [Garreau and Arlot, 2018], which is suited for change point detection. It measures similarity between data points and identifies moments where the data deviates from a normal pattern. In this approach they also make use of SCADA data and divide it into segments, the kernel function is applied to compute the similarity between those segments. To quantify the dissimilarity between

segments a cost function $C(\tau)$ is defined. The objective is to find change points by minimizing this cost function. To ensure that only the most significant changes are detected a regularization term $P(\tau)$ is introduced, alongside the cost function, which penalizes adding too many change points. Lastly, the Laplacian kernel was found to perform best with a batch standard deviation heuristic for bandwidth selection, so that the method can adapt to different signals.

An advantage of this method is again the non parametric flexibility. Apart from that it is robust against diverse data types, which is important when handling the different SCADA signals. On the downside, the method's performance is highly dependent on the choice of kernel, bandwidth and penalty terms and requires extensive preprocessing of the data. Aside from this, the model struggles with detecting multiple change-points of varying significance levels within the same signal.

3 | Power Curve

In this chapter the power curve, which is the relation between wind speed and power, will be described and different representations used in literature will be given.

3.1 Introduction

The output power of a wind turbine is dependent on many factors but most importantly on the wind speed at the hub height. The relation between the power output and wind speed is given by the power curve and is specific for a turbine model. In practise there may even be slight differences between turbines of the same type depending on their location and configuration.

The minimum wind speed at which a turbine produces power is called the cut-in speed (V_{ci}). The rated speed (V_r) is the wind speed at which the maximum power is obtained, called the rated power. The power increases cubical between the cut-in and rated wind speeds. Lastly, the cut-out speed (V_{co}) is the maximum wind speed at which the turbine can generate power, after this is it shut off. This results in the following theoretical curve.

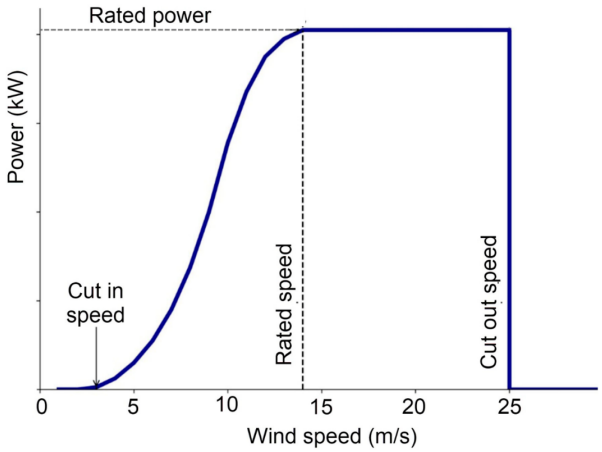


Figure 3.1: Power curve

Usually a manufacturer supplies the power curve specific for their turbine model. This is usually done by cutting the range of wind speeds in intervals of $0.5m/s$. For each of these intervals the according power output is given, which can be interpolated to obtain a smooth curve. Instead of this curve provided by the manufacturer, for many applications a generic mathematical description is preferred. Such as in studies about the modelling of wind turbine power generation, analysis of wind energy potential or cost modelling, listed by [Carrillo et al., 2013]. The mathematical formulations of the curve will be discussed in the next section.

3.2 Types of power curve representations

The modelling techniques for the power curve can be divided into parametric and non-parametric models as discussed by [Lydia et al., 2014]. The most common and relevant for our method are the linearized segmented model, polynomial, cubic, exponential and logistic functions, which are all parametric models. To understand the formulation of these models, we will first provide a brief formulation of the potential power in the wind that crosses a turbine described in [Johnson, 1985]. This value can be obtained by

$$p_w(v) = \frac{1}{2} A \rho v^3, \tag{3.1}$$

where $p_w(v)$ is the power in the wind, ρ is the air density in kg/m^3 and A is the rotor area in m^2. Typically ρ is taken as $1.225 kg/m^3$, [Wind Turbines-Part, 2005]. However, the actual extracted power is a fraction of $p_w(v)$ which is determined by the power coefficient $C_p(V)$ and $p(v) = C_p(V) p_w(v)$. The maximum theoretical value of the power coefficient is know as the Betz limit with a value of 0.593. Because of mechanical imperfections this value is not achievable with real turbines, hence the maximum is normally around 0.5 and 0.4 is considered a good turbine.

The parametric models will first be stated and then compared to find the optimal one. For the cubic, exponential and polynomial curve the general representation is as follows:

$$p(v) = \begin{cases} 0 & v \leq v_{ci} \text{ or } v \geq v_{co} \\ q(v) & v_{ci} \leq v \leq v_r \\ p_{rated} & v_r \leq v \leq v_{co} \end{cases} \tag{3.2}$$

Here p_{rated} is the rated power and the difference in models lies in the formulation of $q(v)$, those will be provided below.

3.2.1 Linearized segmented

The linearized segmented model is the simplest model. It uses piece-wise approximation of the curve and follows the equation $p = mv + c$, where m is the slope of the segment and c a constant. The data is fitted using least squares estimation. This model is in line with the segmented approximation provided by manufacturers. As stated by [Bilendo et al., 2022], since the power curve itself is not linear, the linear method gives the least accurate results.

3.2.2 Cubic power curve

The cubic representation resembles the theoretic power curve closely. As described by [Jangamshetti and Rau, 2001], the cubic equation is

$$q(v) = \frac{1}{2} \rho A C_{p,\text{eq}} v^3, \tag{3.3}$$

where $C_{p,\text{eq}}$ is a constant equivalent to the power coefficient. This however only provides a function for the first part of the power curve, until the rated wind speed and power. This mean in reality you will have to add another constant for the wind speeds that produce the rated power and for the wind speed after the cut out speed, resulting in three different areas which was definer in equation 3.2.

3.2.3 Exponential power curve

An adaptation of the cubic power curve it the exponential method. This is given by,

$$q(v) = \frac{1}{2} \rho A K_p \left(v^\beta - v_{ci}^\beta \right), \tag{3.4}$$

where K_p and β are constants. To derive the cubic curve from this equation we can set β equal to 3 and v_{ci} equals 0.

3.2.4 Polynomial power curve

The next model is a second order polynomial model. As described by [Wen et al., 2009], in this case the power output can be calculated as

$$q(v) = (A + Bv + Cv^2).$$

The constants A, B, and C may be found as functions of v_{ci} and v_r using the following equations:

$$A = \frac{1}{v_{ci} - v_r}\left[v_{ci}(v_{ci} + v_r) - 4(v_{ci} \times v_r)\left(\frac{v_{ci} + v_r}{2v_r}\right)^3\right] \tag{3.5}$$

$$B = \frac{1}{(v_{ci} - v_r)^2}\left[4(v_{ci} + v_r)\left(\frac{v_{ci} + v_r}{2v_r}\right)^3 - (3v_{ci} + v_r)\right] \tag{3.6}$$

$$C = \frac{1}{(v_{ci} - v_r)^2}\left[2 - 4\left(\frac{v_{ci} + v_r}{2v_r}\right)^3\right] \tag{3.7}$$

The performance of the cubic, exponential and polynomial power curves were tested by [Villanueva and Feijóo, 2018]. Using two parameters, the coefficient of determination R^2 and the mean energy production error. It was found that the exponential and cubic curve give good results and perform the best.

3.2.5 Logistic power curve

Another representation for the power curve are logistic curves, also called 's' shaped curves or sigmoid functions. They can have various amounts of parameters, depending on the application and expert knowledge to determine the parameters. In case of the power curve these parameters can include, minimum and maximum power, the infliction point or the slope. The logistic functions are widely used and have provided good results, as presented by [Renani et al., 2016], [Seo et al., 2019] and [Lydia et al., 2015].

The logistic functions with a range of 3 to 6 parameters as a predictor for the power output have been compared by [Villanueva and Feijóo, 2018]. It was found that the functions with 5 and 6 parameters perform the best, however these require for a lot of parameters to be determined. The results of the 3 and 4 parameter functions are not as good but still perform well. Considering the good results of this and the fact that dealing with less parameters is preferable we chose to give the 3 parameter logistic function. This results in the function

$$q(v) = \frac{\alpha}{1 + \exp\left(-\beta \cdot (v - v_0)\right)}, \tag{3.8}$$

where α is the curve's maximum value, thus rated power, v_0 is the value of the inflection point and β is the slope of the curve. The logistic curve not only describes $q(v)$ from equation 3.2 but also q_{rated} and the wind speeds before the cut-in wind speed, due to the shape of the S-curve.

4 | Narrowest Significance Pursuit

In this chapter we introduce a method for change point detection, which is different from the ones discussed in Chapter 2 and will be the basis for the algorithm of this thesis. The Narrowest Significance Pursuit (NSP), proposed by [Fryzlewicz, 2023] is a method to determine intervals within a data sequence, with a global significance level, that each contain a change point. This chapter will provide an introduction to the algorithm in Section 4.1. The actual framework, including the pseudo code, is discussed in Section 4.2. This is then followed by a detailed explanation of the core step in Section 4.3.

4.1 Introduction

The narrowest significance pursuit is designed for identifying change-points in time series data, which are instances where the underlying pattern of the data changes significantly. The ability to pinpoint these changes is crucial in fields such as finance, economics, climatology and signal processing. In short, the method consists of 4 steps; segmentation, hypothesis testing, significance threshold and narrowest significance pursuit. First, the dataset is partitioned into segments, of which each represents a part of the dataset. Next, within each segment, the presence of change-points is determined by performing hypothesis testing. A significant threshold is established and used to assess if the level of statistical evidence, obtained in the hypothesis test, is good enough to declare a change-point is present. Lastly, the method iteratively examines the segments to identify the most narrow one.

The following example illustrates the change in the underlying model and the locations of change points. The blue continuous line resembles the power produced and the vertical dotted lines are the starts and endings of the anomalous periods. In this example the power can only produce 20% of the normal power in between the change-points. This means the relation between the wind speed and power is different which means the underlying pattern of the data has changed. (see Section 6.4 for details)

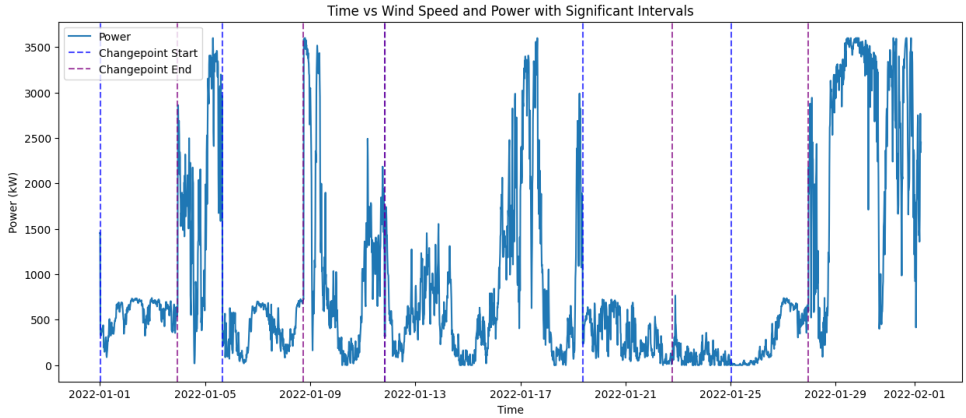


Figure 4.1: Simulated power with change points to 20% remaining capacity.

NSP emphasizes controlling the Type I error, which refers to the probability of falsely identifying a change point when there is none. Fryzlewicz describes three scenarios of underlying models of which each scenario is a generalization of the preceding one. Throughout the description of the method, the third and thus most generalized scenario will be used, which is the linear regression with piece-wise constant parameters. This scenario states; For $X = (X_{t,i})$ a given design matrix, $t = 1, \ldots, T$, $i = 1, \ldots, p$, Y_t follows the model

$$Y_t = X_{t,\cdot}\beta^{(j)} + Z_t \text{ for } t = \eta_j + 1, \ldots, \eta_{j+1},\ j = 0, \ldots, N, \tag{4.1}$$

where the parameter vectors $\beta^{(j)} = (\beta_1^{(j)}, \ldots, \beta_p^{(j)})'$ are such that $\beta^{(j)} \neq \beta^{(j+1)}$. Which means the change-points mark the transitions between the piece-wise segments and thus the change in parameters.

The definition and notation of a change-point is as follows. A **change-point** is the moment where the distribution of a time series changes. These points are denoted as η_j, $j = 0, \cdots N$ where N denotes an unknown number of change-points and the locations are $0 = \eta_0 < \eta_1 < \cdots < \eta_N < \eta_{N+1} = T$.

4.2 The linear framework

First the pseudo code will be provided which will be used to explain the algorithm more elaborately. This is followed by detailed descriptions and proofs of certain parts of the algorithm. The NSP function has input variables $s, e, Y, X, M, \lambda_\alpha, \tau_L, \tau_R$, which are listed in table 4.1.

Variable	Description
s	Starting index of the interval
e	Ending index of the interval
Y	Time series data
X	Covariates or explanatory variables
M	Number of intervals to consider
λ_α	Threshold for significance testing
τ_L	Left-sided degree of overlap with a significance interval
τ_R	Right-sided degree of overlap with a significance interval

Table 4.1: Input Variables for NSP Algorithm

On initialization of the NSP function the interval $[s, e] = [1, T]$; Y with length T and X with dimensions $T \times p$ are as in equation 4.1. Typical values for the global significance level α of the threshold are 0.05 or 0.1. The algorithm is launched by the pair of calls $S = \varnothing$ and $\text{NSP}(1, T, Y, X, M, \lambda_\alpha, \tau_L, \tau_R)$ and in each iteration of the function, an interval of global significance level is added to the output set S.

The algorithm will now be given and described line by line.

Algorithm 1 NSP

1: **function** NSP$(s, e, Y, X, M, \lambda_\alpha, \tau_L, \tau_R)$
2: **if** $e - s < 1$ **then**
3: **return**
4: **else**
5: M-intervals = DrawSampleOfIntervals(s,e,M)
6: **end if**
7: **for** $m = 1, \ldots, M$ **do**
8: $D[s_m, e_m] =$ DeviationFromLinearity(s_m, e_m, Y, X)
9: **end for**
10: $M_0 = \arg\min_m\{e_m - s_m : m = 1, \ldots, M; D[s_m, e_m] > \lambda_\alpha\}$
11: **if** $|M_0| = 0$ **then**
12: **return**
13: **end if**
14: $m_0 = \arg\max_m\{D[s_m, e_m] : m \in M_0\}$
15: $[\tilde{s}, \tilde{e}] =$ ShortestSignificantSubinterval$(s_{m_0}, e_{m_0}, Y, X, M, \lambda_\alpha)$
16: add $[\tilde{s}, \tilde{e}]$ to the set S of significant intervals
17: NSP$(s, \tilde{s} + \tau_L(\tilde{s}, \tilde{e}, Y, X), Y, X, M, \lambda_\alpha, \tau_L, \tau_R)$
18: NSP$(\tilde{e} - \tau_R(\tilde{s}, \tilde{e}, Y, X), e, Y, X, M, \lambda_\alpha, \tau_L, \tau_R)$
19: **end function**

Firstly, the segmentation part of the algorithm is executed in lines 2-6, this makes certain a representative sample of intervals is taken, likely use a sliding window with overlap to cover all the data. Secondly, line 8 computes the local test statistics, this is the core step of the algorithm and will be discussed in Section 4.3. The test is performed by thresholding in line 10 and the smallest interval of significance is stored in M_0. If this set is empty the algorithm is terminated. Otherwise, in line 15-18 the interval in which the greatest deviation occurs is selected and it is checked if a shorter sub interval, $[\tilde{s}, \tilde{e}]$, can be found. This sub interval is then added to S and NSP will be executed twice more on intervals left and right of the one found.

4.3 Deviation from linearity

In this section the DeviationFromLinearity in the NSP function will be discussed in detail. This deviation determines to what extent the obtained data differs from the linear model, which in turn will indicate the intervals containing a change-point. What happens during this determination is that for a chosen interval (from the set drawn earlier in ref) parameters β will be fitted, such that the deviation of the model from the true values is as little as possible. With this parameter fit, the maximum deviation over certain sub intervals is calculated with a loss function, and the value will be returned and continued with in the algorithm.

In order to do so, a scaled partial sum statistic is chosen and a scan (or window) statistic with respect to the interval set. For this measure two results are obtained. The first providing an upper-bound on the deviation under the null hypotheses and the second guarantees an upper bound on the probability of Type I errors occurring. To obtain these results some definitions and a lemma are given.

Definition 4.1. *For an interval set $\mathcal{I}$, the sup-norm of an arbitrary input sequence $\{y_t\}_{t=1}^T$ is defined as*

$$\|y\|_{\mathcal{I}} = \max_{[s,e]\in\mathcal{I}} |U_{s,e}(y)|, \tag{4.2}$$

where $U_{s,e}(y) = (e - s + 1)^{-1/2} \sum_{t=s}^{e} y_t$ is the scaled partial sum statistic.

For computational efficiency instead of using the set of all sub intervals, $\mathcal{I}^a$, the set $\mathcal{I}^d$ of dyadic lengths as in the definition below will be used.

Definition 4.2. *For $j = 0, \ldots, \lfloor log_2 T \rfloor$ the set of all intervals of dyadic lengths and arbitrary locations on [0,T] is defined as*

$$\mathcal{I}^d = \{[s, e]\} \subseteq [1, T] : e - s = 2^j - 1, \tag{4.3}$$

and restrictions to arbitrary intervals [s,e] as

$$\mathcal{I}^d_{[s,e]} = \{[u, v]\} \subset [s, e] : [u, v] \in \mathcal{I}^d\}, \tag{4.4}$$

and analogously for $\mathcal{I}^a_{[s,e]}$.

The norms $\|\cdot\|_{\mathcal{I}^a}$, $\|\cdot\|_{\mathcal{I}^d}$ and their restrictions are referred to as multi-resolution sup-norms. Given that $\mathcal{I}^d_{s,e} \subseteq \dashv^d_{s,e} \subseteq \mathcal{I}^a$ and $\mathcal{I}^d_{s,e} \subseteq \mathcal{I}^d \subseteq \mathcal{I}^a$ Lemma 4.1 is a trivial consequence.

Lemma 4.1. *For any interval [s,e] and input vector y,*

$$\|y_{s:e}\|_{\mathcal{I}^d_{[s,e]}} \leq \|y_{s:e}\|_{\mathcal{I}^a_{[s,e]}} \leq \|y\|_{\mathcal{I}^a} \tag{4.5}$$

and

$$\|y_{s:e}\|_{\mathcal{I}^d_{[s,e]}} \leq \|y\|_{\mathcal{I}^d} \leq \|y\|_{\mathcal{I}^a}. \tag{4.6}$$

Using this multi-resolution sup-norm as the loss function the DeviationFromLinearity is denoted as follows,

$$D_{[s_m,e_m]} = \min_{\beta} \|Y_{s_m,e_m} - X_{s_m,e_m,.}\beta\|_{\mathcal{I}^d_{[s_m,e_m]}}. \tag{4.7}$$

The following property, Proposition 4.1 is fundamental for NSP. It states that under the null hypothesis of no change-point on [s,e], the DeviationFromLinearity $D_{[s_m,e_m]}$ has an upper-bound, defined as $\|Z_{s:e}\|_{\mathcal{I}^d_{[s,e]}}$.

Proposition 4.1. *Let the interval [s,e] be such that it contains no change-points, then the DeviationFromLinearity is bounded by the norm of the true residuals on that interval, i.e.*

$$D_{[s,e]} \leq \|Z_{s:e}\|_{\mathcal{I}^d_{[s,e]}}. \tag{4.8}$$

Proof. The interval [s,e] does not contain a change-point, hence there exists a β^* such that $Y_{s:e} = X_{s:e,.}\beta^* + Z_{s:e}$. Therefor,

$$\begin{aligned} D_{[s,e]} &= \min_{\beta} \|Y_{s:e} - X_{s:e,.}\beta\|_{\mathcal{I}^d_{[s:e]}} \\ &\leq \|Y_{s:e} - X_{s:e,.}\beta^*\|_{\mathcal{I}^d_{[s:e]}} \\ &= \|Z_{s:e}\|_{\mathcal{I}^d_{[s:e]}} \end{aligned}$$

Which holds because the same norm is used for the linear model fit and for the residuals. □

In this proposition the deviation between the measured values and the model based on β is minimized under an optimal β. This deviation is bound from above. In Section 4.3.1 a different optimal β is used when following a different approach. This bound leads to the second important result, Theorem 4.1, providing an upper bound on the probability of NSP falsely returning an interval of significance. In other words, it bounds the probability of Type I errors occurring above by $\alpha = P(\|Z\|_{\mathcal{I}^a} > \lambda_\alpha)$.

Theorem 4.1. *Let $S = \{S_1, \cdots, S_R\}$ be a set of intervals containing change points, returned by the NSP algorithm. Then the probability that there exists a returned interval without a change point is bounded above, i.e.*

$$P(\exists i = 1, \cdots, R \; \forall j = 1, \cdots, N \; [\eta_j, \eta_j + 1] \nsubseteq S_i) \leq P(\|Z\|_{\mathcal{I}^d} > \lambda_\alpha) \leq P(\|Z\|_{\mathcal{I}^a} > \lambda_\alpha). \quad (4.9)$$

Proof. The set S was returned by NSP which, by line 10 of NSP, means $D_{S_i} > \lambda_\alpha$ for all i. By Proposition 4.1, the fact that there exists an interval not containing a change point means there is an i such that $D_{S_i} \leq \|Z_{S_i}\|_{\mathcal{I}^d_{S_i}}$. Thus $\lambda_\alpha < D_{S_i} \leq \|Z\|_{\mathcal{I}^d}$, which proves the first inequality. Inequality 4.6 states $\|Z_{S_i}\|_{\mathcal{I}^d_{S_i}} \leq \|Z\|_{\mathcal{I}^d} \leq \|Z\|_{\mathcal{I}^a}$, which proofs the second inequality. □

4.3.1 X dependent thresholds

The optimal significance threshold λ_α can vary for different segments of the data since it may depend on the level of noise or variability within a segment. Which is why the thresholds obtained in Theorem 4.1 and Proposition 4.1 can be tightened by making them dependent on the matrix X. This leads to alternative versions of both the proposition and theorem in which the sensitivity and accuracy of the change-point detection is improved. In the proposition below the aim is again to find an optimal β, however in this case it is to minimize the difference between the error and the design matrix.

Proposition 4.2. *Let the interval [s,e] be such that it contains no change-points, then*

$$D_{[s,e]} = \min_\beta \|Z_{s:e} - X_{s:e,\cdot}\beta\|_{\mathcal{I}^d_{[s,e]}} \leq \min_\beta \|Z - X\beta\|_{\mathcal{I}^d}. \quad (4.10)$$

Proof. For β fixed, the set $\mathcal{I}^d_{[s,e]}$ is a subset of $\mathcal{I}^d$ and is thus smaller. This means the sup-norm $\|Z_{s:e} - X_{s:e,\cdot}\beta\|_{\mathcal{I}^d_{[s,e]}} \leq \|Z - X\beta\|_{\mathcal{I}^d}$, which proofs the inequality. For the equality it is known since $[s, e]$ does not contain a change point, there exists a β^* such that $Y_{s:e} = X_{s:e,\cdot}\beta^* + Z_{s:e}$. Which gives,

$$\begin{aligned}
D_{[s,e]} &= \min_\beta \|Y_{s:e} - X_{s:e,\cdot}\beta\|_{\mathcal{I}^d_{[s:e]}} \\
&= \min_\beta \|X_{s:e,\cdot}\beta^* + Z_{s:e} - X_{s:e,\cdot}\beta\|_{\mathcal{I}^d_{[s:e]}} \\
&= \min_\beta \|Z_{s:e} - X_{s:e,\cdot}(\beta - \beta^*)\|_{\mathcal{I}^d_{[s:e]}} \\
&= \min_{(\beta - \beta^*)} \|Z_{s:e} - X_{s:e,\cdot}(\beta - \beta^*)\|_{\mathcal{I}^d_{[s:e]}} \\
&= \min_\beta \|Z_{s:e} - X_{s:e,\cdot}\beta\|_{\mathcal{I}^d_{[s:e]}}
\end{aligned}$$

□

Now following this proposition, a tighter bound on the probability of Type I errors is found as well. With the proof similar to that of Theorem 4.1

Theorem 4.2. *Let $S = \{S_1, \cdots, S_R\}$ be a set of intervals returned by the NSP algorithm. Then*

$$P(\exists i = 1, \cdots, R \; \forall j = 1, \cdots, N \; [\eta_j, \eta_j + 1] \nsubseteq S_i) \leq P(\min_\beta \|Z - X\beta\|_{\mathcal{I}^d} > \lambda_\alpha). \quad (4.11)$$

Proof. On the set $\min_\beta \|Z - X\beta\|_{\mathcal{I}^d} \leq \lambda_\alpha$, each interval S_i must contain a change-point. Since if it would not, by Proposition 4.2 that would mean

$$D_{S_i} \leq \min_\beta \|Z - X\beta\|_{\mathcal{I}^d_{[s,e]}} \leq \lambda_\alpha. \quad (4.12)$$

However, since S_i was returned by NSP, by line 14 this means $D_{S_i} > \lambda_\alpha$ which contradicts the inequality above. □

5 | Adaptations of NSP

The existing NSP method is restricted to a range of scenarios in which the most generalized is a linear regression with piece-wise constant parameters. Since this approach is not suitable for all time-series, it will be expanded to non-linear scenarios as well. This can be approached in two ways, they differ in the definition and usage of the noise. Both versions will be discussed in this chapter.

5.1 The non-linear frameworks

The original framework described in Section 4.2 is adapted such that it can be used for non-linear models. This gives several extensions to model 4.1, in which Y_t is described by one of the following models.
Non-linear with output error:

$$Y_t = f(X_{t,\cdot}, \beta^{(j)}) + Z_t \tag{5.1}$$

Non-linear with input error:

$$Y_t = f(X_{t,\cdot} + Z_t, \beta^{(j)}) \tag{5.2}$$

for $t = \eta_j + 1, \ldots, \eta_{j+1}$, $j = 0, \ldots, N$, and f a monotone function.
Isotonic regression:

$$Y_t = g(X_{t,\cdot}) + Z_t, \tag{5.3}$$

where g is an monotone function.
S-Shaped curve:

$$Y_t = h(X_{t,\cdot}) + Z_t \tag{5.4}$$

where h is an S-shaped curve, described in detail in Section 5.6.2.

In all scenarios the noise Z_t is assumed to be Gaussian. The main structure of the algorithm will stay the same, however the step DeviationFromLinearity will be altered. The alterations are slightly different for each version.

5.2 Non-linear with output error

The first non-linear scenario, scenario 5.1, represents a state where X_t can be thought of as a time-series of a measured variable. Y_t is a function of this variable and an error can occur in the transformation which is why Z_t is added. The only requirement for f is that it should be a function on $\mathbb{R}$ and Z_t is noise with 0 mean. Under these circumstances the statistic and loss function of the original framework, stated in definition 4.1, will be reused. The two main results will be stated and proven for this new scenario. In order to do so the 'DeviationFromLinearity'

step will be changed to 'ModelDeviation', since the difference that will be determined is no longer from a linear function. It is denoted as follows,

$$D_{[s,e]} = \min_{\beta} \|Y_{s:e} - f(X_{s:e,.}\beta)\|_{\mathcal{I}^d_{[s:e]}}. \tag{5.5}$$

where f is a function on $\mathbb{R}$.

Recall proposition 4.1 stating that on an interval that doesn't contain a change-point, the deviation $D_{[s,e]}$ is bounded from above, this proposition is adapted to the version below.

Proposition 5.1. *Let the interval [s,e] be such that it contains no change-points, then the model deviation is bounded by the norm of the true residuals on that interval, i.e.*

$$D_{[s,e]} = \min_{\beta} \|Y_{s:e} - f(X_{s:e,.}\beta)\|_{\mathcal{I}^d_{[s:e]}} \leq \|Z_{s:e}\|_{\mathcal{I}^d_{[s,e]}}, \tag{5.6}$$

for f a function on $\mathbb{R}$.

Proof. Since the interval [s,e] does not contain a change-point, there exists a β^* such that $Y_{s:e} = f(X_{s:e,.}, \beta^*) + Z_{s:e}$. Therefore,

$$\begin{aligned} D_{[s,e]} &= \min_{\beta} \|Y_{s:e} - f(X_{s:e,.}, \beta)\|_{\mathcal{I}^d_{[s:e]}} \\ &\leq \|Y_{s:e} - f(X_{s:e,.}, \beta^*)\|_{\mathcal{I}^d_{[s:e]}} \\ &= \|Z_{s:e}\|_{\mathcal{I}^d_{[s:e]}} \end{aligned}$$

Which holds because the same norm is used for the model fit and for the residuals, and f is a function on $\mathbb{R}$. □

The proof is very similar to that of the original scenario and almost no restrictions are given to the choice of function f. This also goes for the second proof, of theorem 4.1, bounding the probability of falsely detecting a change point.

Theorem 5.1. *Let $S = \{S_1, \cdots, S_R\}$ be a set of intervals containing change points, returned by the NSP algorithm and f is a function on* $\mathbb{R}$. *Then the probability that there exists a returned interval without a change point is bounded from above, i.e.*

$$P(\exists i = 1, \cdots, R \; \forall j = 1, \cdots, N \; [\eta_j, \eta_j + 1] \nsubseteq S_i) \leq P(\|Z\|_{\mathcal{I}^d} > \lambda_\alpha) \leq P(\|Z\|_{\mathcal{I}^a} > \lambda_\alpha). \tag{5.7}$$

Proof. The set S was returned by NSP which, by line 8 of NSP, means $D_{S_i} > \lambda_\alpha$ for all i. By proposition 5.1, the fact that there exists an interval not containing a change point means there is an i such that $D_{S_i} \leq \|Z_{S_i}\|_{\mathcal{I}^d_{S_i}}$, which proves the first inequality. Inequality 4.6 states $\|Z_{S_i}\|_{\mathcal{I}^d_{S_i}} \leq \|Z\|_{\mathcal{I}^d} \leq \|Z\|_{\mathcal{I}^a}$, which proofs the second inequality. □

5.3 Non-linear with input error

In this section the second non-linear scenario, described in 5.2 will be discussed. This situation differs from the first one in the definition of the noise Z_t. The noise will represent the measurement error by being added, inside the function f, to the measured variable X_t. It is required that f is not only a continuous function on $\mathbb{R}$ but also invertible. The definition for the model deviation is as follows,

$$D_{[s,e]} = \min_{\beta} \|f^{-1}(Y_{s:e}, \beta) - X_{s:e,\cdot}\|_{\mathcal{I}^d_{[s,e]}}. \tag{5.8}$$

where f is a monotone function on $\mathbb{R}$ and should be invertible. This deviation represents the difference between the measured value of X and the value that generated the actual Y.

This leads the a third version of proposition 4.1.

Proposition 5.2. *Let the interval [s,e] be such that it contains no change-points, then the model deviation is bounded by the norm of the true residuals on that interval, i.e.*

$$D_{[s,e]} = \min_{\beta} \|f^{-1}(Y_{s:e}, \beta) - X_{s:e,\cdot}\|_{\mathcal{I}^d_{[s,e]}} \leq \|Z_{s:e}\|_{\mathcal{I}^d_{[s,e]}}, \tag{5.9}$$

for f a monotonic function on $\mathbb{R}$.

Proof. Since the interval [s,e] does not contain a change-point, there exists a β^* such that $Y_{s:e} = f(X_{s:e,\cdot} + Z_{s:e}, \beta^*)$. Therefore,

$$\begin{aligned} D_{[s,e]} &= \min_{\beta} \|f^{-1}(Y_{s:e}, \beta) - X_{s:e,\cdot}\|_{\mathcal{I}^d_{[s:e]}} \\ &\leq \|f^{-1}(Y_{s:e}, \beta^*) - X_{s:e,\cdot}\|_{\mathcal{I}^d_{[s:e]}} \\ &= \|Z_{s:e}\|_{\mathcal{I}^d_{[s:e]}} \end{aligned}$$

Which holds because the same norm is used for the model fit and for the residuals. □

From this formulation of the proposition, the new version of theorem 4.1 follows as well. The formulation and proof are again analogous to that of theorems 4.1 and 5.1.

Theorem 5.2. *Let $S = \{S_1, \cdots, S_R\}$ be a set of intervals containing change points, returned by the NSP algorithm and f is a monotonic function on $\mathbb{R}$. Then the probability that there exists a returned interval without a change point is bounded from above, i.e.*

$$P(\exists i = 1, \cdots, R \; \forall j = 1, \cdots, N \; [\eta_j, \eta_j + 1] \not\subseteq S_i) \leq P(\|Z\|_{\mathcal{I}^d} > \lambda_\alpha) \leq P(\|Z\|_{\mathcal{I}^a} > \lambda_\alpha). \tag{5.10}$$

Proof. The set S was returned by NSP which, by line 8 of NSP, means $D_{S_i} > \lambda_\alpha$ for all $S_i \in S$. By proposition 5.2, the fact that there exists an interval not containing a change point means there is a set S_i such that $D_{S_i} \leq \|Z_{S_i}\|_{\mathcal{I}^d_{S_i}}$, which proves the first inequality. Inequality 4.6 states $\|Z_{S_i}\|_{\mathcal{I}^d_{S_i}} \leq \|Z\|_{\mathcal{I}^d} \leq \|Z\|_{\mathcal{I}^a}$, which proofs the second inequality. □

5.4 Multi-model non-linear NSP

In certain cases the underlying data model can be described with different functions f. In this case you can either choose one of these functions, for example based on which is expected to perform the best, or use them all to find an optimal function for each interval. In this case there is a set F of functions modelling the underlying data and in the model deviation step, not only will we look for the optimal parameters, but also for the optimal model. The following steps will be taken:

Step 1. For each $f \in F$, find the β^*_f that minimizes the deviation as in equation 5.5,

$$\beta^*_f = \arg\min_{\beta} \|Y_{s:e} - f(X_{s:e,\cdot}, \beta)\|_{\mathcal{I}^d_{[s:e]}}. \tag{5.11}$$

Step 2. Next, find the function $g \in F$ such that

$$g = \arg\min_{f} \|Y_{s:e} - f(X_{s:e,\cdot}, \beta^*)\|_{\mathcal{I}^d_{[s:e]}}. \tag{5.12}$$

This model deviation will be denoted as,

$$D_{[s,e]} = \min_{f} \min_{\beta} \|Y_{s:e} - f(X_{s:e,\cdot}, \beta)\|_{\mathcal{I}^d_{[s:e]}}. \tag{5.13}$$

Proposition 5.3. *Let the interval [s,e] be such that it contains no change-points, then the model deviation is bounded by the norm of the true residuals on that interval, i.e.*

$$D_{[s,e]} = \min_f \min_\beta \|Y_{s:e} - f(X_{s:e,\cdot}, \beta)\|_{\mathcal{I}^d_{[s:e]}} \leq \|Z_{s:e}\|_{\mathcal{I}^d_{[s,e]}}, \tag{5.14}$$

for f a function on $\mathbb{R}$.

Proof. Since the interval [s,e] does not contain a change-point, there exists a β^* and $g \in F$ such that $Y_{s:e} = g(X_{s:e,\cdot}, \beta^*) + Z_{s:e}$. Therefor,

$$\begin{aligned} D_{[s,e]} &= \min_f \min_\beta \|Y_{s:e} - f(X_{s:e,\cdot}, \beta)\|_{\mathcal{I}^d_{[s:e]}} \\ &\leq \min_\beta \|Y_{s:e} - g(X_{s:e,\cdot}, \beta)\|_{\mathcal{I}^d_{[s:e]}} \\ &\leq \|Y_{s:e} - g(X_{s:e,\cdot}, \beta^*)\|_{\mathcal{I}^d_{[s:e]}} \\ &= \|Z_{s:e}\|_{\mathcal{I}^d_{[s:e]}} \end{aligned}$$

Which holds because the same norm is used for the model fit and for the residuals. □

Theorem 5.3. *Let* $S = \{S_1, \cdots, S_R\}$ *be a set of intervals containing change points, returned by the NSP algorithm and f is a function on* $\mathbb{R}$. *Then the probability that there exists a returned interval without a change point is bounded from above, i.e.*

$$P(\exists i = 1, \cdots, R \; \forall j = 1, \cdots, N \; [\eta_j, \eta_j + 1] \nsubseteq S_i) \leq P(\|Z\|_{\mathcal{I}^d} > \lambda_\alpha) \leq P(\|Z\|_{\mathcal{I}^a} > \lambda_\alpha). \tag{5.15}$$

For which the proof is analogous to that of Theorem 5.2

5.5 Isotonic NSP

In certain cases where the underlying signal of the data is monotonic, it could be preferred to use isotonic regression, which is a technique where the independent variable is monotonically related to the dependent variable. This means that when the independent variable increases, the dependent variable is non-decreasing as well. As described by [Barlow and Brunk, 1972], isotonic regression can mathematically be formulated as an optimization problem of the following form,

$$\hat{x} = \min_{x^*} \sum_i w_i (x_i - x_i^*)^2 \tag{5.16}$$

$$\text{s.t. } x_i^* \leq x_j^*, \; \forall i \leq j. \tag{5.17}$$

When we want to use isotonic regression to compute the model deviation within the NSP method we need to exchange the least squares estimator in the model fit for the multi-resolution sup-norm that is used for the residual check as well. This means we obtain a non-parametric model, denoted as

$$Y_t = g(X_t) + Z_t, \tag{5.18}$$

where $g(X_t)$ is estimated as the interpolated monotone function obtained from the optimized vector $\hat{Y}$,

$$\hat{Y} = \arg\min_{Y^*} \|Y_{s:e} - Y^*_{s:e}\|_{\mathcal{I}^d_{[s,e]}} \tag{5.19}$$

$$\text{s.t. } Y_a^* \leq Y_b^*, \; \forall X_a \leq X_b. \tag{5.20}$$

Hence, instead of the parameter β changing at a change-point, now the determined $g\hat{Y}_t$ will differ. This means the model deviation will now be represented as,

$$D_{[s,e]} = \min_{g} \|Y_{s:e} - g(X_{s:e})\|_{\mathcal{I}^d_{[s,e]}}. \tag{5.21}$$

where $g \in G$, a class of isotonic functions.
Which in turn leads to a new formulation and proof of proposition 4.1.

Proposition 5.4. *Let the interval [s,e] be such that it contains no change-points and M a class of monotone functions, then the model deviation is bounded by the norm of the true residuals on that interval, i.e.*

$$D_{[s,e]} = \min_{g} \|Y_{s:e} - g(X_{s:e})\|_{\mathcal{I}^d_{[s:e]}} \leq \|Z_{s:e}\|_{\mathcal{I}^d_{[s,e]}}, \tag{5.22}$$

for $g \in M$.

Proof. Since the interval [s,e] does not contain a change-point, there exists an $g^*(X_{s:e})$ such that $Y_{s:e} = g^*(X_{s:e}) + Z_{s:e}$. Therefor,

$$\begin{aligned} D_{[s,e]} &= \min_{\hat{Y}} \|Y_{s:e} - g(X_{s:e})\|_{\mathcal{I}^d_{[s:e]}} \\ &\leq \|Y_{s:e} - g^*(X_{s:e})\|_{\mathcal{I}^d_{[s:e]}} \\ &= \|Z_{s:e}\|_{\mathcal{I}^d_{[s:e]}} \end{aligned}$$

Which holds because the same norm is used for the model fit and for the residuals. □

Theorem 5.4. *Let $S = \{S_1, \cdots, S_R\}$ be a set of intervals containing change points, returned by the NSP algorithm. Then the probability that there exists a returned interval without a change point is bounded above, i.e.*

$$P(\exists i = 1, \cdots, R \; \forall j = 1, \cdots, N \; [\eta_j, \eta_j + 1] \nsubseteq S_i) \leq P(\|Z\|_{\mathcal{I}^d} > \lambda_\alpha) \leq P(\|Z\|_{\mathcal{I}^a} > \lambda_\alpha). \tag{5.23}$$

Proof. The set S was returned by NSP which, by line 8 of NSP, means $D_{S_i} > \lambda_\alpha$ for all i. By proposition 5.1, the fact that there exists an interval not containing a change point means there is an i such that $D_{S_i} \leq \|Z_{S_i}\|_{\mathcal{I}^d_{S_i}}$, which proves the first inequality. Inequality 4.6 states $\|Z_{S_i}\|_{\mathcal{I}^d_{S_i}} \leq \|Z\|_{\mathcal{I}^d} \leq \|Z\|_{\mathcal{I}^a}$, which proofs the second inequality. □

5.6 S-shaped non-parametric NSP

In specific cases the relation between a dependent variable and covariate is S-shaped. This means there exists an inflection point in the data, where the regression function flips from convex to concave. When in these instances we want to use non-parametric estimation of the regression function, it can be preferred to use S-shaped non-parametric regression instead of isotonic regression. An approach to doing so is described by [Feng et al., 2021], using a least squares approach. It is tuning-free, which means it does not require the selection of additional tuning parameters. First the approach will be described, followed by the adaptation and application for NSP.

5.6.1 Computation of S-shaped least squares estimators

The computation is done for data points (x_i, y_i) for $i = 1, \cdots, n$, with x_i the independent input variable and y_i the dependent output variable. We call $f(x)$ the S-shaped functions used to fit the data, which as one inflection point m where the functions curve changes from convex to concave. The objective is to find the function $\hat{f}$ that minimizes the residual sum of squares,

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} (y_i - f(x_i))^2, \tag{5.24}$$

where $\mathcal{F}$ is the class of all S-shaped functions with an inflection point m.

Initialization. The computation starts with a given data set (x_i, y_i) for $i = 1, \cdots, n$. Consider each data point $x_j, j = 2, \cdots, n-1$ as a potential inflection point or define another subset that should be considered.

Splitting data and fitting segments. For each potential inflection point $m = x_j$, split the data into two segments. For the left segment i ranges from 1 to j and for the right from $j+1$ to n. On the left segment a least squares approach will be used to fit a convex increasing function f_l and on the right a concave increasing function f_r. This means a function is fitted with the constraints

$$f_l"(x) \geq 0, \text{ for } x \leq m, \tag{5.25}$$

$$f_r"(x) \leq 0, \text{ for } x \geq m. \tag{5.26}$$

These left and right functions should be combined in the inflection point to obtain a continuous S-shaped function. For the combined fit the residual sum of squares (RSS) will be calculated.

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{5.27}$$

Optimize. The above step will be repeated for each potential inflection point and the one minimizing the RSS gives the best fit for the data and will be selected. This is denoted as,

$$\hat{m} = \arg\min_{m} RSS(m) \tag{5.28}$$

to find the inflection point and the $\hat{f}$ computed before that accompanies it. In order to improve computational efficiency, a primal-dual algorithm is used. This way the least squares fit updates when new data points are added, instead of completely computing it again each time.

5.6.2 S-shaped adaptation

In order to use this method when measuring the model deviation in NSP, two adaptations will need to be made. The multi-resolution sup-norm will replace both the least squares, for fitting the convex and concave increasing functions, and the RSS for the combined fit. This way the method is very similar to the isotonic version. However, the constraints in the optimization step are different and since in most cases we do not know the inflection point beforehand, we have to optimize the residual step to find it. The model is formulated as,

$$Y_t = h(X_t) + Z_t \tag{5.29}$$

The deviation is denoted as,

$$D_{[s,e]} = \min_{h} \|Y_{s:e} - h(X_{s:e,\cdot})\|_{\mathcal{I}^d_{[s:e]}}, \tag{5.30}$$

where h is the S-shaped function. This again is a non-parametric model and thus the structure is analogous to that of isotonic regression but with an extra constraint in the optimization step of the function.

Proposition 5.5. *Let the interval [s,e] be such that it contains no change-points, then the model deviation is bounded by the norm of the true residuals on that interval, i.e.*

$$D_{[s,e]} = \min_{h} \|Y_{s:e} - h(X_{s:e})\|_{\mathcal{I}^d_{[s:e]}} \leq \|Z_{s:e}\|_{\mathcal{I}^d_{[s,e]}}, \tag{5.31}$$

for h an S-shaped function.

Proof. Since the interval [s,e] does not contain a change-point, there exists an h_t^* such that $Y_{s:e} = h^*(X_{s:e}) + Z_{s:e}$. Therefor,

$$\begin{aligned} D_{[s,e]} &= \min_h \|Y_{s:e} - h(X_{s:e})\|_{\mathcal{I}^d_{[s:e]}} \\ &\leq \|Y_{s:e} - h^*(X_{s:e})\|_{\mathcal{I}^d_{[s:e]}} \\ &= \|Z_{s:e}\|_{\mathcal{I}^d_{[s:e]}} \end{aligned}$$

Which holds because the same norm is used for the model fit and for the residuals. □

The theorem and proof for the upper bound on the Type I error is in line with Theorem 5.4.

6 | Application of NSP to wind energy

In this section it will be explained how the NSP method and it's adaptations can be applied to the change points in wind energy generation. First, we will describe the available data for a single wind turbine, and how each aspect relates to the variables of the NSP method. After this, it will be detailed how the original NSP method can be applied to the data and how each of the adaptations compare to each other.

6.1 Data description

To apply NSP to produced wind power, we describe how to match the available data to the mathematical model. One model will be used to explain the application. Recall the NSP framework as described in equation 5.1 in Chapter 5,

$$Y_t = f(X_t, \beta^{(j)}) + Z_t.$$

In this scenario we have four variables which describe the production of power. Firstly, Y_t denotes the power. Next, X_t represents the wind speeds and the parameter vector β determines the power output of each wind speed. Lastly, Z_t subsumes the model error and potential measurement errors. For now, we assume Z_t to be identically symmetrically distributed, but not necessarily independent. For the wind speed X_t there are multiple options to consider to use. The data used in this thesis and also provided along are from an onshore wind turbine in the Netherlands with a height of 112.0 meters and a maximum capacity of 3600.0kW.

Generated power:
As stated above, Y_t represents the generated power of a wind turbine. It is a data frame where the index t corresponds to time in multiples of 15 minutes and the measurements are in kilowatt hour (kWh) of generated electricity.

Wind speed:
The design matrix X_t is constructed using the wind speeds recorded at each time index. There are three main sources of wind speed data to consider:

1. Wind speeds measured directly at the wind turbine.
2. Wind speeds recorded at various weather stations throughout the country.
3. Forecasted wind speeds provided on a grid covering the Netherlands.

Ideally, we use the wind speeds measured directly at the wind turbine, as this data offers the most accurate representation of the local conditions. In this research these actual wind speeds were available and thus those are used. However, such measurements are often unavailable for turbines, making it necessary to explore the use of alternative data sources.

When this turbine data is unavailable it leaves a choice between weather station measurements and forecasted wind speeds. The key advantage of weather station data is that it represents

actual, measured conditions, making it highly reliable for the specific location of the station. However, the Netherlands has 48 weather stations of which 34 on land and 14 at sea, as noted by [KNMI, 2024a]. Consequently, there isn't always a weather station near a given wind turbine, which can reduce the accuracy of the data for that location.

In contrast, forecasted wind speed data is available for a dense grid of locations across the Netherlands, with a resolution of 2.5 km, according to [KNMI, 2024b]. The main drawback is that forecasted data, by nature, is less reliable than direct measurements. However, these forecasts are updated hourly, which allows for the timely incorporation of sudden weather changes.

Given the higher spatial resolution of the forecast data and its frequent updates, we have decided to use the forecasted wind speeds in cases where data from the wind turbine itself is unavailable. This approach provides a good balance between coverage and reliability.

For future application, forecasted wind speeds at a height of 100 meters are recommended, when turbine measurements are unavailable, as this height most closely matches the actual height of most wind turbines. For turbines with significantly different heights, adjustments can be made by using forecast data from a height that is closer to the turbine's actual height.

6.2 Model comparison for wind energy

In this section each of the models applied to wind energy will be described and compared to each other.

6.2.1 Non-linear regression with output error

The first extension of NSP is a non-linear model with a model error, described as follows,

$$Y_t = f(X_t, \beta^{(j)}) + Z_t. \tag{6.1}$$

In this scenario, instead of a linear model between the wind speed and the power generation, a non-linear relation is imposed. This framework was introduced in Section 5.2. Since the true power function is not a linear function, this is a more realistic scenario for this application. The function f will thus resemble the power function, for which one of the variations described in Chapter 3 can be used. The parameter vector β can be of different sizes, based on the amount of parameters in the chosen power function, and will be optimized to find the best function fit on the true values Y_t. Then the error Z_t represents the measurement error. The power function is an S-shaped curve consisting of four segments, each with a different error distribution. The exact distributions are uncertain but assumed to have mean 0.

This approach however can easily turn into a model misspecification problem. Model misspecification occurs when the statistical model used does not accurately represent the underlying data process. In the context of wind energy it can happen in several ways. Firstly, the distribution can be misspecified. With careful analysis of different power functions this problem will be minimized. Nevertheless, using different power functions may give different results. Secondly, a misestimation of the noise can affect the threshold used for detecting change points, leading to false positives or negatives. This is why a careful analysis of the residuals should be done, in order to minimize the false outcomes. Lastly, since this method relies on parameters, the selection of these should be done carefully to avoid a bad fitted power curve.

In the case of a multi-model non-linear scenario, we are considering the scenario described above, but now with a set F of descriptions of the power curve. In this variation, in addition the parameter vector β, the power function f will also be optimally fitted. This way the model misspecification problem in the distribution is attempted to be minimized.

6.2.2 Non-linear regression with input error

The second non-linear scenario differs from the first one in the error type. Instead of a model error we are now considering an error in the input variable X_t which resembles the wind speed.

$$Y_t = f(X_t + Z_t, \beta^{(j)}) \tag{6.2}$$

There are multiple aspects that influence the error in the weather component and that should be considered when selecting the data for this variable. The input can either be a forecast or an actual measurement, of which the actual measurement will be more accurate for the location of the measurement. However, since there are not many measuring stations, the wind turbine will likely be far away and thus a large error may be present. The forecasts on the other hand are done for a large grid, which means the location of the forecasts is closer to an asset and multiple forecasts can be combined for an optimal value. The downside of this is that there is uncertainty in the forecast, which also results in an error. These forecasts are done for multiple heights, so depending on how close the wind turbine height is to the forecast height, there is some error here as well. For both possibilities of wind speed input, the measurement and forecasts, the error is assumed to be mean 0. A disadvantage of this approach is that since the noise is part of the input of the function f, it might be harder to estimate the parameters well.

6.2.3 Isotonic regression

Isotonic regression is a non-parametric regression form. Whereas in the scenarios above we had an input variable and a parameter vector to optimize, we now look for an optimal monotone function to fit the actual generated power data. This model has a model error and is denoted by,

$$Y_t = g(X_t) + Z_t, \tag{6.3}$$

where g is an isotonic function. This framework looks similar to the non-linear scenario with a model error. However, fitting the optimal function $g(X_t)$ is not restricted to a power curve this time, hence also the error is distributed differently.

The isotonic regression fit is a piece-wise constant function. When the fit is done with the leas-squares solution, the residuals within each constant segment will vary around the mean of that segment, however with the multi-resolution that is not the case. In the application to generated power, the data follows a monotonic trend with some variation, until the cutout wind speed is reached. In that last segment of the power curve, monotonicity is not valid and thus the residuals might be very large here. Elsewhere they will be mean 0 and clustered around the isotonic fit. This last segment, starting at the cutout wind speed, can pose a problem for the isotonic fit if it is reached often. In these cases a different approach might be needed to provide a good performance. An example to do so is to use isotonic regression up to the cutout speed, and a constant zero function for the wind speeds beyond that point.

Although the error has a similar meaning to one of the non-linear models, there is an important difference between isotonic regression and the non-linear models. Unlike the models described above, isotonic regression does not make use of parameters to relate the wind speed to generated power. This makes it a non-parametric regression model without any specific form for the underlying relationship, which makes is flexible in modelling monotonic relationships. This piece-wise constant nature of the estimate is relevant to the generated power, since that is also the shape of the original deterministic relation between wind speed and power, provided by the manufacturer. However, in reality the underlying relationship is much smoother.

6.2.4 S-shaped non-parametric

The S-shaped non-parametric regression is similar to the isotonic version but with a shape constraint. This S-shape constraint resembles the power curve. In this approach convex and concave

constraints are imposed in the fitting process, which makes sure to maintain the S-shape.

In [Feng et al., 2021] it is described how in some examples this non-parametric estimator showed better performance than parametric methods such as segmented linear regression and logistic regression. They do however state that when enough domain specific knowledge is available to determine the parameters, the superior performance might not hold. For the application to wind turbines this domain knowledge will be present in most assets, thus the logistic regression might perform better in these cases. Another advantage of this method, as opposed to the parametric methods described before, is that it is robust to model misspecification within the cubic part of power curve, for which Feng has provided results.

6.3 Statistical metrics

In order to assess and compare the performance of each method on the simulations, they will be evaluated on 5 aspects. We use the possible test outcomes as shown in the confusion matrix in Figure 6.1 to describe the metrics.

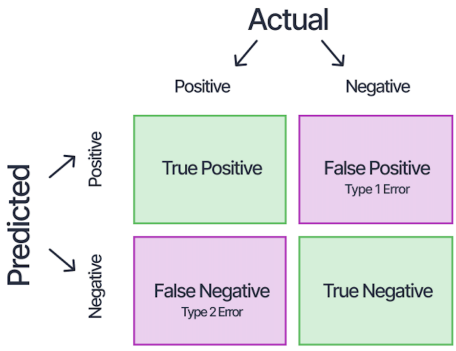


Figure 6.1: Confusion matrix of test results.

The first is empirical coverage, whether at least $(1-\alpha)100\%$ of the simulations are such that any of the returned significant intervals contain at least one true change point.

There are many other metrics to track model performance, [Luque et al., 2019]. The second we will compute is the *true positive value*, also called the *sensitivity* (SNS). It is the fraction of known change points that is detected and calculated as,

$$SNS = \frac{TP}{TP + FN}. \tag{6.4}$$

Next, the *positive predictive value*, or *precision* (PRC), is the fraction of positive results that is indeed an actual change point. The fraction is denoted as,

$$PRC = \frac{TP}{TP + FP}. \tag{6.5}$$

Since the NSP method gives confidence intervals around a detected change point as output, we also want to track the length of these intervals. In this case the smaller the intervals are the better the performance. This metric will be the average length of genuine intervals.

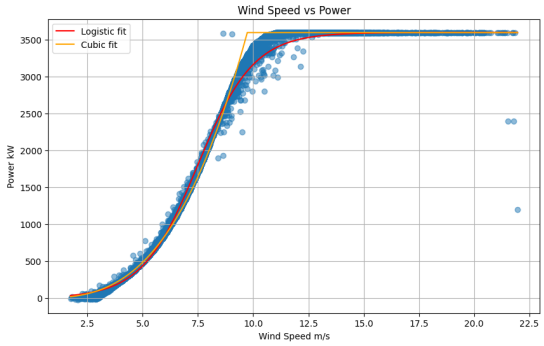
Lastly, the runtime of the models will be tracked. In case different models perform equally, the one with a lower runtime is preferred.

6.4 Simulations and thresholds

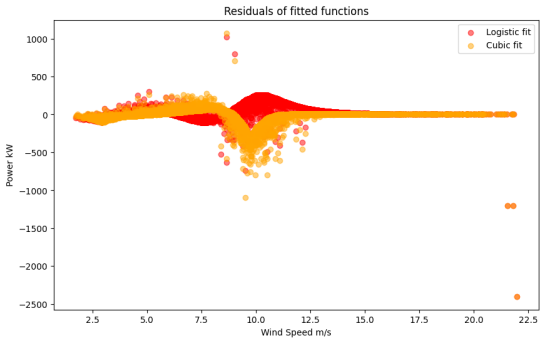
In order to test the performance of each version of NSP, we made use of simulations of power output based on actual wind speeds and with manually added change points. First, the construction of the simulations will be explained, followed by the calculation of the thresholds. The simulations are based on the data in *clean_power_data*, which includes for each timestamp the wind speed and power.

6.4.1 Output error simulations

In this scenario we simulate data with the error in the power output. In this case it is assumed that the input variable, the wind speed, is accurate and the error lies solely in the power output. To determine this error we use the actual wind speeds and power output of a period without anomalies, where the full range of wind speeds and power output has been reached. To this data we fitted a cubic and logistic function from which for both the residuals could be calculated. A scatter plot of these values and fitted functions can be seen in Figure 6.2.



(a) Scatter plot of wind speed and power output with cubic and logistic fitted functions.



(b) Scatter plot of wind speed and power output with cubic and logistic fitted functions.

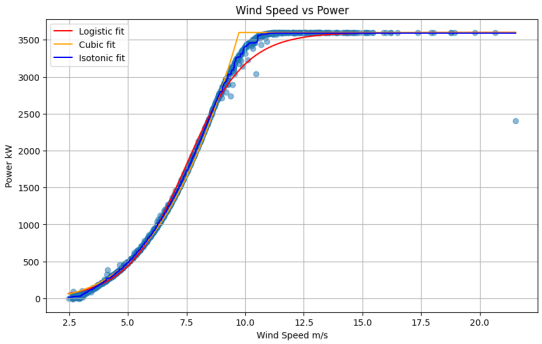
Figure 6.2: Function fits and residuals of the cubic and logistic functions.

For both functions the fit can be seen in Figure 6.2a and their residuals are plotted in Figure 6.2b. The total sum of residuals for both fits are given in Table 6.1.

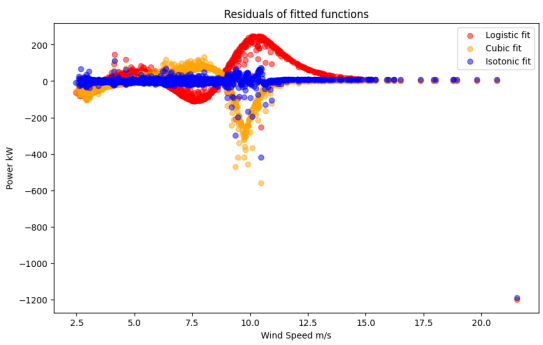
Fitted function	Total residuals	Mean	Std. Dev.
Cubic fit	-253818.06	-8.55	71.44
Logistic fit	97793.22	3.29	67.25

Table 6.1: Total residuals, mean and standard deviation of the fitted functions in kW.

In addition to the cubic and logistic functions we are also interested in the isotonic fit. As shown by [Mehrjoo et al., 2020], non-parametric monotone power curve fitting can provide good results. However, due to the non-parametric nature of the isotonic fit it required computationally too much time to fit on the entire set of data. To be able to compare the isotonic fit to the two parametric fits it was decided to use a largest possible subset of data. This fit in comparison to the cubic and logistic fit on the subset is visible below.



(a) Cubic, logistic and isotonic fitted functions on subset of data.



(b) Residuals of cubic, logistic and isotonic fitted functions on subset of data.

Figure 6.3: Function fits and residuals of the cubic, logistic and isotonic functions on data subset.

In Figure 6.3a we see that taking a subset of data has caused for a more perfect power curve and the noise is lost. The residuals of the cubic and logistic fit are similar in distribution to those in Figure 6.2b. The isotonic fit has smaller residuals with a more normal distribution, except for the wind speeds around the change to the rated power, where the fit overestimates a bit. However, since a large portion of the noise is lost we prioritize the function fits of the full data set to give a more realistic view.

From Table 6.1 and the fitted functions we see the cubic fit overestimates, whereas the logistic fit underestimates. If we only consider the total absolute residuals we can conclude the logistic function had the best fit with the lowest total amount, hence the logistic function will be used for the simulations.

For the residuals of the logistic fit we plotted again the scatter plot and added a histogram and QQ-plot to analyze the distribution, these plots are visible below.

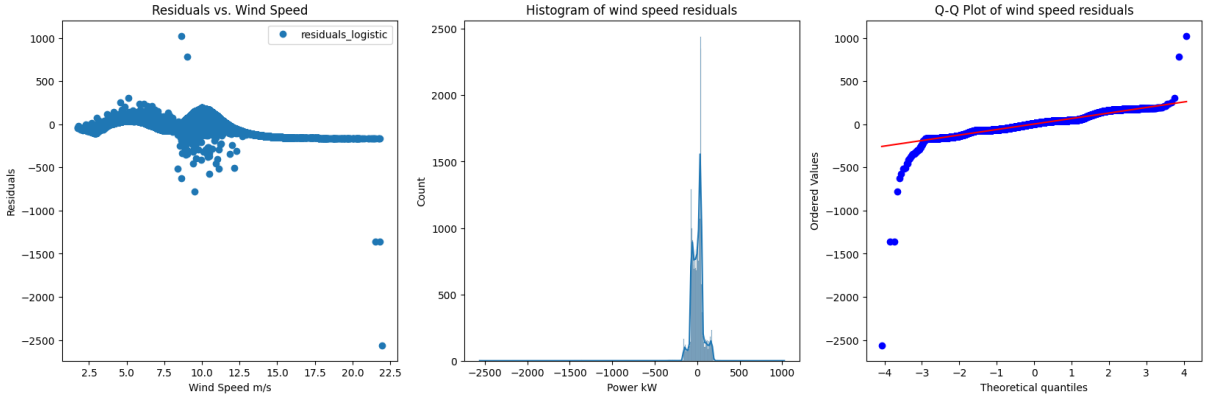


Figure 6.4: Scatter plot, histogram and QQ-plot of the residual of the logistic function.

From the scatter plot, histogram and QQ-plot we can clearly see the distribution is not completely normal and especially has a few outliers, thus large tails. However, for our purpose of simulating data, Gaussianity will be assumed. For the application on wind energy we know the errors might not be independently distributed. Intuitively, we know it is unrealistic that within an hour with the similar wind speeds, we would have great differences in power output. Thus a dependence in the error could be expected. To determine if the errors are indeed dependent and an autoregressive (AR) model should be used, we compute the autocorrelation of the error series. In the figure below, the autocorrelation function (ACF) and partial autocorellation function (PACF) are shown.

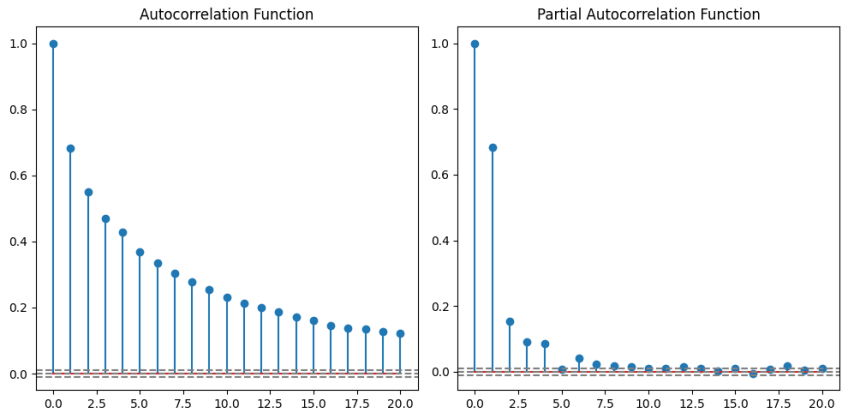


Figure 6.5: The ACF and PACF for the errors of the logistic function fit.

In Figure 6.5 we see the ACF is decreasing slowly and the PACF has a spike at lags 1, 2, 3 and 4. This indicates it is an AR model of order 4, AR(4). Using the AutoReg function of statsmodels in python we obtain the following coefficients for the lags and constant:

Variable	**Coefficient (coef)**
constant	0.7297
residuals logistic.L1	0.5554
residuals logistic.L2	0.0930
residuals logistic.L3	0.0433
residuals logistic.L4	0.0862

Table 6.2: Regression Coefficients

The basis of the simulations is made with real wind speed data with a length of 3000. This was split into 30 simulation sets of 100 data points, which combined represents one month of data. The logistic function, with the determined parameters by the model fit, was used to calculate the generated power of each point in the wind speed dataset. After this, the noise, based on a AR(4) model, is added to the power output of each of the 30 simulations. Thus, at this point there are 30 simulations, based on a different set of wind speeds and with the power output based on the logistic power curve fit with output noise.

To these simulations with normal power output, change points will be added. A total of 5 anomalous periods is generated, each with different lengths. Since these anomalous periods each have a start and end point, this results in 10 change points randomly divided over the simulations. The change in these periods lies in the power capacity that remains. For example, a 20% capacity means if normally the turbine could generate 3500kW at a certain wind speed, it will now only generate 700kW. To efficiently test the sensitivity of the models 5 different capacity changes, 0.0%, 20%, 40%, 60% and 80%, will be applied on the same periods. Hence, these the 30 simulations will be made in five fold to add the different capacities.

In the figure below we see the five fold of simulations with different capacity changes in one plot. The plot adjoins all 30 simulations of 100 points, which results in approximately one month of simulations in total.

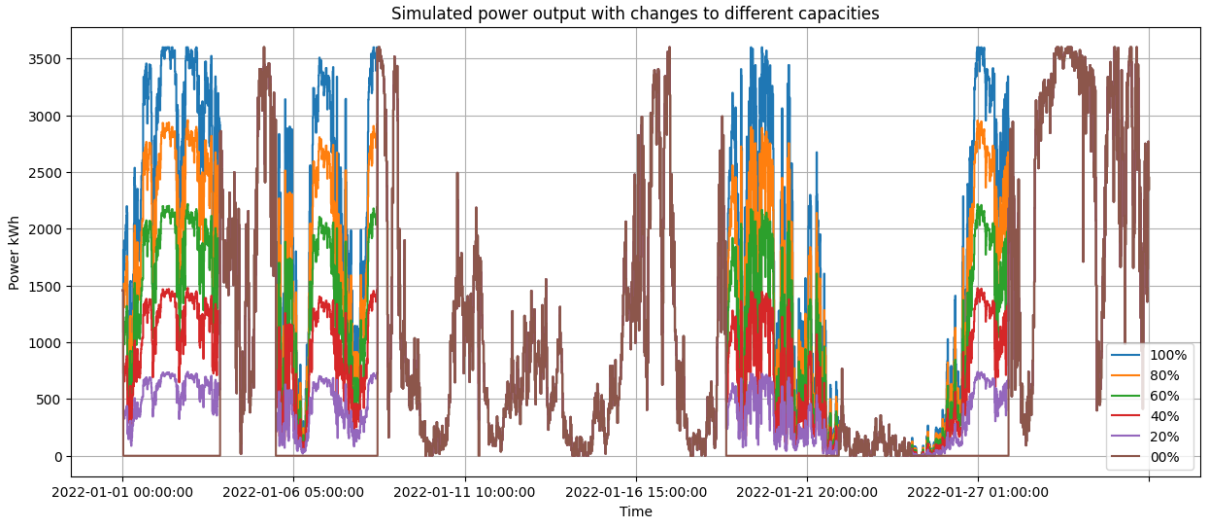


Figure 6.6: Five fold adjoined simulations with different capacity changes.

6.4.2 Threshold

For each model, a general threshold is determined that will be used across all simulations, ensuring that the threshold is independent of any specific simulation instance. This threshold is based on the model residuals, as detailed earlier in Chapter 5. The residuals are derived by fitting the function of each model, two non-linear models with output error, two with input error, and one isotonic regression model, on the power curve of our data and calculating the difference between the model's predictions and the observed data.

The process for determining the thresholds of each model involves the following steps:

1. **Fitting AR(4) model to residuals:** For each model, the residuals are extracted by fitting the model to the power curve. An AR(4) model is then fitted to these residuals to capture the temporal dependencies within the noise. The AR(4) model, which assumes that each residual is influenced by the previous four residuals, provides the necessary autoregressive coefficients to simulate new errors.

2. **Simulating residuals:** Once the AR(4) coefficients are obtained, the residuals are simulated with a Gaussian noise term. This method generates new residual samples that preserve the AR(4) autocorrelation structure, maintaining the time-dependent error pattern modeled by the AR process.

3. **Simulating new time series:** For each simulated series of residuals, the test statistic $\|Z\|_{\mathcal{I}^d}$ is computed, which measures deviations within the structure of the model's error over time.

4. **Determining the threshold λ_α:** From the test statistics of the simulated samples, the threshold λ_α is chosen as the empirical $100(1-\alpha)\%$ quantile of the distribution of the test statistics. This threshold represents the critical value beyond which deviations are considered significant for the given model.

The thresholds determined for each model are presented in Table 6.3.

	Log. Out.	Cub. Out.	Log. In.	Cub. Out	Isotonic
0.05	547.82	831.59	10.26	6.16	125.41

Table 6.3: Threshold determined for each model for alpha=0.05.

These thresholds are determined for simulated data with Gaussian noise. However, in application to data where the noise distribution is unknown it is also possible to use bootstrapping on the residuals to determine the threshold for a specific turbine. In this case the residuals are randomly sampled with replacement. For these bootstrapped samples the statistic can be determined to find the threshold.

7 | Results

This chapter presents the results for each model, evaluated using the simulations and the metrics outlined in Section 6.3. First, a reproduction of one of the datasets from the original NSP method is discussed. Following this, the results for both non-linear models, incorporating output and input errors for logistic and cubic fits, are presented. Finally, the outcomes of the isotonic model are analyzed. All results are based on an alpha value of 0.05, as described in Section 4.3.

7.1 Reproduction of the Fryzlewicz results

In order to compare the performance of the implemented NSP algorithm with the original implementation in R-studio, we reproduced the results from the data example from [Fryzlewicz, 2023] on the US ex-post real interest rate[1]. Similar to the article approach we estimate σ^2 via $\hat{\sigma}^2_{MAD}$ and run with the same parameters: M=1000, $\alpha = 0.1$, $\tau_L = \tau_R = 0$. The value of $\hat{\sigma}^2_{MAD}$ is calculated to be 6.69. With this value and the other parameters, the implemented NSP algorithm with the original framework was run. The following two significant intervals were returned: $[35, 69], [75, 82]$. In the article the set of significant intervals is $S_0 = [24, 55], [76, 83]$ and the real change points are estimated to be $\tilde{\eta}_1 = 47$ and $\tilde{\eta}_2 = 82$. This means the intervals returned by our model do both contain a change point. A visualization is given in Figure 7.1.

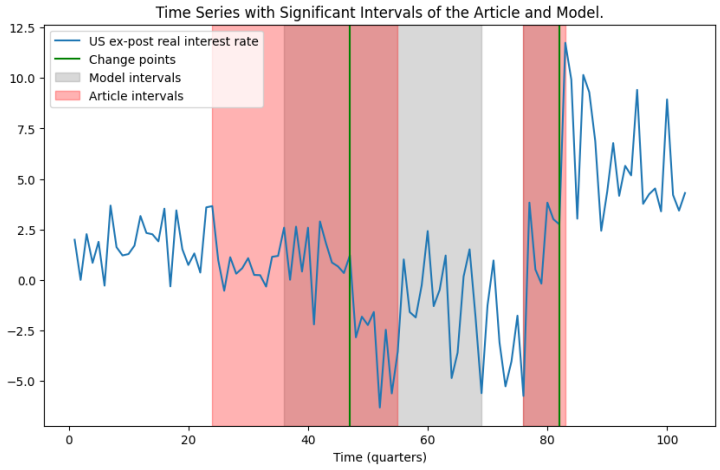


Figure 7.1: Significant intervals from the article and our model for the US ex-post real interest rate.

The reason for the difference in significant intervals of our model and the article is the random interval sampling. The results in the article are obtained from a random sample without a seed and it is therefore not possible to duplicate them.

[1]Dataset: http://qed.econ.queensu.ca/jae/datasets/bai001/

7.2 Global analysis of results

First, an overview of the results for each model will be given in Table 7.1. In this table the results are provided for each model and capacity simulation. The capacity in this case means the remaining capacity in an anomalous period. Each model and simulation combination was assessed on 5 metrics. The coverage is the percentage, out of 30 simulations, where no false intervals were returned. The fraction detected change points denotes the fraction, out of 10 points that was detected, hence 1.0 would be the perfect result. The fraction true change points within the returned intervals is the fraction of intervals that indeed contain a change, hence again 1.0 is the optimal result. Lastly are the average length of genuine intervals and the runtime. The last row provides the average results of all capacities per model.

Capacity	Metric	Log. Out	Cub. Out.	Log. In.	Cub. In.	Isotonic
0.0%	coverage	100.0	100.0	100.0	96.67	100.0
	frac. det. cp.	0.7	0.6	0.4	0.8	0.8
	frac. true. pos.	1.0	1.0	1.0	0.9	1.0
	av. int. length	5.57	2.5	47.2	18.78	2.63
	runtime	10:42	4:47	8:16	2:20	37:37
20%	coverage	100.0	100.0	100.0	93.33	100.0
	frac. det. cp.	0.6	0.6	0.0	0.3	0.8
	frac. true. pos.	1.0	1.0	1.0	0.6	1.0
	av. int. length	8.17	7.33	0	35.33	2.63
	runtime	13:30	6:01	11:52	4:03	37:03
40%	coverage	100.0	96.67	100.0	93.33	100.0
	frac. det. cp.	0.5	0.5	0.0	0.1	0.8
	frac. true. pos.	1.0	0.83	1.0	0.33	1.0
	av. int. length	8.4	15.4	0	31.0	2.75
	runtime	9:23	5:09	10:47	3:06	19:11
60%	coverage	100.0	100.0	100.0	93.33	100.0
	frac. det. cp.	0.3	0.1	0.0	0.0	0.8
	frac. true. pos.	1.0	1.0	1.0	0.0	1.0
	av. int. length	20.33	5.0	0	0	3.9
	runtime	8:00	3:08	9:50	2:56	19:17
80%	coverage	100.0	100.0	100.0	93.33	100.0
	frac. det. cp.	0.1	0.0	0.0	0.0	0.2
	frac. true. pos.	1.0	1.0	1.0	0.0	1.0
	av. int. length	27.0	0	0	0	4.0
	runtime	7:08	3:04	10:54	2:56	18:22
Average	coverage	100.0	99.33	100.0	93.99	100.0
	frac. det. cp.	0.44	0.36	0.08	0.24	0.68
	frac. true. pos.	1.0	0.97	1.0	0.36	1.0
	av. int. length	9.91	7.83	47.2	23.94	3.03
	runtime	9:44	4:25	10:19	3:04	26:18

Table 7.1: Results of the 5 metrics on all capacity + model combinations and their average result.

From these general results we know the isotonic model on average performs best. With a coverage of 100.0 and a fraction true positives of 1.0, the results it provides are very reliable. In addition to this the average fraction of detected points is the highest of all models with a value of 0.68 and the length of the significant intervals are the lowest of the models. The only downside is the long run time of almost half an hour on average. Following the isotonic model, both the methods with output error perform good as well. In comparison to the isotonic model they have especially a lower detection rate and return longer intervals of significance. However, their runtime is lower. The methods with input error have the least promising results, with the one based on the cubic model with an under performing coverage, which could be due to a numerical error. The logistic fit does have a high coverage which is largely due to the low amount of returned intervals. When comparing the logistic and cubic model in general we see the logistic has a better coverage but does take more than twice as long. Besides this, the models perform similar on the amount of detected changes and the average interval length of the cubic function is lower.

The simulations were made based on the logistic function with output error. This could have caused for the logistic model with output error to perform slightly better then the cubic one. In addition to this, it also explains the on average worse performance of the models with input error. But in spite of that the cubic input model did perform well on the 0% simulation, hence the worse results can not fully be accounted to the construction of the simulations. The problem here is that the inverse functions have trouble finding a good fit on two horizontal parts of the power curve, hence the moment before the cubic increase starts and the rated power. This means the inverse fit is not ideal for this application.

Next, we look at each of the change points separately, numbered from 1 to 10. An indication is given in the Figure 7.2, where the 20% simulation is visible with the numbered changes. In this figure all 30 simulations of length 100 are adjoined to be easily displayed in one figure. The fifth and sixth points lie so closely together that their labels are given on either side of the changes.

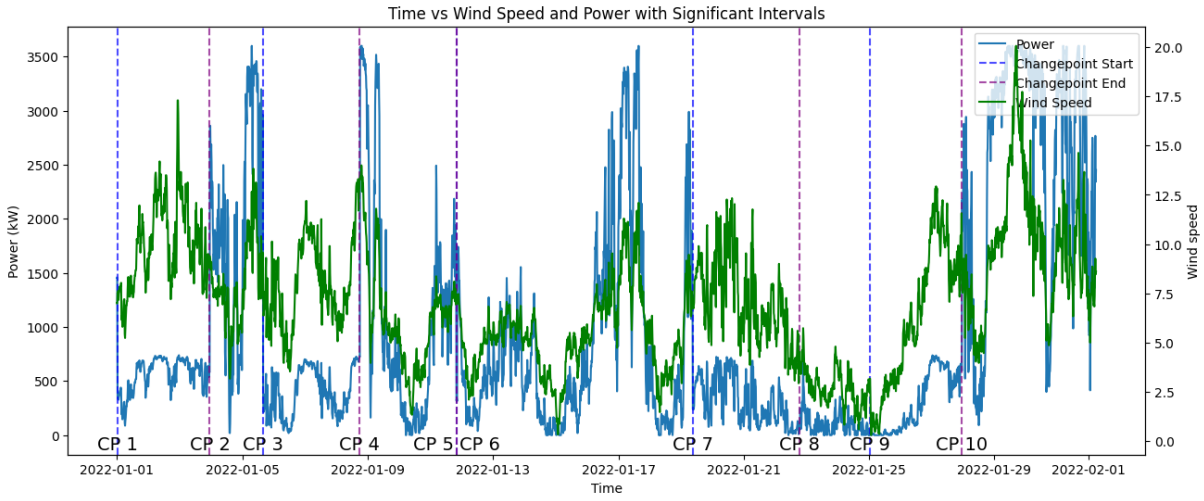


Figure 7.2: Numbered change points in the 20% capacity simulation.

In Table 7.2 the results per model and change are given. A few remarks to be made immediately are that points 1, 2, 3, 4, 7 and 10 are often detected, with the detection count varying over most models. Changes 5 and 6 are almost exclusively detected by the isotonic model, and for that model they were detected in all but one simulations. These were the changes that followed each other closely. It shows the isotonic model is almost exclusively the model that is ably to detect these short changes. In contrast to this, changes 8 and 9 are only detected by the inverse models, each only in one simulation, and are the only points the isotonic model has missed.

Change Point	Metric	Log. Out	Cub. Out.	Log. In.	Cub. In.	Isotonic	Average
1	det. count	3	2	1	1	4	11
	av. length	4.33	5.0	34.0	23.0	1.25	7.73
2	det. count	4	3	0	1	5	13
	av. length	13.0	5.0	0	15.0	1.0	6.69
3	det. count	2	3	0	2	4	11
	av. length	20.5	10.0	0	32.5	3.0	13.45
4	det. count	5	4	0	3	5	17
	av. length	7.8	2.25	0	16.0	7.0	7.71
5	det. count	1	0	0	0	4	5
	av. length	3.0	0	0	0	6.25	5.60
6	det. count	0	0	0	0	4	4
	av. length	0	0	0	0	6.25	6.25
7	det. count	3	3	1	2	4	13
	av. length	14.33	15.67	91	57.5	4.75	24.23
8	det. count	0	0	1	1	0	2
	av. length	0	0	44	25.0	0	34.50
9	det. count	0	0	1	1	0	2
	av. length	0	0	58	8.0	0	33.00
10	det. count	4	3	0	1	4	12
	av. length	6.75	10.0	0	2.0	1.5	5.42

Table 7.2: Number of times a change point was detected, out of the 5 capacity runs, and average length of these detected intervals.

In the following sections we analyze the models separately in more detail.

7.3 Logistic model with output error

For the logistic model with output error, the method was carried out with a threshold of 547.82. From Table 7.1 we see the coverage of the model is 100.0 for all the simulations, resulting in a fraction of true positives of 1.0 as well. This means there are no false positives returned. The fraction of detected change points, or sensitivity, is on average 0.44. We can see the fraction is 0.7 for the 0.0% simulation and then decreases to 0.1 for the 80% run. It is clear that for the model is more sensitive for larger changes. This is also visible in the average length of the genuine intervals detected, this value increases with the capacity that is left. It shows the model needs more data to determine a change when the change is smaller. In context of power production in wind energy the interval length translates to PTUs, where 1 PTU is 15 minutes. Since this model estimates 4 parameters it also needs a minimum of 4 data points to estimate those, which means the minimum interval length would be 3. This value is not reached on average but it is returned multiple times. In the histogram in Figure 7.3 we see that indeed most intervals have a length of 3 but the average is higher due to some outliers.

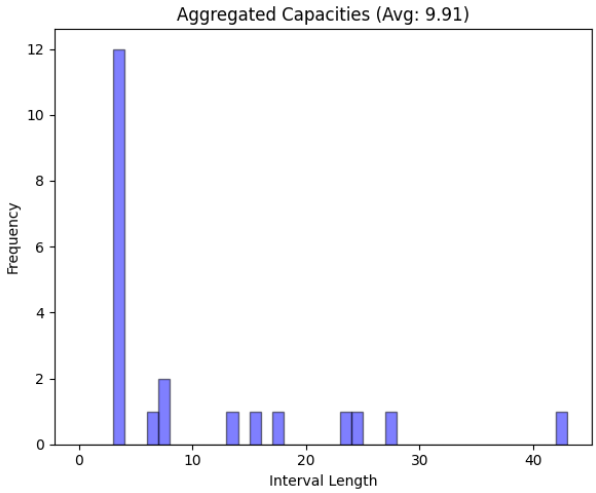


Figure 7.3: Lengths of the genuine intervals returned by the logistic model with output error, aggregation of all capacity simulations.

In Figure A.1 in Appendix A, the histograms per capacity simulation are given. After careful analysis of the returned intervals we see that the largest intervals returned by the 0 and 20 percentage runs are not returned in the other simulations. This suggests the model had difficulty detecting these points when the change was large but when it got smaller they were missed completely.

7.4 Cubic model with output error

This model was run with a threshold of 831.59. In Table 7.1 we see all simulations ran with a coverage of 100 percent, except for the 40% run, resulting in an overall coverage of 99.33. On average the model returns 36% of the known change points. This value is influenced downwards by the two models with the least change, they return only 10% or none of the changes. On the contrary do the other three models with a larger change 50% or 60%. Hence, the model performs much better for capacity changes below 60%.

The difference between those three simulations lies mostly in the length of the returned intervals. This increases significantly, starting with an average length of 2.5 for the change to 0 percent up to 14.4 for the change to 40 percent. On average over all the simulations this leads to a length of 7.83. In the figure below a histogram of the lengths of the genuine intervals is given. We see the minimum detected length is 1, which is in line with the theoretic minimum. Since the cubic model needs to fit 2 parameters, it only needs two points to estimate them and thus a length of 1 PTU is the minimum.

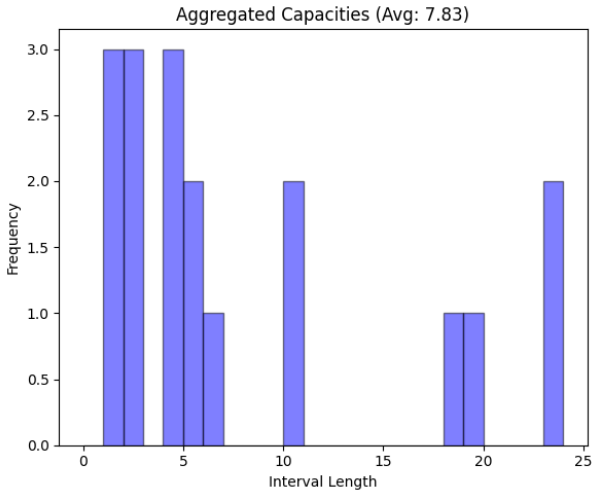


Figure 7.4: Lengths of the genuine intervals returned by the cubic model with output error, aggregation of all capacity simulations.

The histogram shows most intervals are of lengths 1 to 6 but there are some outliers, all of these are returned by the 40% and 20%, which is in line with the average returned lengths.

Lastly we look at the run time, which is very quick with an average of 4:25 minutes.

7.5 Logistic model with input error

This model ran with a threshold of 10.26. The most important result here is the value of the fraction detected points for the 20 to 80 percent simulations, which are all 0.0 and means no changes were detected. Only at the 0.0 run, 4 out of 10 changes were detected. Aside from the fact that no changes were detected there were also no false positives, leading to a 100.0 percent coverage. The average length of the returned intervals is only based on the first simulations since the others did not return any.

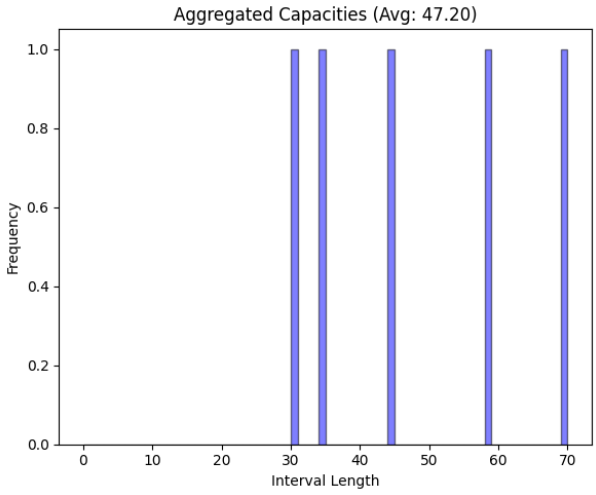
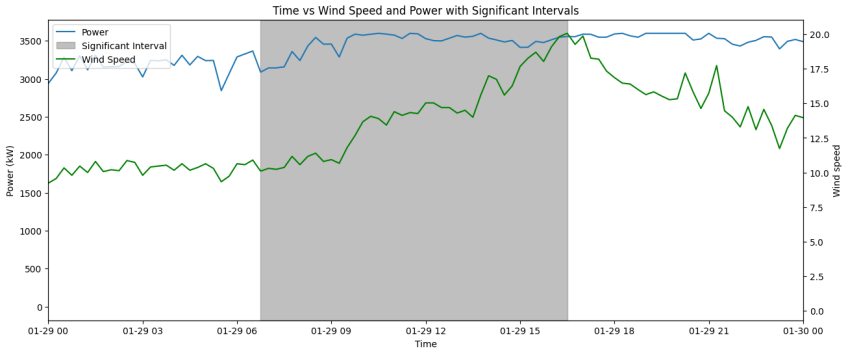


Figure 7.5: Lengths of the genuine intervals returned by the logistic model with input error, average only based on 0.0% simulation.

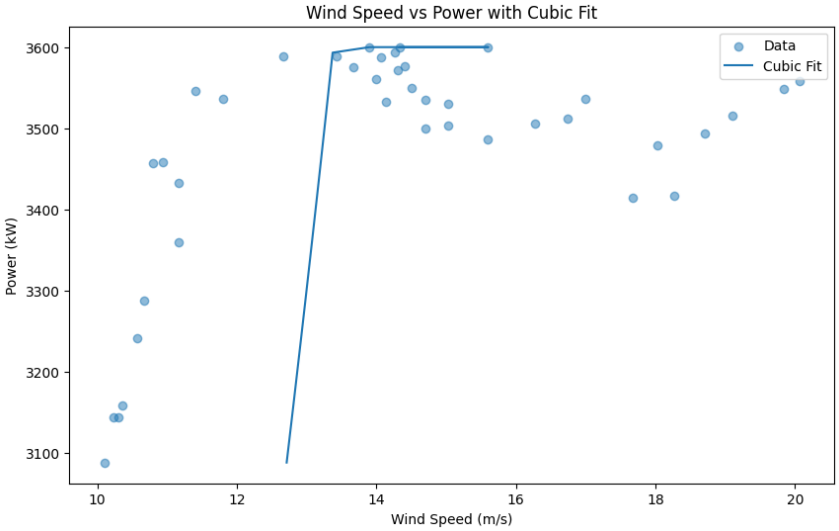
We see all of the intervals are very large which means the model is not precise enough to determine changes based on little data. This also results in no detected points for the smaller changes.

7.6 Cubic model with input error

The threshold of the cubic model with input error was 6.16. The most important result is that the average cover is only 93.99%, which is less then the required 95% on which the threshold is based. When we take a look at the separate simulations we notice the 0% capacity run has fairly good results. The coverage is 96.67 and also the fraction of detected changes and true positives are high. When we compare this result with the other simulations we see they all have a coverage of 93.33 and low values for the sensitivity and true positives. This model clearly only performs well for the change to 0%. If we simulate more paths there is a possibility the average coverage does reach 95% however this does not take away that apart from the first simulation the other perform poorly on detecting changes. This could have multiple reasons, starting with the build of the simulations. Since they are based on a logistic fit and the error is added to the power output, hence output error, this model has little in common with the basis on which the simulations are made. Nonetheless, the model does perform well on the changes to no power. To get a better idea of the wrongly returned intervals we look into this more closely. There appears to be one interval outside the anomalous periods that is returned by this model, since it is in the normal data it is returned by each capacity run. In Figure 7.6 the location and fit of this false positive is given.



(a) Falsely returned interval by the cubic model with input error.



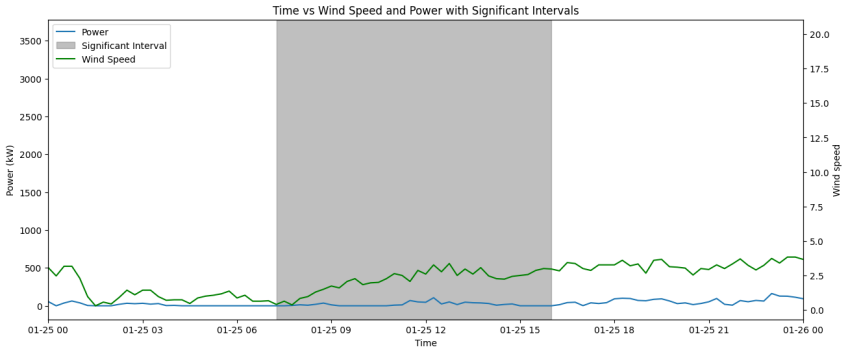
(b) Fit of the inverse cubic function at a wrongfully returned interval

Figure 7.6: Falsely returned interval by the cubic model with input error and the fitted inverse function causing the high deviation near the rated wind speed.

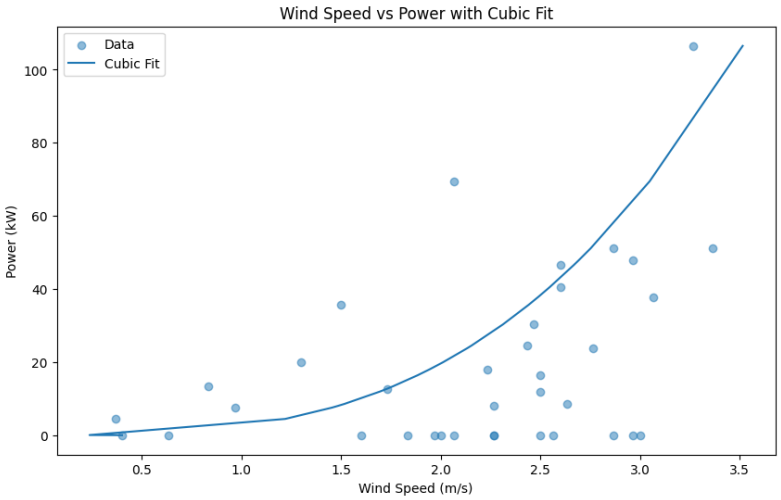
We see in Figure 7.6a the interval is situated at a moment where the wind speed increases and the power reaches its rated power. The inverse cubic fit for this period does not fit the

data perfectly but it the function does seem to be implemented well. Hence this seems to be an unlucky case of a false error and one could investigate by running more simulations of this is an isolated problem or if the average coverage is indeed below standard.

They second falsely returned interval is detected at low wind speeds, given in the figures below.



(a) Falsely returned interval by the cubic model with input error.



(b) Fit of the inverse cubic function at a wrongfully returned interval

Figure 7.7: Falsely returned interval by the cubic model with input error and the fitted inverse function causing the high deviation at low wind speeds.

In this case we see there might be an implementation fault handling zero power. When looking at the left end of the curve it is visible the fitted function displays imperfect behaviour that is related to the function implementation.

When we have a look at the average lengths of the genuine intervals in Figure 7.8 returned we notice the range is very large. The minimum length of 2 is found once but the others are larger. It shows the model has difficulty detecting changes based on little data, regardless of the capacity change.

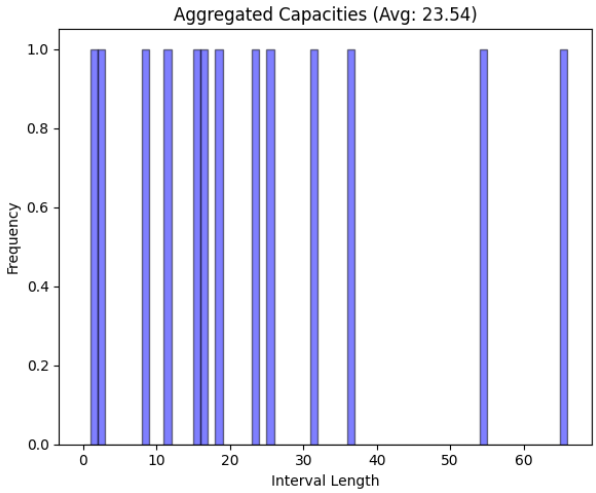


Figure 7.8: Lengths of the genuine intervals returned by the cubic model with input error, aggregation of all capacity simulations.

Lastly, the average runtime is very low. Since most runs also return little intervals it would be expected that the algorithm needs less computing time. Nonetheless, also the 0% capacity run is fast and this one has returned a similar amount of intervals. In this case we do know the returned interval is large, where normally finding the smallest interval takes time as well that this model has not used.

7.7 Isotonic regression

The NSP algorithm was applied with isotonic regression with a threshold of 125.41, based on $\alpha = 0.05$. In Table 7.1 the results for each of the capacity simulations are given. Notable is the coverage is 100.0 for all runs and thus also the fraction of true positives is 1.0. This means all returned intervals are genuine. The fraction of detected intervals stays at a constant of 0.8 for the first four simulations. We do see a small increase in interval length between these runs but there are no other significant changes between them, only a longer runtime for the 0 and 20 percentage simulations. The similarity between these runs suggests the performance of the isotonic fit does not depend highly on the severity of the capacity change. To see why two changes were missed they were separately analyzed. It was found that in these cases there was too little data to find the change. After increasing the maximum amount of points used to fit a function it was found that these points would be detected as well. The only downside is a significant increase in computation time, which is why this was not done before. To avoid this long runtime a choice could be made to use a larger order for the dyadic intervals to be able to efficiently use larger search windows.

In Figure 7.9 we see the histogram of lengths of the genuine intervals of all simulations combined. These lengths range from 1 to 6 points, with no outliers at all. Given the isotonic regression doe not estimate parameters it only needs a minimum of 2 points to determine where the change lies, which would result in an interval length of 1.0. The average length of 3.03 is a bit higher then the minimum but still a very short interval.

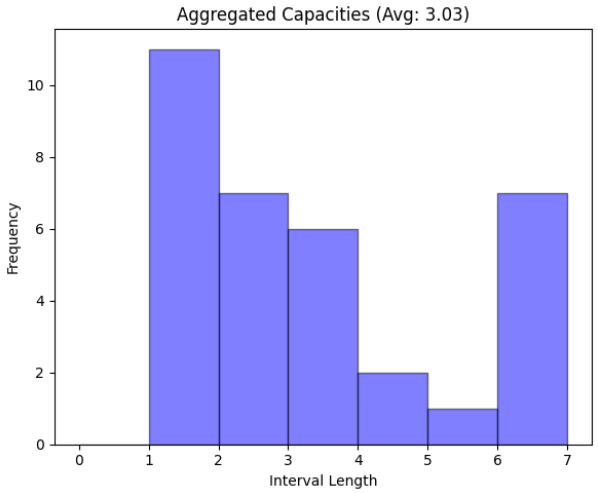


Figure 7.9: Lengths of the genuine intervals returned by the isotonic model, aggregation of all capacity simulations.

An interesting result is the difference between the first four runs and the one with 80% capacity. We see a large decrease in fraction of changes that is detected, dropping from 0.8 to 0.2. The average length stays relatively constant. This sudden drop in sensitivity suggests the model is capable of detecting changes up to a percentage between 60 and 80 and is weaker after this.

7.8 Validation on real data

In this section each of the models was applied to a month of real data from a wind turbine. For this turbine, data is available on the actual wind speed, produced power and the moments where the production was not 100% due to maintenance or curtailment, which is plotted in Figure 7.10. This data can be found in the *clean_power_data* file. The change points in this case are moments where there is a loss of power due to maintenance or downtime. Different to the simulated events is that the remaining capacity does in some cases not drop to a certain percentage in 1 PTU, but it can take multiple PTU's to reach the remaining capacity. This means the detected intervals with change points will be longer.

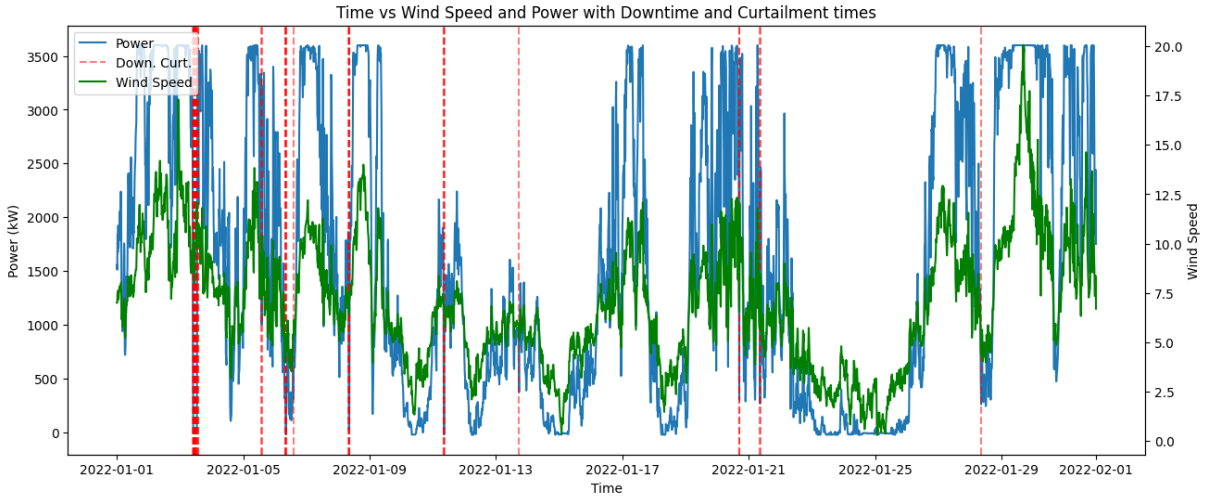


Figure 7.10: Wind speed and power data with moments of curtailment or downtime.

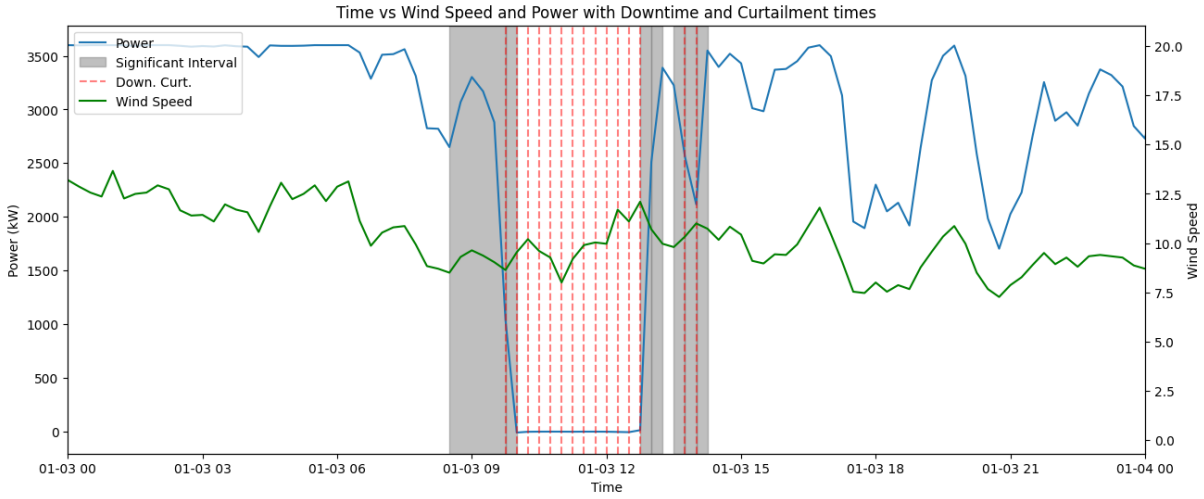


Figure 7.11: Segment of detected changes in the validation set by the isotonic model.

Since this is the same turbine as the one used for the simulations, the same thresholds for each of the models can be used. In Table 7.3 the performance of each of the models is given on this data. The logistic and cubic model with output error both have a perfect coverage of 100.0. However, they both return only a small fraction of the changes. The models with input error both perform under the required 95% coverage hence these will not be considered. After careful analysis this under performance is due to function fit in the low regime of the power curve. In reality a turbine can have negative production, when it is turned off and uses energy for internal systems, this situation was not accounted for in the implementation of the inverse functions.

Metric	Log. Out	Cub. Out.	Log. In.	Cub. In.	Isotonic
coverage	100.0	100.0	80.0	73.33	96.67
frac. det. cp.	0.2	0.34	0.34	0.57	0.6
frac. true. pos.	1.0	1.0	0.58	0.39	0.94
av. int. length	6.0	4.0	1.0	2.5	7.82
runtime	8:13	3:21	12:00	6.16	32:53

Table 7.3: Performance of the 5 metrics per model on real data.

The isotonic model has a coverage of higher than 95%, has the highest fraction of detected changes and a high fraction of true positives. In Figure 7.11 a segment of the detected changes by the isotonic model is given. Here it becomes clear the model detects the start and finish of an anomalous period and also finds an interval with two subsequent changes. The general downside for the isotonic model is the longer run time, but this is also partly due to its ability to detect more changes and thus the need to perform more computations. The high average length of the intervals is a consequence of the higher fraction of detected change.

An important difference between the results on the real data and the once of the simulation is that the simulation was based on a logistic function with output error. Yet, for the results in Table 7.3 the underlying model was unknown. Hence, here it becomes clear the models with input error do not perform well and are thus not representative for the data. With regard to the models with output error the cubic model performs better, which is in contrast to the results of the simulations. This means the fact that the logistic function was used for the simulations has indeed influenced the difference in performance of these models.

In general the isotonic model is the preferred model based on the simulation and validation results. In addition to this it is the only non parametric method. This is preferred since there

are differences in characteristics of wind turbines thus minimizing the amount of parameters that need to be estimated provides the most efficient method to scale the method to multiple turbines. The long runtime is not significant enough to be of influence on the model preference.

In Table 7.4 we see all changes in this data set and for each the detection count and average interval length. When taking a closer look to the points that were undetected by all models we see those were at indices 515, 536, 1222, 2626. Except for 515, each of those points was a change of a duration of 1 PTU, which makes it harder to detect them and in addition it does not form the start or finish of an anomalous period but is a single outlier.

Change Point	Log. Out	Cub. Out.	Log. In.	Cub. In.	Isotonic
231	1	1	1	0	1
243	1	1	1	1	1
247	0	0	0	0	1
248	0	0	0	0	1
439	0	0	0	0	1
440	0	0	0	0	1
512	0	0	0	0	1
515	0	0	0	0	0
536	0	0	0	0	0
704	0	0	1	0	1
707	1	1	1	0	1
993	0	1	1	0	1
995	0	0	1	0	1
1222	0	0	0	0	0
1891	0	0	0	0	1
1892	0	0	0	0	1
1954	0	1	0	0	1
1955	0	1	0	0	1
2626	0	0	0	0	0

Table 7.4: Detection count per model per change point, 1 if detected and 0 otherwise.

8 | Conclusion

In this chapter the conclusion on each of the research questions will be provided.

How to apply data segmentation methods to the wind energy power data to detect (partial) downtime?
To answer this question we expanded the existing Narrowest Significance Pursuit to be able to handle non-linear situations, and thus wind energy power data. Of the five proposed new frameworks for NSP, the non parametric approach applying isotonic regression performed the best in simulations and for the validation data set. Furthermore, this method requires the least knowledge on the characteristics of the wind turbine it will be applied to and avoids the problem of model misspecification the other methods face.

Can the rigorous method be used to improve upon industry standard?
It was impossible to test the method against methods currently in use, since those are not made public within the industry. However, comparison to other researched methods clearly shows the ability of NSP to detected multiple changes is an advantage over other described methods. In addition to this, the non parametric nature of the isotonic model is an advantage, especially when scaling the model to many different turbines. At this moment is was not possible to benchmark against alternative approaches such as those proposed by [Mohr and Neumeyer, 2020] and [Xu et al., 2020].

9 | Discussion

The research on multiple change point detection for wind energy using non-linear NSP has produced insightful results, encountering some limitations that necessitate further refinement and exploration. Additionally, expanding the scope of the research beyond wind energy could further enhance its relevance and application.

One limitation of the results lies in the reliance on simulations. While simulations provide useful insights they cannot fully capture the circumstances of real-world wind energy data. As a result, many more simulations are needed to cover a wider range of operational scenarios and refine the performance of the different models. For further research it is also recommend to investigate the use of simulations based on the isotonic fit, since this fit had smaller, more normally distributed residuals. Additionally, the performance of the models has not been compared to any industry benchmarks. It was not possible to assert the effectiveness of the method in comparison to existing models in the field since companies are not actively sharing their methods or results.

In practise another significant issue is the accuracy of wind data. The wind speeds are measured at several locations in the Netherlands but not always at a turbine itself. When the method is applied to turbines that do not provide the accurate wind speed, the error in wind speed between the weather station and the turbine needs to be taken into account as well. These inaccuracies can lead to higher thresholds for detecting changes, making the model less sensitive to subtle variations. Furthermore, the challenge of model misspecification persists. Inaccurate assumptions about the underlying structure of the data can negatively impact model outcomes. However, isotonic regression and other nonparametric approaches offer a solution, as they do not rely on assumptions.

Looking ahead, there are several recommendations to help improve the model and its applications. The use of isotonic regression, effective in wind energy, could be expanded to include other forms of renewable energy, such as solar power. The variability in solar energy production, similar to wind, makes it suitable for advanced change point detection methods. By broadening the scope to other renewable sources, these models could have a greater impact on the overall energy sector.

For the non-linear models with in- or output errors it would be interesting to determined the threshold based on the wind speed. This was described and proven in this research but time did not permit the implementation. Regarding the isotonic regression model, improving the efficiency is another key recommendations. While this approach is powerful in detecting changes, it can be computationally demanding as the size of datasets grows. Optimizing the model for faster computations will be necessary for its practical use.

Additionally, the use of S-shaped models could offer a good fit for the non-linear relationship in wind power. The performance of this model could not be tested in the during of this project but could offer good results as well and offers a first framework for nonparametric methods with specific shapes.

In conclusion, the current research has shown good results in advancing change point detection for wind energy, but challenges remain, such as the need for more simulations, better weather data accuracy and addressing model misspecification. Expanding the application of isotonic regression to other renewable energy forms, optimizing the computational efficiency of the model

and exploring S-shaped models are important steps to refine the approach.

Bibliography

[Barber et al., 2011] Barber, S., Wang, Y., Jafari, S., Chokani, N., and Abhari, R. S. (2011). The impact of ice formation on wind turbine performance and aerodynamics. *Journal of Solar Energy Engineering*.

[Barlow and Brunk, 1972] Barlow, R. E. and Brunk, H. D. (1972). The isotonic regression problem and its dual. *Journal of the American Statistical Association*, (337):140–147.

[Bilendo et al., 2022] Bilendo, F., Meyer, A., Badihi, H., Lu, N., Cambron, P., and Jiang, B. (2022). Applications and modeling techniques of wind turbine power curve for wind farmsâĂŤa review. *Energies*, 16(1):180.

[Bird et al., 2016] Bird, L., Lew, D., Milligan, M., Carlini, E. M., Estanqueiro, A., Flynn, D., Gomez-Lazaro, E., Holttinen, H., Menemenlis, N., Orths, A., et al. (2016). Wind and solar energy curtailment: A review of international experience. *Renewable and Sustainable Energy Reviews*, 65:577–586.

[Carrillo et al., 2013] Carrillo, C., Montaño, A. O., Cidrás, J., and Díaz-Dorado, E. (2013). Review of power curve modelling for wind turbines. *Renewable and Sustainable Energy Reviews*, 21:572–581.

[Carroll et al., 2016] Carroll, J., McDonald, A., and McMillan, D. (2016). Failure rate, repair time and unscheduled o&m cost analysis of offshore wind turbines. *Wind energy*, 19(6):1107–1119.

[Chinchilla et al., 2006] Chinchilla, M., Arnalte, S., Burgos, J. C., and Rodríguez, J. (2006). Power limits of grid-connected modern wind energy systems. *Renewable energy*, 31(9):1455–1470.

[Costa et al., 2021] Costa, Á. M., Orosa, J. A., Vergara, D., and Fernández-Arias, P. (2021). New tendencies in wind energy operation and maintenance. *Applied Sciences*, 11(4):1386.

[Dao, 2021] Dao, P. B. (2021). A cusum-based approach for condition monitoring and fault diagnosis of wind turbines. *Energies*, 14(11):3236.

[Eurostat, 2024] Eurostat (2024). Share of energy from renewable sources. https://ec.europa.eu/eurostat/databrowser/view/nrg_ind_ren/default/table?lang=en. [Accessed 11-09-2024].

[Feng et al., 2021] Feng, O. Y., Chen, Y., Han, Q., Carroll, R. J., and Samworth, R. J. (2021). Nonparametric, tuning-free estimation of s-shaped functions.

[Fryzlewicz, 2023] Fryzlewicz, P. (2023). Narrowest significance pursuit: inference for multiple change-points in linear models. *Journal of the American Statistical Association*, pages 1–14.

[Garreau and Arlot, 2018] Garreau, D. and Arlot, S. (2018). Consistent change-point detection with kernels. *Electron.*

[GWEC, 2024] GWEC (2024). Gwec | global wind report 2024.

[Haac et al., 2022] Haac, R., Darlow, R., Kaliski, K., Rand, J., and Hoen, B. (2022). In the shadow of wind energy: Predicting community exposure and annoyance to wind turbine shadow flicker in the united states. *Energy Research & Social Science*, 87:102471.

[Honda, 1997] Honda, T. (1997). The cusum tests with nonparametric regression residuals. *Journal of the Japan Statistical Society*, 27(1):45–63.

[Hossin and Sulaiman, 2015] Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2):1.

[Jangamshetti and Rau, 2001] Jangamshetti, S. H. and Rau, V. G. (2001). Normalized power curves as a tool for identification of optimum wind turbine generator parameters. *IEEE Transactions on Energy Conversion*, 16(3):283–288.

[Jing et al., 2021] Jing, B., Qian, Z., Zareipour, H., Pei, Y., and Wang, A. (2021). Wind turbine power curve modelling with logistic functions based on quantile regression. *Applied Sciences*, 11(7):3048.

[Johnson, 1985] Johnson, G. L. (1985). *Wind Energy Systems*. Citeseer.

[Kabluchko, 2007] Kabluchko, Z. (2007). Extreme-value analysis of standardized gaussian increments. *arXiv: Probability*.

[KNMI, 2024a] KNMI (2024a). Automatische weerstations.

[KNMI, 2024b] KNMI (2024b). Weather model - harmonie-arome cy43 forecast netherlands meteorological parameters.

[Latiffianti et al., 2022] Latiffianti, E., Sheng, S., and Ding, Y. (2022). Wind turbine gearbox failure detection through cumulative sum of multivariate time series data. *Frontiers in Energy Research*, 10:904622.

[Leahy et al., 2019] Leahy, K., Gallagher, C., OâĂŹDonovan, P., and OâĂŹSullivan, D. T. (2019). Issues with data quality for wind turbine condition monitoring and reliability analyses. *Energies*, 12(2):201.

[Letzgus, 2020] Letzgus, S. (2020). Change-point detection in wind turbine scada data for robust condition monitoring with normal behaviour models. *Wind Energy Science Discussions*, 2020:1–29.

[Lin et al., 2020] Lin, Z., Liu, X., and Collu, M. (2020). Wind power prediction based on high-frequency scada data along with isolation forest and deep learning neural networks. *International Journal of Electrical Power & Energy Systems*, 118:105835.

[Liu et al., 2008] Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE.

[Luque et al., 2019] Luque, A., Carrasco, A., Martín, A., and de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91:216–231.

[Lydia et al., 2014] Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P. (2014). A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30:452–460.

[Lydia et al., 2015] Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P. (2015). Wind resource estimation using wind speed and power curve models. *Renewable Energy*, 83:425–434.

[Mehrjoo et al., 2020] Mehrjoo, M., Jozani, M. J., and Pawlak, M. (2020). Wind turbine power curve modeling for reliable power prediction using monotonic regression. *Renewable Energy*, 147:214–222.

[Mohr and Neumeyer, 2020] Mohr, M. and Neumeyer, N. (2020). Consistent nonparametric change point detection combining cusum and marked empirical processes. *Electronic Journal of Statistics*.

[Renani et al., 2016] Renani, E. T., Elias, M. F. M., and Rahim, N. A. (2016). Using data-driven approach for wind power prediction: A comparative study. *Energy Conversion and Management*, 118:193–203.

[Schlechtingen and Santos, 2011] Schlechtingen, M. and Santos, I. F. (2011). Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical systems and signal processing*, 25(5):1849–1875.

[Seo et al., 2019] Seo, S., Oh, S.-D., and Kwak, H.-Y. (2019). Wind turbine power curve modeling using maximum likelihood estimation method. *Renewable Energy*, 136:1164–1169.

[Stute, 1997] Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, pages 613–641.

[Su and Xiao, 2008] Su, L. and Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. *Statistics and its Interface*, 1(2):347.

[Tomé et al., 2017] Tomé, R., Canário, F., Leitão, A. H., Pires, N., and Repas, M. (2017). Radar assisted shutdown on demand ensures zero soaring bird mortality at a wind farm located in a migratory flyway. *Wind Energy and Wildlife Interactions: Presentations from the CWW2015 Conference*, pages 119–133.

[Villanueva and Feijóo, 2018] Villanueva, D. and Feijóo, A. (2018). Comparison of logistic functions for modeling wind turbine power curves. *Electric Power Systems Research*, 155:281–288.

[Wang et al., 2023] Wang, Z., Liu, W., and Wang, X. (2023). Abnormal data cleaning of wind turbine power curve using bayesian change point-quartile combined algorithm. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 237(3):495–503.

[Wen et al., 2009] Wen, J., Zheng, Y., and Donghan, F. (2009). A review on reliability assessment for wind power. *Renewable and Sustainable Energy Reviews*, 13(9):2485–2494.

[Wind Turbines-Part, 2005] Wind Turbines-Part (2005). *12-1: Power Performance Measurements of Electricity Producing Wind Turbines*, volume 61400-12 of *IEC Standard*. British Standard IEC, London.

[Xu et al., 2020] Xu, Q., Lu, S., Zhai, Z., and Jiang, C. (2020). Adaptive fault detection in wind turbine via rf and cusum. *IET Renewable Power Generation*, 14(10):1789–1796.

A | Appendix

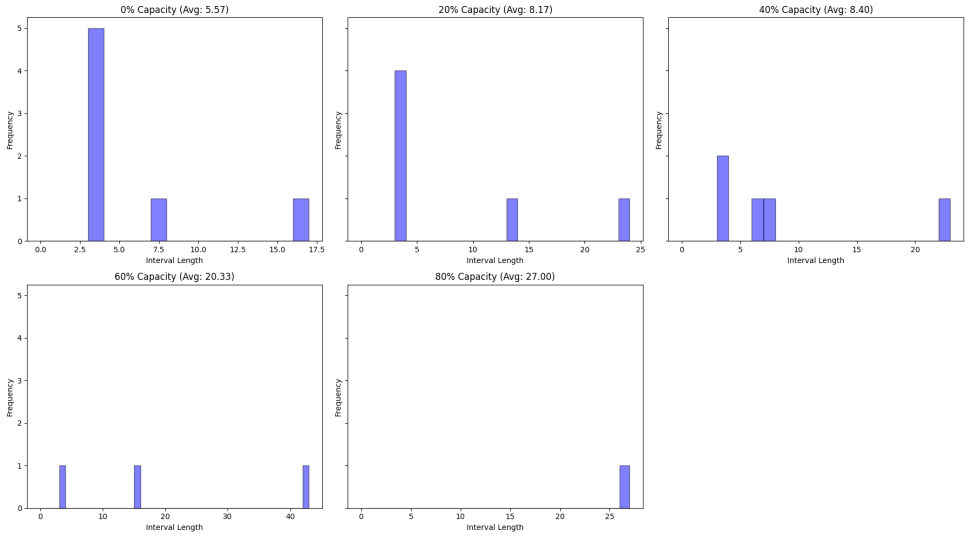


Figure A.1: Lengths of the genuine intervals per capacity simulations, returned by the logistic model with output error.

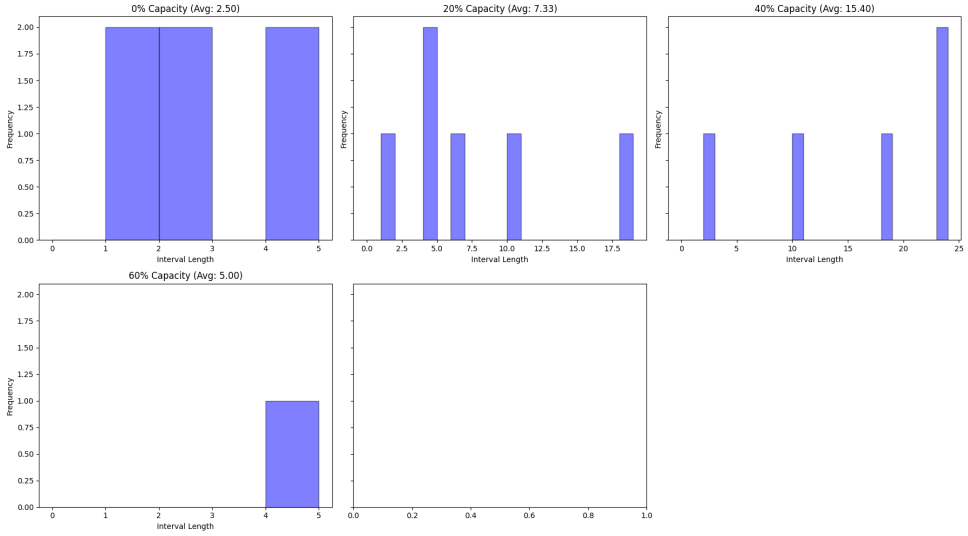


Figure A.2: Lengths of the genuine intervals per capacity simulations, returned by the cubic model with output error.

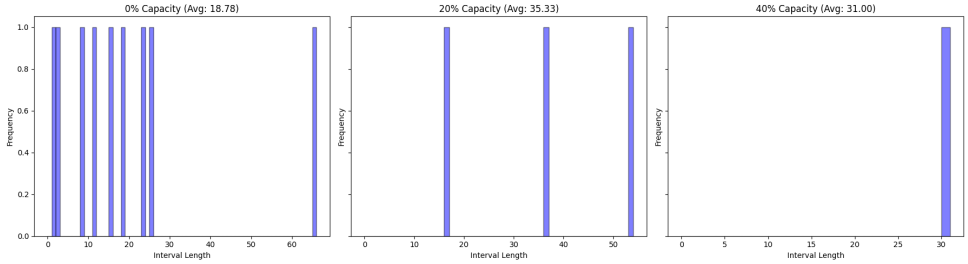


Figure A.3: Lengths of the genuine intervals per capacity simulations, returned by the cubic model with input error.

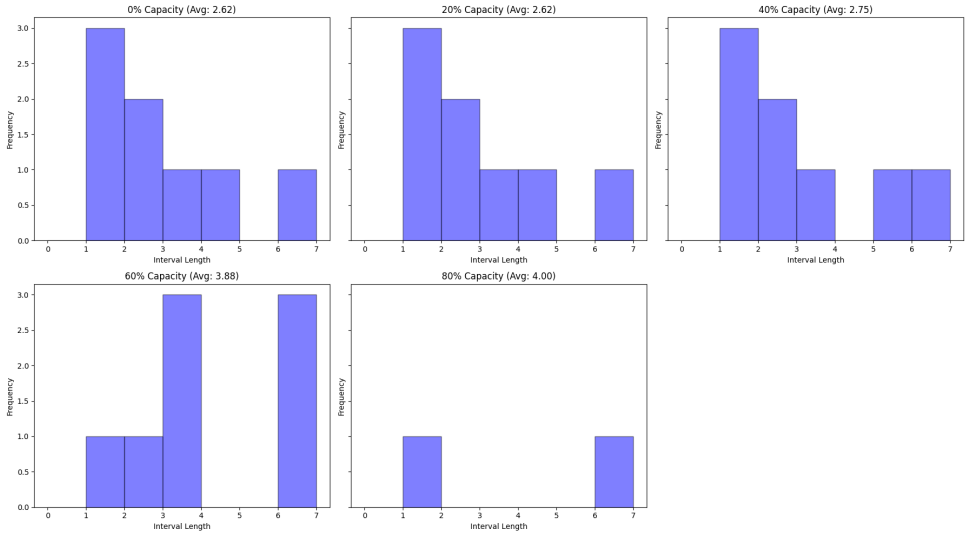


Figure A.4: Lengths of the genuine intervals per capacity simulations, returned by the isotonic model.