Utrecht University

Faculty of Science
Department of Information and Computing Sciences

Master Thesis

Thesis number: ICA-4016548

Outlier Detection in Energy Climate Data

Author
Erik Duijm

Daily Supervisor
Laurens P. Stoop MSc.

First Supervisor
Dr. Ad Feelders

Second Supervisor
Prof. Dr. Arno Siebes

February 12, 2021

Abstract

The energy transition, moving away from fossil fuels to renewable energy sources, introduces an increasing variability in energy generation. In order to prepare for the future, significant improvements need to be made to the energy grid. We investigate the use of outlier detection algorithms to improve the assessment of future energy systems. Outliers represent critical conditions that should be taken into consideration by policy makers when designing the future energy grid. We combine the MDI algorithm and the SLOM algorithm with novel post-processing to detect temporal and spatial-temporal outliers. These algorithms are applied to energy generation data derived from ERA5 historical climate reanalysis data using energy conversion models. Using the MDI algorithm we found temporal outliers that are potential risks for the energy grid. We found that the application of SLOM, a spatial outlier detection algorithm, and the post-processing, provided no new insights. Historical trends that could be attributed to climate change were investigated but not found. For the historical period we found that outlier intensity might be influenced by multidecadel variability. We conclude that our method shows that outlier detection might help the assessment of the future energy grid by highlighting the most extreme situations. Researchers and policy makers could use information on the discovered outliers to improve the future development of the electricity network.

Contents

1 Introduction

A consequence of climate change is a significant increase in predicted average temperature [1]. This will have devastating consequences [2]. There is high confidence in an increased risk of extreme weather. Such weather might cause the breakdown of infrastructure networks and critical services, food and water insecurity, loss of rural livelihoods, and more [2].

In attempts to combat this increase, the world is moving away from fossil-fuels as an energy source. The Paris agreement, currently ratified by 189 parties [3], represents a political response to climate change. One of the goals in the agreement is the limiting of global temperature increase well below two degrees, pursuing efforts to limit the increase to 1.5 degrees [4]. In line with their commitment to the Paris agreement, the European Union adopted the European Green Deal [5]. A key objective of the European Green Deal is to be climate-neutral by 2050 . This means net-zero greenhouse gas emissions, which would require an energy transition.

With the energy transition from fossil-fuel driven generators to renewable energy sources like wind and solar energy, comes the introduction of additional variability on the generation side of the electricity grid [6]. At the same time the demand side of the electricity grid is already very variable [7, 8]. Additionally this demand side is expected to become more variable due to the expected electrification of space heating [9, 6]. This enhanced variability of the energy system can lead to weather driven critical conditions that can damage the electrical grid [10]. Such critical conditions can be avoided with adequate balancing [11] and back-up systems, given enough knowledge about their cause and the effect they have on the system. Thorough knowledge of critical conditions and the systems that cause them will provide policy makers with the information needed to allocate risk reducing investments during the energy transition [12].

The goal of this research is to gain insight into critical systems that might impact the future energy grid, using outlier detection. These critical systems indicate inadequacies in the energy grid, and are vital for policy makers to make decisions about the future energy grid.

Outlier detection algorithms are utilized to find these critical systems. Outlier events found by the detection algorithm represent the rarest and most extreme events that the models and data predict. Any risk analysis performed on the power grid should include these extremes, as these are indications of the maximum stress the power grid must be expected to withstand. Knowledge of these outlier events can improve risk analysis, as well as improve our understanding of extreme climate events.

To achieve this goal a literature study was performed to discover outlier detection algorithm well suited for our research. Based on this literature study, two algorithms were selected. The first outlier detection algorithm is the Maximally Divergent Interval algorithm by Rodner et al. (2016) [13] and Barz et al. (2017, 2018) [14, 15]. This algorithm is capable of finding temporal outliers, temporal outliers and their

spatial location, and spatial-temporal outliers. The second algorithm is the SLOM algorithm by Chawla and Sun (2006) [16], which detects spatial outliers.

These algorithms are used to detect outliers in renewable energy generation data. This data is based on historical weather reanalysis data. Energy conversion models are used to turn the weather data into renewable energy data. Several experiments are performed to investigate outlying events and their properties.

First we provide a description of the problem in chapter 2. In chapter 3 we provide a discussion on the dataset used for this research. This is followed by an overview of the related work and methods for outlier detection in chapter 4. Chapter 5 contains an in depth explanation of the outlier detection algorithms and how they were used. In chapter 6 the performed experiments and their results are discussed. Finally a conclusion and discussion are provided in chapters 7 and 8.

2 Problem Statement

Algorithmic data analysis is the field that studies how to analyse data and extract useful information. It is very important to keep the area or field being studied in mind. The properties and sources of the available data and the peculiarities of a field of research determine what algorithms can be applied. Therefore we provide a brief introduction into the energy climate field and the associated challenges and problems.

First we discuss the European energy system in section 2.1. The future energy grid and potential associated risks are discussed in section 2.2. Finally, in section 2.3 the research questions are discussed.

2.1 The European Energy System

The European energy system is a highly interconnected and complicated system[17, 18]. In figure 1 the European electricity transmission network is depicted in a simplified form. A detailed and interactive map can be found at `https://www.entsoe.eu/data/map/`. Note the many connections between nations and the many different types of power lines, stations and powerplants that are all connected.

A reason for the high connectivity of the network is the energy market and the stability it brings. National shortfalls can be met by neighbouring surpluses which increases the stability of the entire European energy grid. However since the network is connected, it needs to be working on the same frequency. The European energy grid runs on a frequency of 50 Hertz [20]. This frequency limitation causes the European energy system to be a very complex system, since fluctuations in supply and demand influence the frequency of the whole system. Additionally, as machines and appliances in Europe are tuned to a 50 Hertz frequency, shifts in the frequency can cause malfunctions.

Due to the interconnected nature of the European energy system, and the margins on the allowed frequency, local shortfalls and surpluses have significant impact on the entire system. This means that it is difficult to analyze small parts of the energy system, as the bigger picture is required to gain proper understanding.

The consequence of this is that datasets relating to the energy system are very large, as they need to contain the entire network, spread across Europe, as well as have a high resolution [21].

Due to the variability of renewable energy sources, especially when compared to fossil fuel generators, the energy system must become more flexible and robust in dealing with variance. This requires the network to become even more connected in order to increase the size and flexibility of the energy market. In order to make the required changes to the energy grid, it is important to have a proper understanding of the risks the energy system faces.

2.2 Risks Facing the Future Energy System

Due to the increasing demand for energy [22], as well as the electrification of space heating, alterations and improvements need to be made to the energy grid. These

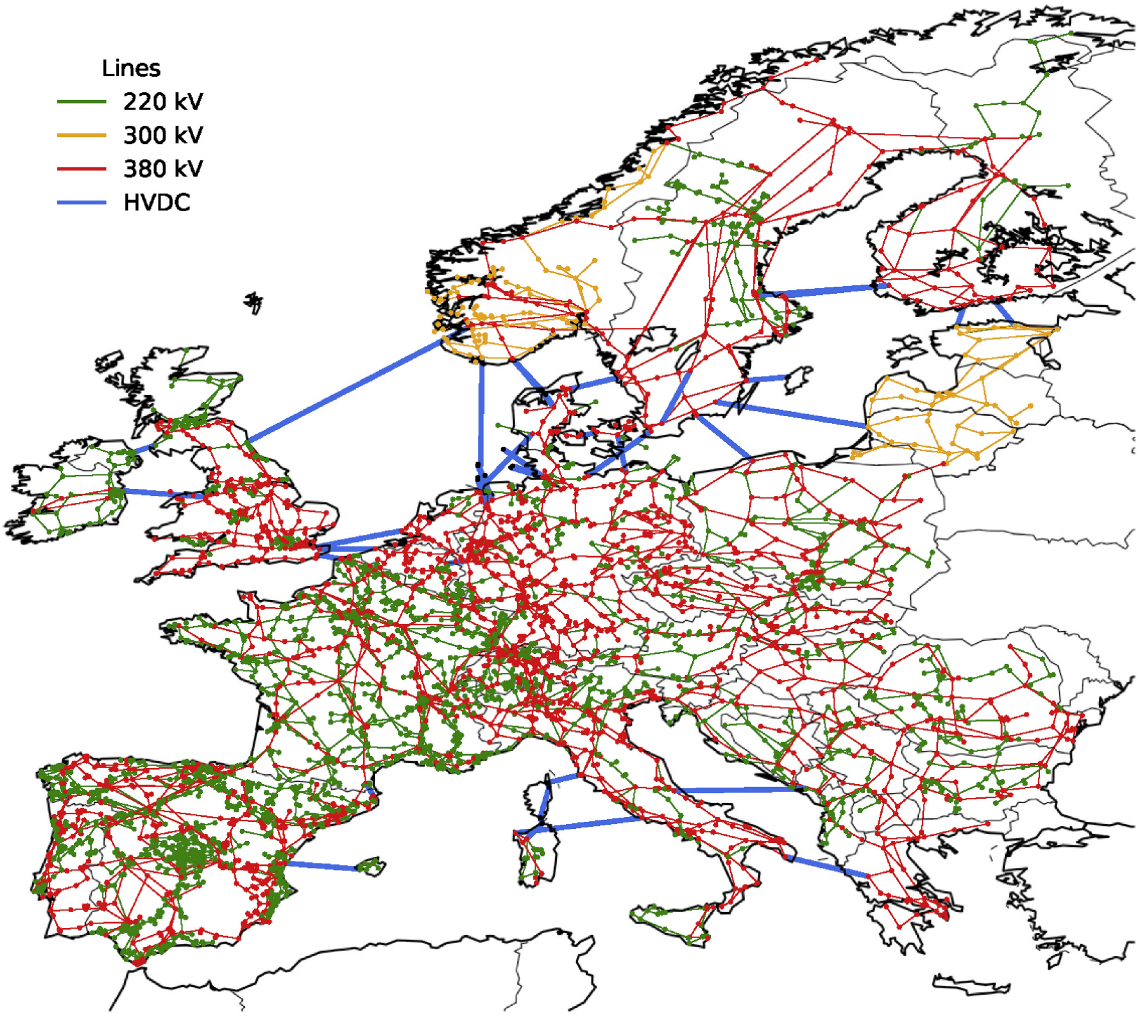


Figure 1: European transmission network model, includes lines that are planned and under construction. Image from Hörsch et al. [19].

changes must ensure that the energy system is ready for the future, as blackouts and frequency shifts have catastrophic consequences [23]. In order to determine the capacity and demand of the future energy grid, simulations and optimizations are being performed for various possible scenarios of the future [24]. These can advise policy makers in their choices for grid development, such as what type of additional energy generation and storage capacity are required and what transmission lines should be extended or added.

The complexity of the European energy system causes the analysis of the future energy grid to be a computationally difficult task [25]. In order to assess risks and make policy for the future energy system, scenarios of the future are studied. For each future scenario an optimized and operable schedule of operation needs to be defined for all generators that are connected to the grid. Additional variability is introduced into this already complex scheduling problem by the energy transition to renewable energy sources. Because of this complexity, additional methods for investigating the future energy grid are being studied. A promising method to reduce the impact of variable weather on the optimization problem is the subsampling approach developed by Hilbers et al. [26]. However this method relies on knowledge of outliers to reduce the problem.

In this research outlier detection is used to discover outlying events that represent concerns towards the future energy grid. First of all, these situations are potentially interesting for policy makers to investigate as they represent risks in the system [27]. Secondly, knowledge of outliers could be used in the subsampling approach by Hilbers et al. [26] to reduce climate based uncertainty.

For this research we are interested in several situations that can occur in the future energy grid. These situations represent deficiencies of the energy grid. Such deficiencies pose risks to the future energy security of Europe.

We have identified three different situations as events that pose a risk to the future energy grid.

- Low energy generation.
  This situation, regardless of its location, is interesting as it indicates the need for more energy sources or better backup systems.

- Insufficient energy generation with regards to demand.
  Similar to the first situation, but depending on the demand. This could mean local shortages such as per nation, or group of nations. This situation identifies the need for more energy capacity and possibly backup systems. In addition this could identify local deficiencies in the system.

- Sufficient energy generation, but local deficiencies need to be compensated by spatially distant energy surpluses.
  Situations like these, where for example an energy shortage in Spain, must be supplied from a surplus in Germany puts a lot of strain on the energy grid. Such situations show possible deficiencies in local energy capacity and could indicate the need for better connectivity of the grid.

In addition, the outlier detection algorithms might detect outlier events that do not match any of these situations. Finding these new scenarios is very interesting as

they might represent previously unthought of situations that threaten future energy security. Such outlier events can lead to new insights.

2.3 Research Questions

For this research into outliers in the future energy grid, the following research question is posed:

In what way can outlier detection algorithms be used to improve the assessment of future energy systems?

In order to specify the research question, we have divided it in three sub-questions:

1. What adaptations to the algorithms need to be made to apply state-of-the-art spatial-temporal outlier detection algorithms to energy system data?
2. What are the discovered spatial-temporal outliers and do they pose a threat to energy security?
3. Do current levels of climate change have an effect on outlier events, and if so, what is this effect?

In order to answer the first sub-question, a literature study was performed, the results of which are presented in chapter 4, as the related work. Based on that study two promising state-of-the-art outlier detection algorithms are selected.

The outlier detection algorithms will be used to detect extreme events. These events will be analyzed to see if they match any of the previously described situations that pose risks for the energy grid are investigated, or are risky situations not considered beforehand.

The effect of climate change will be studied by applying the outlier detection algorithms to several different decades of data. In the data used, the effect of climate change over those decades is incorporated. The outlier events detected per decade are investigated in order to study if, and how, they change over time.

3 Data

To investigate future energy systems, several data sets are available. Our work will focus on ERA5 reanalysis data [28, 29, 30]. In section 3.1 we discuss the properties of the ERA5 dataset. The energy conversion models used to create energy generation data based on ERA5 data are discussed in section 3.2.

3.1 ERA5

The ERA5 dataset contains historical climate data, and will be used to investigate what adaptations need to be made to the proposed outlier detection algorithms. Since the data is historical, outlier events represent historical events. Thus the detection of extreme historical events can be used as some sort of ground truth for the algorithms.

Working with this dataset is challenging, as it contains climate data. Climate data is autocorrelated as is stated by Tobler's first law of geography [31]: "*Everything is related to everything else, but near things are more related than distant things*". This correlation invalidates the assumption that data is independent and identically distributed, on which many modern models are based [32].

In addition, climate data is heteroscedastic. This fluctuating variance needs to be factored into the outlier detection algorithms. The high variability of climate data needs to be taken into account as well. This variability makes it hard to distinguish outlier events from regular variation in the data. All these challenges need to be overcome by the outlier detection algorithms.

ERA5 reanalysis data stretches from 1950 to the present, with a two month delay. The period between 1950 and 1979 is the preliminary version of the ERA5 back-extension [33]. The ERA5 back-extension has undergone significant quality control and is considered state-of-the-art. It should be noted that using the ERA5 back-extension for tropical cyclones is not advised [34]. Since our region of interest, Europe, is typically far removed from tropical cyclones we will be using the back-extension.

For this research, data up to, and including, 2019 will be used. The spatial resolution of the data is $\pm$ 31 kilometers or 0.25 degree. The temporal resolution is hourly.

In order to limit the size of the data we will be looking at a spatial subsection of Europe. We define the region of Western Europe as the area with latitude -14.75 to 40 and longitude 35 to 74.75. For the remainder of this thesis, when referring to the full region, we are talking about the Western Europe region. This area contains nearly all renewable energy generation and thus the most relevant data.

In order to create energy generation data, conversion models are needed. The energy conversion model used for this research is described in section 3.2. The converted dataset has a size of 568GB. After conversion we are left with three variables. A wind-onshore (WON), wind-offshore (WOF) and solar photovoltaic (SPV) energy generation variable.

An example of the resulting data can be seen in figure 2. This data shows the

onshore wind energy generation on the Western European region on August 16th 1994 at 20:00.

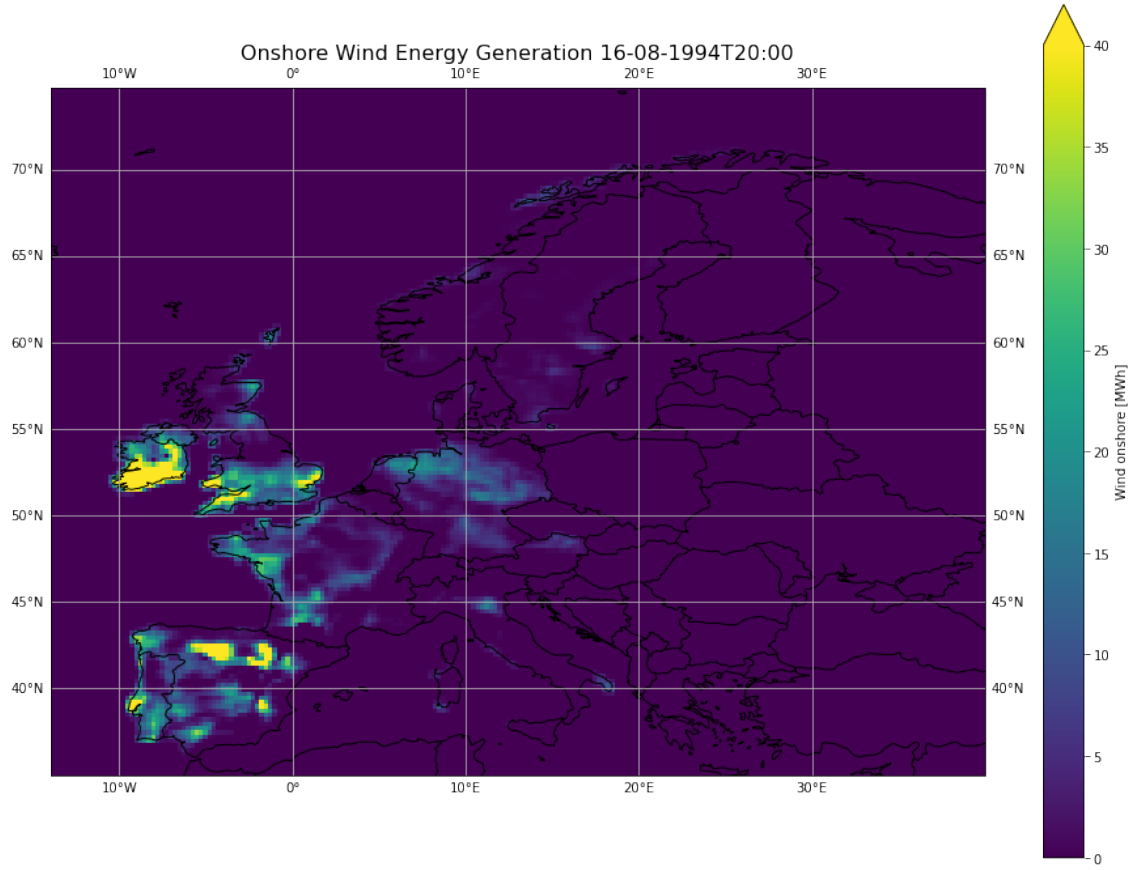


Figure 2: Single time step of energy generation based on ERA5 reanalysis data showing onshore wind energy generation of Western Europa on 16/08/1994 20:00.

3.2 Energy Conversion Models

To calculate the energy generation based on climate model data you need to know two things; the efficiency of the wind turbine/solar photovoltaic panel and the distribution of these over the region you want to study. The first can be done by using conversion models that give you a capacity factor for each grid cell based on the climate variables in that grid cell. The second, a distribution of renewable energy sources, can be made through a variety of methods. We have chosen here to use the cost-optimal distribution that was developed by van Zuijlen et al [35]. By multiplication of the capacity factor and the installed capacity from the cost-optimal distribution you obtain the energy generation per grid cell.

The conversion model used to obtain wind turbine energy capacity factors follows the power curve method [36]. As there is a variety of methods for the conversion of wind to energy, several methods were compared [37, 38, 39, 40] and small adjustments where made to the method as presented in [36]. The adjustment procedure and model comparison will be described in full detail in [41]. In summary the power

curve is adjusted by limiting the maximum capacity factor to 95% and by the addition of a quadratic decay in the capacity factor in the tail of the distribution.

The conversion model used to obtain the solar photo-voltaic energy capacity factors follows the method as set out by [42]. By using this method no assumptions have to be made about the specific properties of the solar panels, while keeping the thermal effects on the solar panel efficiency tractable. A small adjustment in the implementation was made as rounding errors caused small negative values in summer twilight periods. By forcing negative values to zero after the conversion this issue was solved.

4 Related Work

Outlier detection is a well studied and popular field, as it provides insight into the most extreme cases in data [43, 44, 45]. The detection of outliers has many applications. It is used in the medical field, with for example disease cluster detection [46]. Another application is traffic flow analysis [47], which can be used by policy makers to make more informed decisions on road network improvements. In climate research, outlier detection is very important, as climate extremes can cause high impact events. Using outlier detection to detect hurricanes [48], could for example be used to give hazard warnings and timely warn society to reduce their impact.

A different application of outlier detection, is the discovery of invalid data points. These invalid data points can be caused by inaccurate measurements, or broken equipment. Adam et al. [49] use Voronoi diagrams to detect outliers caused by malfunctioning sensors. The Voronoi diagrams are used to create micro neighborhoods around data points. These neighborhoods are merged into larger macro neighborhoods. They consider data points isolated from other data points in the same macro neighborhood to be outliers.

Both looking for invalid data, and correct but extreme data points are considered outlier detection. It should be noted however that these have significantly different goals and approaches. Therefore it is important to clearly state what one considers an outlier. In the remainder of this thesis, when discussing outliers, we are referring to extreme situations in data.

In the following section an overview of outlier detection is given, and the main algorithmic approaches are discussed. This is followed in section 4.2 by a discussion of outlier detection on spatial, temporal, and spatial-temporal data.

4.1 Outlier Detection Overview

In order to gain a proper understanding of outlier detection, it is important to have an understanding of the definitions and common algorithmic approaches used in the field.

A data point that is very different from the remaining data can be considered an outlier. A popular definition of an outlier is given by Hawkins [50]:

"*An observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism.*"

To detect outliers, there are several different algorithmic approaches as described in chapter 8 of *Data mining: the textbook* [51]. There are extreme value, density, clustering, probabilistic, information theoretic and distance based methods. Each of these approaches is discussed in the following sections.

Extreme value analysis

Extreme value analysis looks at the statistical tails of probability distributions to find outliers. These extreme values at the tails of the distribution, are considered

outliers. However not every outlier is an extreme value. Hence Extreme value analysis is suited for specific applications.

Extreme value analysis is well suited for multivariate data, as the distance measure used to determine extreme values, e.g. Euclidean, can handle multivariate distances. The Mahalanobis distance can be used to scale the importance of variables along its principle component, to incorporate the shape of the data. Laurikkala et al. [52] performed extreme value analysis using the Mahalanobis distance to discover outliers in female urinary data and vertigo data. They created box plots based on these measures, and considered points outside of the box plots outliers. The impact of distance measures indicate the importance of the selection of the correct distance measure for specific problems.

Density based methods

When using a density based approach to determine outliers, the idea is to determine unusually sparse or dense regions in data. For one dimensional data histogram based methods might be used. These count the number of other data points in the same bin as the tested data point as outlier score. The difficulty here is the correct tuning of bin sizes. The histogram approach can be seen as a special case of the multivariate grid based method, where a grid structure is used. The number of data points in the same grid as the tested data point is used as an outlier score. The grid based method also requires a careful tuning of grid sizes, as otherwise outliers might be missed or normal data might be classified as an outlier. Another difficulty arises with high dimensional data, as N-dimensional data requires a minimum of 2^N grid cells.

Kernel density estimation methods are comparable to histogram based methods, in the sense that they also model density profiles. However kernel density estimation uses continuous density estimates, estimated by summing smoothed kernel functions. The estimated density at a point is reported as the outlier score. Kulldorf introduces a scan statistic in his 1997 paper [53]. This scan statistic compares data points that are in some state, to a baseline of an expected state. For example the number of diseased people compared to the complete population. In the paper this statistic is used to find regions with an unusually high number of sudden infant death syndrome in North Carolina.

Clustering based methods

When applying a clustering algorithm, data points not belonging to any cluster might be considered outliers. This is at the core of clustering based outlier detection. However just clustering is not sufficient, as a mechanism that causes outliers might repeat itself. This could create a small cluster of outliers, and thus clusters should be checked to see if they are outlier clusters. Even if every point lies within a larger cluster, there might still be outliers in the data. Clustering based methods need to check points within clusters. A possibility is checking the distance of data points to the centroid of the cluster. Data points with a high distance to the cluster centroid are potential outliers. Loreiro et al. [54] use a cluster based outlier detection algorithm to discover errors in foreign trade data.

Using the Mahalanobis distance, calculated using only points in the same cluster, the shape of each cluster is incorporated in the outlier detection. Depending on the application this is desirable. In their 2011 report, Warren et al. [55] use the Mahalanobis distance in cluster based outlier detection to discover anomalous traffic behaviour on the I-95 highway in New York.

Probabilistic methods

Probabilistic based methods line up with the definition given by Hawkins very well. These approaches look at a data point, and fit a model to the remaining data points. The probability of the original data point being generated by the model is then calculated. If this probability is low enough, it is considered to be generated by a different model.

This method is very flexible, as it allows for fitting of any model and works for multivariate data. Probabilistic approaches benefit from prior knowledge of the data, as knowledge of the generative model simplifies the process. Bauder et al. (2017) [56] use a probabilistic outlier detection algorithm on output of a regression model trained on their data. Doing this they discover fraudulent claims in health insurance declarations.

Information theoretic methods

Information theoretic methods are based on compression. The idea is that regular, non outlying, data follows some distribution, and that patterns in this distribution can be used to efficiently compress the data. Outlying points, that do not fit the distribution, do not match the patterns of the remaining data. The data points that are responsible for the biggest increase in the model size that is needed to compress the data are potential outliers.

Smets and Vreeken (2011) [57] detect outliers in transactional and binary data sets based on the minimum description length principle. They train a compressor, KRIMP [58], on regular data. If the trained compressor finds that a sample needs an unusually large number of bits to compress, it is considered an outlier.

Distance based methods

The distance from a data points to its neighbors, can be considered a measure of outlierness. This is at the core of distance based outlier detection algorithms. Many variants exist, such as using the distance to the k-nearest neighboring data point. Setting this parameter allows for finding the single most outlying points by setting k equal to one, or identifying small outlying groups of data points by selecting k larger than one. Another variant is using the average distance to the k-nearest neighbors. The *ODIN* algorithm introduced by Hautamäki et al. [59] finds outliers based on the k-nearest neighbor distance, or the average k-nearest neighbor distance, using a k-nearest neighbor graph.

Using local distance measures is an important alternative, as local outliers are different than global outliers, and provide different insights. In their 2000 paper, Breunig

et al. [60] present one of the earlier works on local outliers. They argue the importance of local outliers over global outliers, and introduce the Local Outlier Factor (LOF). The LOF encapsulates the degree to which a data point is an outlier with regards to its neighboring data.

The Local Outlier Factor is one of the first quantifiers of outlierness. This scoring of outliers allows for ranking of outliers, which could provide additional insight in the discovered outliers. Several other measures of local outlierness exist, such as the multi-granularity deviation factor, introduced by Papadimitriou et al. [61], which can account for variations in the local density. The connectivity-based outlier factor by Tang et al. [62] is based on LOF. It improves effectiveness when outliers have low density neighborhoods. Zhang et al. [63] introduce the local distance-based outlier factor. This factor is better suited for scattered data sets, and is calculated using relative distance to local neighborhoods.

4.2 Outlier Detection for Spatial and Temporal Data

When working with climate data, or impact data derived from climate variables, challenges arise [64]. In large part these challenges are caused by the spatial, temporal, or the spatial-temporal nature of the data. These additional difficulties need to be taken into account when detecting outliers. In the following section we discuss several algorithms that discover outliers in spatial, temporal or spatial-temporal data and why they are or are not suited for our research. As discussed in chapter 3, the ERA5 data we will be using is derived from climate data and has a regular grid structure.

4.2.1 Spatial Data

The type of data that is being researched has a significant impact on what defines an interesting outlier. Chawla and Sun (2006) [16] work with spatial data. Their Spatial Local Outlier Measure (SLOM) expresses how a data point behaves with regards to their spatial neighborhood. The SLOM score is defined to take autocorrelations and heteroscedasticity present in spatial data into account. This is desirable for our research, as climate data is both autocorrelated and heteroscedastic.

Wu et al. (2008) [65] detect and track heavy rainfall events. Their method first detects spatial outliers using Kulldorf's scan statistic [53]. The statistic is used to find the top k highest outlying regions. These outlying spatial regions are then tracked through time, to discover spatial-temporal outliers. This is achieved using a tree structure of all possible time series of overlapping outlier regions. However this approach is exponential in time. Their tree building algorithm can create potentially very large trees and the complexity of the algorithm to traverse the tree to find all outlier sequences has a running time of $\mathcal{O}(n^y)$. Here n is the number of spatial outliers being analyzed and y is the number of time steps being analyzed. The method by Wu et al. is therefore not suited for data that cover larger time series.

The approach by Lu et al. (2004) [48] is similar to the approach by Wu et al. (2008) [65], as they first discover spatial outliers, and then track these through time. To discover spatial outliers, they first perform a wavelet transform of their data. A fuzzy edge detector is used to detect boundaries in the wavelet transformed

data. These boundary regions encapsulate potential outliers. The tracking through time was performed manually on specific outliers. The authors used this method to detect and track hurricane Isabel. An advantage of this method is, that the boundary classification performed by the fuzzy edge detector is relatively flexible. Where Wu et al. are restricted to rectangular regions, and Chawla and Sun to individual data points, Lu et al. are capable of finding arbitrarily shaped outlier regions. Their method however isn't suited for larger time series found in climate data, as each time slice would need to be checked by an edge classifier, and the tracking of outliers is performed manually.

4.2.2 Temporal Data

There exist many problems and approaches in the temporal outlier detection field, as can be seen in Gupta et al. (2013) [45]. Knowledge of these problems, solutions and algorithms can be beneficial when working with climate data. Climate data often has a spatial component, but Barz et al. [15] demonstrate how to turn a spatial-temporal problem into a temporal one. They use their Maximally Divergent Intervals (MDI) algorithm [13, 14, 15], which is a probabilistic method, to detect storms over the North Sea. They turn their spatial-temporal data into temporal data, by averaging over the spatial component of the data. They argue that the outliers they are trying to detect, large storms, will cover the entire spatial extent of their data. Their data covered the entire North Sea, which matched the spatial extend of large storms.

The MDI algorithm is also capable of finding temporal outlying intervals with their spatial location and spatial-temporal outlying intervals. To discover outliers, the probability distribution of tested intervals is compared to the probability distribution of the remaining data. Accounting for autocorrelations in data is achieved by using context embedding. By detecting intervals of outlying data points the algorithm is capable of discovering events that are unusual. The algorithm is designed for data on a regular grid.

4.3 Spatial-Temporal Data

Early work on detecting spatial-temporal outliers is presented by Kut et al. (2006) [66]. Their algorithm combines a cluster based approach with a density based approach to discover extreme wave heights in the Mediterranean Sea. The algorithm first detects spatial outliers. These are then checked to see if they are also temporal outliers. The discovered outliers are thus spatial-temporal outliers. In comparison, the algorithms by Wu et al. and Lu et al. discover spatial outliers that they track through time. The algorithm by Kut et al. doesn't take autocorrelations into account, which is desirable when working with climate data.

The Spatio-Temporal Behavioral Density-based Clustering of Applications with Noise (ST-BDBSCAN) algorithm introduced by Duggimpudi et al. (2019) [67] detects spatial-temporal outliers. The algorithm calculates a Spatio-Temporal Behavioral Outlier Factor (ST-BOF). This score is inspired by the LOF by Breunig et al. [60]. It uses the local spatial and temporal neighborhoods of data points to determine the outlierness of its non spatial-temporal attributes. This approach directly

discovers spatial-temporal outliers rather than first detecting spatial outliers, then temporal outliers like the approach by Kut et al. [66]. ST-BDBSCAN was used to track the path hurricane Katrina took, using buoy data. The ST-BDBSCAN algorithm contains no procedures to deal with autocorrelation in data which is a downside when working with climate data, and does not leverage the grid based nature of our data.

Rogers et al. (2009) [68] present a statistical approach to discover spatial-temporal outliers. They introduce the Strangeness-based Outlier Detection algorithm (StrOUD). It compares the strangeness of a data point to that of a baseline. The baseline needs to be defined by the user, and thus this algorithm requires prior knowledge of what constitutes normal behaviour. The algorithm is based on Transductive Confidence Machines [69]. These machines make the assumption that data is independent, thus this method isn't well suited for highly correlated climate data.

Cheng et al. (2006) [70] use semantic based classification to detect outliers. A clustering algorithm is used to classify data in predefined classes. Data points that have been assigned a different class in different time steps are potential outliers. Observations that change class when the spatial resolution is downscaled are also considered potential outliers. All potential outliers are then manually checked to see if they represent actual outliers. Chen et al. use a broad definition of spatial-temporal outliers, as they do not differentiate between spatial, temporal and spatial-temporal outliers. Their method requires manual checking and isn't suited for large data sets. This makes it unsuitable for our research.

An outlier detection framework that focuses on remote sensed data is presented by Liu et al. [71]. This framework combines outlier detection with data cleaning and feature extraction. Liu et al. consider an object an outlier if its non spatial-temporal features divert from its spatial or temporal neighbors. Thus spatial, temporal and spatial-temporal outliers are treated the same. The outliers are detected using a clustering algorithm. Individual outliers are than grouped to form outlying events based on their non spatial-temporal features. The framework was used to detect events on south pole temperature data, such as sensor failure, and extreme ice melt. Autocorrelation in the data isn't taken into account.

Learning the driving forces behind outliers is an important related task. It might provide insight in the behaviour and causes of anomalies. The paper by Jorge et al. [72] combines outlier detection with subgroup discovery. In this paper the authors look for outliers in social interactions on a playground using LOF. The LOF values are then used as target variables to discover subgroups, to investigate what might cause outlying behaviour.

4.4 Algorithm Selection

Based on this literature study, two algorithms are selected that are well suited for this research. The MDI algorithm by Barz et al. (2018) [15] is designed for data on a regular grid. It accounts for autocorrelations in data and works for spatial-

temporal data. It can find both temporal and spatial-temporal outliers. The SLOM algorithm by Chawla and Sun (2006) [16] compensates for both heteroscedasticity and autocorrelations and is designed for data on a regular grid. The SLOM algorithm is purely spatial, and will be used to investigate temporal outliers found by the MDI algorithm. The algorithms are explained in more detail in chapter 5.

5 Methodology

In order to find outlying events, two existing state-of-the-art outlier detection algorithms are utilized. Preprocessing and post-processing methods are added to the algorithms, to make them better suited for finding spatial-temporal outlier events relevant for energy grid security. In section 5.1 the Maximally Divergent Interval algorithm, introduced by Rodner et al. [13] and expanded by Barz et al. [14, 15], will be discussed. This algorithm will be used to detect temporal outliers. Section 5.2 provides an overview of the preprocessing steps performed on the data before the MDI algorithm is used. The second algorithm, discussed in section 5.3, is based on SLOM scores, introduced by Chawla and Sun [16]. This algorithm detects spatial outliers and will be used to investigate the temporal outliers detected by the MDI algorithm. The original algorithm is extended to group SLOM outliers, and track them through time. Details regarding the implementation of the algorithms are discussed in section 5.4.

5.1 The Maximally Divergent Intervals Algorithm

The Maximally Divergent Intervals (MDI) algorithm was first introduced by Rodner et al. (2016) [13]. The algorithm was then improved and expanded upon by Barz et al. [14, 15]. The algorithm is based on defining outliers as statistically unlikely events. In the algorithm, the statistical distribution, or most likely estimate, of a potential outlier interval is compared to the distribution on the rest of the data. If it is unlikely that both are generated by the same distribution, the interval is considered an outlying interval. The initial MDI algorithm by Rodner et al. [13] focused on finding outliers in temporal data sets. In the later works by Barz et al. [14, 15] the algorithm was generalized to find outliers in spatial-temporal data.

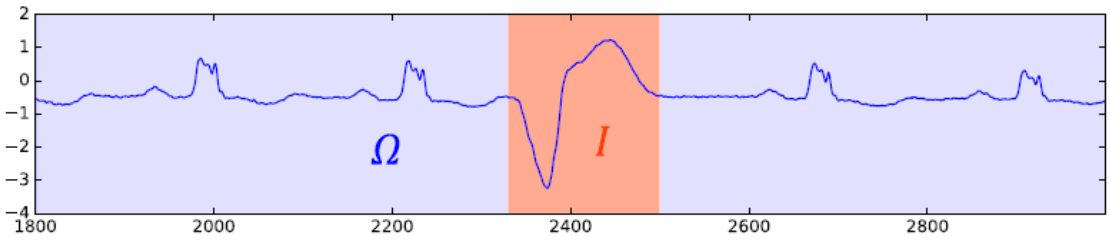


Figure 3: Example of an interval I, whose distribution will be compared to the remaining data Ω. Example taken from Barz et al. (2018) [15].

The MDI algorithm finds outliers by comparing intervals to the remaining data. If it is statistically unlikely that both are created using the same generative process, the interval is considered an outlier. In the algorithm all partial tensors within some user defined minimum and maximum size parameter values are compared to the remaining data set. An example of an interval in temporal data is depicted in figure

3. In this example the (estimated) distribution of interval I, would be compared to the (estimated) distribution of all remaining data Ω.

A sample use case, hurricane detection, is depicted in figure 4. In this example outliers were detected on temporal data from a weather buoy near the Bahama's from June 2012 to November 2012. The dataset contains significant wave height, which is the average height of the highest one-third of the waves, sea level pressure and wind speed data. The largest outliers detected coincide with hurricanes that passed by the weather buoy.

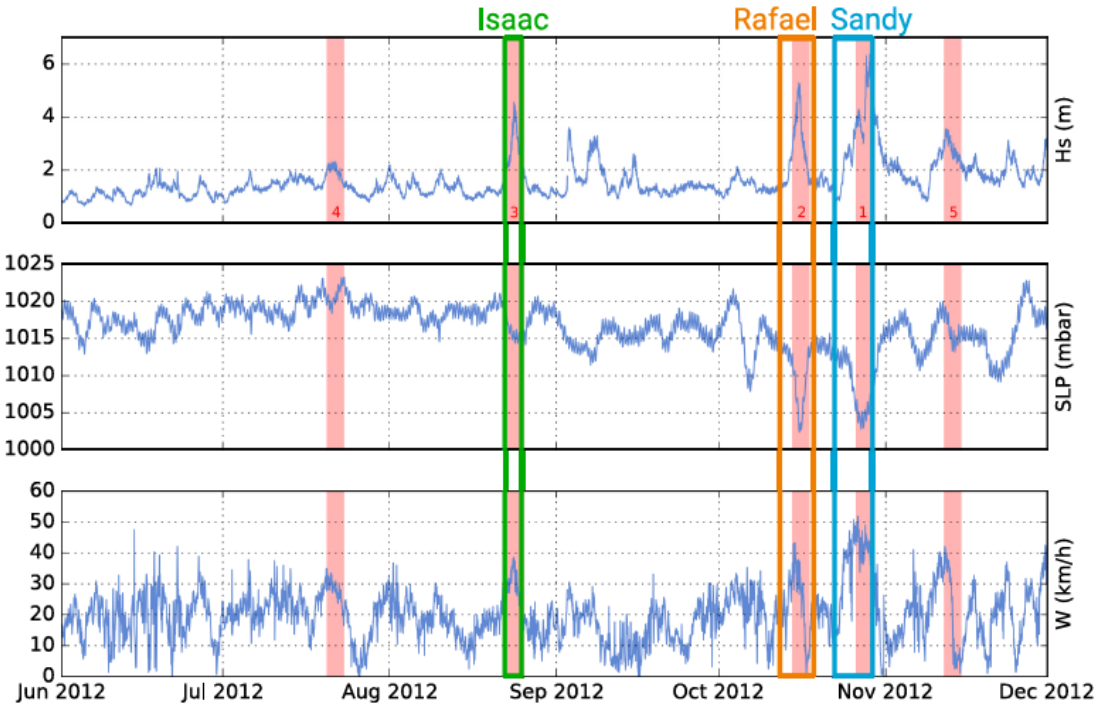


Figure 4: Outliers detected by the MDI algorithm in temporal data. Data measured near the Bahama's, from June 2012 to November 2012. Contains significant wave height (Hs), sea level pressure (SLP) and wind speed (W). Top 5 outliers are shown. The top 3 match hurricanes that passed the measuring location. Image taken from Barz et al. (2017) [14].

In section 5.1.1 the details of distribution estimation are discussed. The scoring of interval outlierness is discussed in section 5.1.2. This is followed by a discussion on the complexity of the algorithm in section 5.1.3. The context embedding procedure that deals with the autocorrelation in the data is discussed in section 5.1.4. A heuristic used to speed up calculations, based on Hotteling's T^2 score is discussed in section 5.1.5.

5.1.1 Distribution Estimation

For the MDI algorithm to work, distributions need to be compared. This can potentially be known distributions, but more often than not these need to be estimated. The choice of distribution to fit to the data has significant impact on the detected outliers. By using a certain distribution to estimate the data, assumptions are made about the data.

Since the MDI algorithm estimates a distribution for every possible interval in the data, one important concern is the complexity of the distribution, and its efficient calculation. It is crucial to use a distribution that can be calculated and updated efficiently, especially given the size of our data. Thus, like the original paper, we limited the potential distributions to two models that are computationally feasible. These are Kernel Density Estimation (KDE) using a Gaussian kernel, and multivariate normal distributions, also known as Gaussian distributions. The probability density function of the multivariate normal distribution is given in equation 1:

$$p(\mathbf{x}) = \frac{exp(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu))}{\sqrt{(2\pi)^n|\mathbf{\Sigma}|}} \tag{1}$$

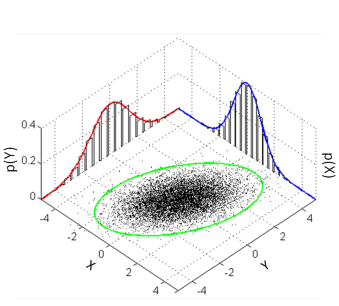
Here $\mathbf{x}$ is an n-dimensional column vector, $\mathbf{\Sigma}$ is the covariance matrix, $|\mathbf{\Sigma}|$ the determinant of $\mathbf{\Sigma}$ and μ the n-dimensional mean vector.

The Kernel Density Estimation of an n-dimensional data point $\mathbf{x}$ is given by equation 2. It is calculated as a summation of kernel functions K_h. A Gaussian kernel function will be used, given by equation 3.

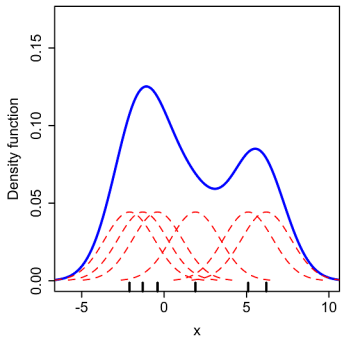
$$p(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} K_h(\mathbf{x}, \mathbf{x}_i) \tag{2}$$

$$K_h(\mathbf{x}, \mathbf{x}_i) = (2\pi h^2)^{-\frac{n}{2}} e^{-\frac{\|\mathbf{x}-\mathbf{x_i}\|^2}{h^2}} \tag{3}$$

Here N is the number of samples drawn from the data to estimate the probability density function p. The dimension of the data, in our case the number of energy variables, is indicated by n. The width of the Gaussian kernel function is indicated by h, which is the standard deviation of the Gaussian distribution. Examples of the probability density functions can be seen in figure 5.



(a) Multivariate Normal Distribution With 2 Variables



(b) Univariate Kernel Density Estimation Using Gaussian Kernels

Figure 5: Examples of the multivariate normal distribution and univariate Kernel Density Estimation using Gaussian Kernels. Images taken from Wikipedia.

5.1.2 Divergence Measures

In order to compare probability distributions a divergence measure $\mathcal{D}(p_I, p_\Omega)$ is used. This measure indicates the discrepancy between the probability distribution of the investigated interval I and the distribution of the remaining data Ω. The intervals with the highest divergence values $\mathcal{D}$ are reported as outliers. The authors propose several divergence measures: cross entropy, Kullback-Leibler divergence and the Jensen-Shannon divergence. We will discuss each in the following paragraphs.

Cross Entropy

Cross entropy is based on information theory and can be used as a divergence measure. Consider two probability distributions, p and q. Assume a sample is drawn from q, the cross entropy of p and q is given by:

$$\mathcal{D}_{CE} := H(p, q) = \mathbf{E}_p[-log\ q] \tag{4}$$

Where $\mathbf{E}_p$ is the expected value with respect to distribution p. $H(p, q)$ indicates how surprising a sample drawn from q is, if it were drawn from p. The cross entropy is a relatively simple divergence measure, and can be approximated from data:

$$\mathcal{D}_{CE}(I, \Omega) \approx \frac{1}{|I|} \sum_{i \in I} \log p_\Omega(\mathbf{x}_i) \tag{5}$$

It should be noted that this approximation only requires the estimation of one probability density. Let I be the interval being checked, and Ω the remaining data, only the probability density function of Ω, p_Ω needs to be estimated. This could be advantageous when working with small interval sizes, as these are difficult to accurately estimate due to the lack of data points.

Kullback-Leibler

The Kullback-Leibler divergence is similar to cross entropy. In addition to how well a sample drawn from p is explained by q, the intrinsic entropy of p is taken into account. The idea is that a stable interval receives a higher score than a highly variable interval. Thus an interval that is divergent but stable is considered the more significant outlier. Outlying behaviour caused by intrinsic variability is thus penalized. Equation 6 shows the Kullback-Leibler divergence. For clarity the same notation as in the original paper is used, however a typo in equation 6 was corrected.

$$\mathcal{D}_{KL}(p, q) := H(p, q) - H(p, p) = \mathbf{E}_p[log\frac{p}{q}] \tag{6}$$

The $H(p, p)$ term is the intrinsic entropy of p. So probability distributions with high intrinsic entropy, thus high variability, receive lower divergence scores. $\mathcal{D}_{KL}(I, \Omega)$ can be estimated empirically, using both the interval distribution p_I and distribution p_Ω of the remaining data:

$$\mathcal{D}_{KL}(I, \Omega) \approx \frac{1}{|I|} \sum_{i \in I} \log p_I(\mathbf{x}_i) - \log p_\Omega(\mathbf{x}_i) \tag{7}$$

The Kullback-Leibler divergence is not a symmetrical measure. Barz et al. investigated other versions of the Kullback-Leibler divergence such as $\mathcal{D}_{KL}(p_\Omega, p_I)$ rather than $\mathcal{D}_{KL}(p_I, p_\Omega)$. A symmetrical combination of both was also considered:

$$\mathcal{D}_{KLsym}(p, q) := \frac{1}{2}\mathcal{D}_{KL}(p, q) + \frac{1}{2}\mathcal{D}_{KL}(q, p) \tag{8}$$

The original authors found that the calculation of $\mathcal{D}_{KL}(p_\Omega, p_I)$ was dominated by the variance of the interval. Since smaller intervals are more likely to have low variance, there existed an inherit bias that caused poor results where the Divergence measure blew up as the variance approached zero. The $\mathcal{D}_{KL}(p_\Omega, p_I)$ also dominates the calculation of the symmetrical measure, thus it suffers from the same problem.

The authors noted that the best performing Kullback-Leibler divergence $\mathcal{D}_{KL}(p_I, p_\Omega)$, suffers from a bias towards small intervals, as can be seen in figure 6. To adjust for the bias towards smaller intervals, the authors proposed an unbiased Kullback-Leibler divergence:

$$\mathcal{D}_{U-KL}(p_I, p_\Omega) := 2 \cdot |I| \cdot \mathcal{D}_{KL}(p_I, p_\Omega) \tag{9}$$

The theoretical derivation of the factor $2 \cdot |I|$ can be found in Barz et al. (2018) [15].

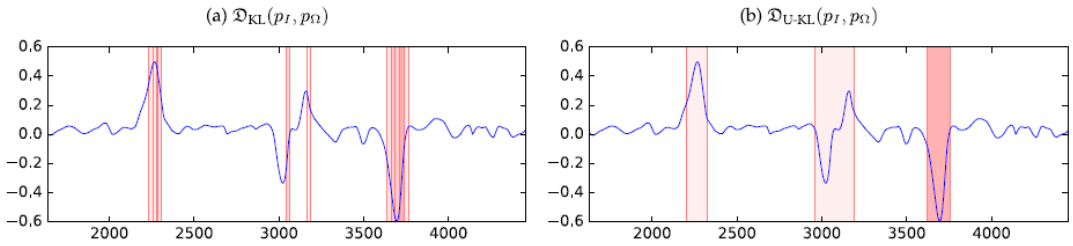


Figure 6: High divergence regions with the regular Kullback-Leibler divergence (a), compared to high divergence regions calculated using the unbiased Kullback-Leibler divergence proposed in [14]. Note the smaller interval sizes in a), compared to b). Image taken from Barz et al. (2018) [15].

Jensen-Shannon

The final divergence measure used by the authors is the Jensen-Shannon divergence. This measure uses mixture distributions, where it is equally likely to draw a sample from either p or q. These are combined with the Kullback-Leibler divergence:

$$\mathcal{D}_{JS}(p, q) = \frac{1}{2}\mathcal{D}_{KL}\left(p, \frac{p+q}{2}\right) + \frac{1}{2}\mathcal{D}_{KL}\left(q, \frac{p+q}{2}\right) \tag{10}$$

The Jensen-Shannon divergence can be approximated by combining approximations of the Kullback-Leibler divergences.

What divergence measure provides the best solutions isn't known a priori and depends on factors such as the underlying distribution of the data and computational complexity.

5.1.3 Computational Complexity

Since we are working with rather large datasets, the computational complexity of the algorithm is important. There are two factors that contribute to the complexity of the MDI algorithm. The estimation of the probability density function and the calculation of the divergence score.

Consider a straight forward approach, let N denote the number of data points and L denote the maximum size of an interval. So for data with a temporal, latitude and longitude component $L = (max_t - min_t + 1) \cdot (max_{lat} - min_{lat} + 1) \cdot (max_{lon} - min_{lon} + 1)$. Here $max_{t/lat/lon}$ and $min_{t/lat/lon}$ indicate the maximum and minimum allowed temporal duration and spatial length of an interval I. These are user defined parameter values.

There are $\mathcal{O}(N \cdot L)$ intervals to check. For the Gaussian model, each interval requires a summation over $\mathcal{O}(N)$ data points to calculate the mean and covariance matrices. For the KDE model each interval requires the summation of $\mathcal{O}(N)$ kernel functions. As we have seen in the previous section, each divergence measure can be approximated summing over the interval. This results in a computational complexity of $\mathcal{O}(N \cdot L(N + L))$ for the Gaussian model and $\mathcal{O}(N^2 \cdot L^2)$ for KDE using Gaussian kernels.

This simple approach is not very efficient. It often sums the same data. Clever use of cumulative sums using the integral image concept from Viola et al. (2004) [73] reduces the complexity. First calculating matrices containing cumulative sums reduces the complexity to $\mathcal{O}(N \cdot L^2)$ for the Gaussian model and $\mathcal{O}(N^2 + N \cdot L^2)$ for the KDE model.

The Kullback-Leibler divergence has a closed form solution when comparing multivariate Gaussian distributions [74]. This means that the divergence can be calculated in constant time, when using the Gaussian model combined with Kullback-Leibler or cross entropy, since it can be derived from that closed form, as a divergence measure. The closed form solution of the Kullback-Leibler and cross entropy divergence measures can be seen in equation 11 and equation 12 respectively.

$$\mathcal{D}_{KL}(p_I, p_\Omega) = \frac{1}{2}\big((\hat{\mu}_\Omega - \hat{\mu}_I)^T S_\Omega^{-1}(\hat{\mu}_\Omega - \hat{\mu}_I) + trace(S_\Omega^{-1} S_I) + \log \frac{|S_\Omega|}{|S_I|} - n\big) \quad (11)$$

$$\mathcal{D}_{CE}(p_I, p_\Omega) = \frac{1}{2}\big((trace(S_\Omega^{-1} S_I) + \log |S_\Omega| + n \cdot 2\log(2\pi) + (\hat{\mu}_\Omega - \hat{\mu}_I)^T S_\Omega^{-1}(\hat{\mu}_\Omega - \hat{\mu}_I)\big) \quad (12)$$

In these equation S denotes the estimated covariance matrix, $\hat{\mu}$ the estimated average and n the number of variables. The trace of a matrix is the sum of all its diagonal components. Using these closed form solutions reduces the complexity to $\mathcal{O}(N \cdot L)$, the number of intervals.

This means that for experiments we perform on large parts of our dataset we are forced to use the Gaussian model with either cross entropy or the (unbiased) Kullback-Leibler divergence. Since we want to compare the results of the different experiments with each other, and the Jensen-Shannon divergence measure lacks a

closed form solution, we will use both cross entropy and unbiased Kullback-Leibler as divergence measures.

5.1.4 Context Embedding

In order to incorporate the fact that climate data is highly correlated, both spatial and temporal context are used to ensure that these correlations are accounted for. This context is added, by so called delay embedding [75]. The embedding changes data points to vectors in phase space by adding the data point, together with nearby data points to a vector.

For temporal data the added data points are data points from previous time steps, that are relatively close, as can be seen in figure 7. Spatial embedding uses neighboring data points, this is demonstrated in figure 8.

These resulting vectors are no longer autocorrelated. This is achieved by ensuring that the embedded points are far enough apart to not only contain autocorrelation, and not so far that they are not correlated at all.

This embedding is a preprocessing step of the algorithm. The number of data points to add as context, as well as the time lag or distance to neighbors are parameter values of the algorithm. The context is added as attributes to the original data point.

For temporal context embedding K-1 points will be added, resulting in vectors of size K. Each of these points will be T time apart. Similarly for spatial context embedding K_x-1 and K_y-1 points will be added. These will be D_x and D_y apart.

So, for example, consider a data point $\mathbf{x}$ with n variables at position (t, lat, lon). If we were to use temporal context embedding on $\mathbf{x}$ with values T and K, the resulting tensor would be extended along the attribute dimension. So $\mathbf{x}$ would be at position (t, lat, lon) with $n * K$ attributes. The attributes would be the values of the n attributes, at times $(t - T, t - 2T, ..., t - (K - 1)T)$. These attribute vectors are what will be used to estimate the distributions, and calculate the divergence.

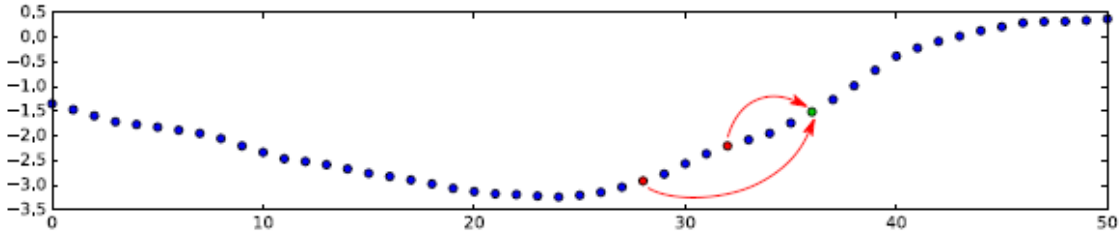


Figure 7: Time embedding example. The attributes of the red data points are added to the green data point's attribute vector. Here K=3 and T =4. Image taken from Barz et al. (2017) [14].

Trial runs were performed to determine what parameter values to use for our final experiments. First the use of both spatial and temporal context was examined. Using these on spatial-temporal data would enable the MDI algorithm to find spatial-temporal outliers rather than spatial or temporal outliers.

An experiment was performed with minimal spatial context embedding settings: $K_x = 2$, $K_y = 2$, $D_x = 3$,$D_y = 3$ and minimal temporal context embedding: K=2,

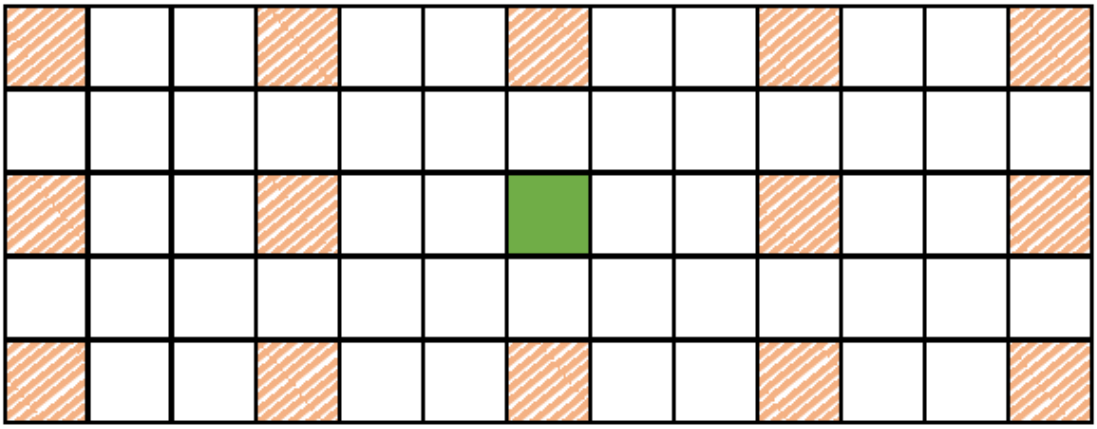


Figure 8: Spatial embedding example. The attributes of the red striped data points are added to the green data point's attribute vector. Here $K_x = 3$, $K_y = 2$, $D_x = 3$ and $D_y = 2$. Image taken from Barz et al. (2017) [14].

T=1. For this experiment we looked at a very small subset of our data. The spatial resolution was reduced from 0.25 to 0.5 degree, which results in an 80 by 110 grid. The temporal resolution was 1 hour. Rather than the 70 years of available data, 96 hours were looked at.

This small sample used over 20GB of memory. Even with 125 GB of available RAM memory on our system, we found that using spatial context embedding isn't feasible for this specific application due to the size of the dataset.

The direct result is that the MDI algorithm will not be used to find spatial-temporal outliers. If used to analyze spatial-temporal data without spatial context embedding, temporal outliers and their spatial location will be discovered.

In order to accurately discover temporal outliers, the temporal context embedding parameters need to be investigated. The idea behind the temporal context embedding is to pick points that are correlated but not points so far that they are not correlated at all. To discover the correlation length of our data we looked at temporal data derived from the original data. In order to create a time series, we summed the energy generation of the Western Europe region as defined in section 3.1.

To investigate the autocorrelation length on this time series we used the partial autocorrelation plot provided by the statsmodels Python package [76]. The results of the partial auto correlation plots can be seen in figure 9. The correlation coefficients were calculated per variable.

Based on these plots we have decided to use K=4 and T=8 as temporal embedding settings as these capture most of the autocorrelation. These settings ensure that the correlations in solar photovoltaic and wind-onshore energy generation at the larger lags of approximately 24 hours are accounted for. These settings also ensure that at least one day and night cycle is embedded as context, which has a big impact on climate data, Solar Photovoltaic energy generation in particular.

We assume that this temporal series captures the auto correlated behaviour of our data, and will be using these settings for all further experiments.

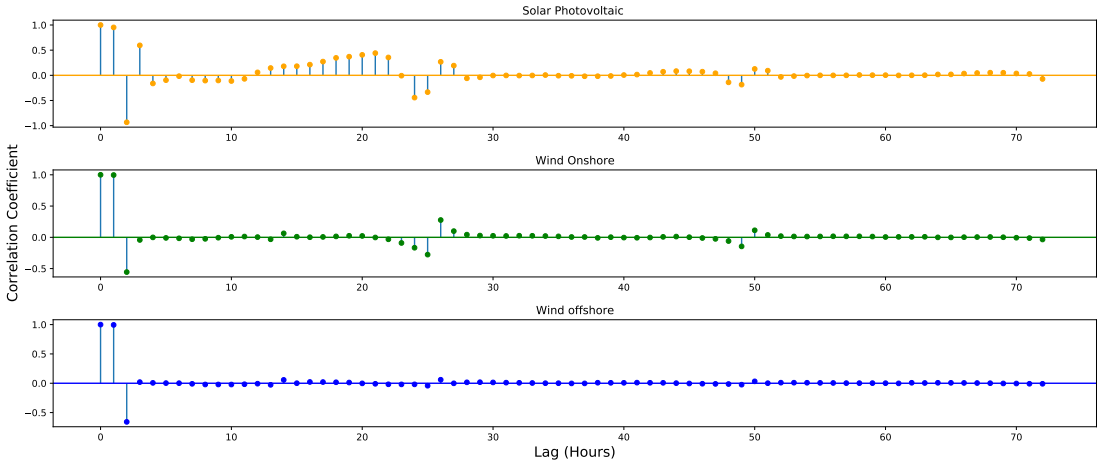


Figure 9: Partial auto correlation plots created using statsmodels. Depicts the correlation coefficient of the data with itself at different time lags. Yule-Walker with sample size adjustment method used.

5.1.5 Hotteling's T^2 Proposal method

The closed form complexity of the MDI algorithm, $\mathcal{O}(N \cdot L)$, is not optimal for large data sets. To make the algorithm better suited for large data sets, an interval proposal method is provided. This proposal method will propose intervals that are likely to contain outliers. The proposal method ensures that not every possible interval in the data has to be checked, which can be unfeasible for large data sets.

The idea behind the interval proposals is that an outlier interval would contain several data points that would receive high scores when considering point wise outlier detection. One such point wise scoring method is Hotteling's T^2 score [77], which can be seen as the multivariate generalization of Student's T-test.

An outlying interval would thus see high point wise scores. Notably at the start and end of an outlying interval, respectively an increase and decrease of the point wise scores is expected. This is the core principle of the outlying interval proposal method. Only intervals that start and end with data points whose Hotteling's T^2 score gradient magnitude surpass some threshold are considered. This can achieve a significant speedup as this potentially reduces the number of intervals to check significantly, and the distribution checking and divergence calculation are the most time consuming processes.

There is no a priori knowledge of the number of intervals that will be proposed using this method. Worst case all intervals satisfy the conditions thus the theoretical complexity remains $\mathcal{O}(N \cdot L)$. In order to determine speedup we could achieve using this heuristic, small experiments were performed.

To check the method we used the MDI algorithm to detect outliers in ERA5 temperature at 2 meters data. We used data from latitude 35 to 74.75 and longitude -14.75 to 40. A 3 hourly temporal resolution was used, and 0.25 degree spatial resolution. We used the Gaussian model with unbiased KL divergence. The minimum

length of an interval was 24 hours. The maximum length we varied. No spatial context embedding was used. For temporal context embedding we use K=2 and T=1 as parameter values. Intervals that start and end with a Hotteling's T^2 gradient magnitude larger than the average gradient + 1.5 times the standard deviation of the gradient magnitude are proposed.

Analyzed Years	All Intervals (s)	Hotteling (s)	Max interval length (h)
1987	829	16	72 (3 days)
1987	1904	30	168 (1 week)
1987	20660	284	2160 (90 days)
2000-2001	43795	31	72 (3 days)

Table 1: Table with running times of the MDI algorithm for sample data. An Intel Xeon Gold 6130 CPU with 16 dual-cores at 2.1 GHz clock speed was used. Reported time is wall clock time in seconds, rounded down.

We note a significant speedup due to the interval proposal. This finding is in line with the original work by Barz et al. where speedups from 4 hours to 5 seconds and 13 hours to 5 minutes are mentioned. Relatively few of these experiments were performed as the initial results were convincing enough of the time improvement caused by the proposal method.

The potential drawback of the interval proposal method is that not all intervals are considered. One would expect this causes a loss of accuracy. However the original authors noted an increase in accuracy when using the proposal method. They suspect that this interval proposal method prevents intervals that are potentially confusing for the algorithm to be considered.

Since the interval proposal method increases accuracy and improves computation time we use the interval proposal method for all our experiments.

5.1.6 Non maximum suppression

In order to ensure that the top detected outliers aren't all small variations of the same event non maximum suppression is used. Starting with the top outlier the intersection with other outliers intervals divided by the union of the intervals is checked. If this is larger than 0.5 the intervals are considered overlapping, and only the interval with the highest score is reported.

After discovering temporal outliers, the outlying time steps will be investigated further using the SLOM algorithm, which is capable of finding spatial outliers.

5.1.7 Temporally Driven Approach

To summarize, the MDI algorithm will be used to detect temporal outliers. These will highlight times that will be investigated with the SLOM algorithm to detect any potential spatial outliers. We have to split the process up, rather then using the MDI algorithm to detect spatial-temporal outliers as this is not feasible for our data.

The reason we look for temporal outliers first and see if they are spatial-temporal outliers second rather than spatial outliers first and temporal second, is that in the context of energy generation temporal outliers are more interesting than spatial outliers. The European energy system is a very connected and complicated system, as we have seen in section 2.1. Temporal outliers can influence the entire system, while localized spatial outliers can be more easily contained and addressed. Thus it is better to find temporal outliers that are not spatial outliers than spatial outliers that are not temporal outliers.

5.2 Preprocessing

Before we can input our data into the MDI algorithm, several steps have to be performed to prepare the data. These have to do with seasonal effects in our data as well as highlighting potentially interesting regions.

5.2.1 Seasonality

Since the MDI algorithm is used to investigate temporal data, the seasonality of the data needs to be investigated. The seasons potentially cause patterns in the data that might influence the results. That the seasons impact our data can be seen in figure 10. The presence of seasonal effects in energy generation data is in line with findings by van der Wiel et al. [78].

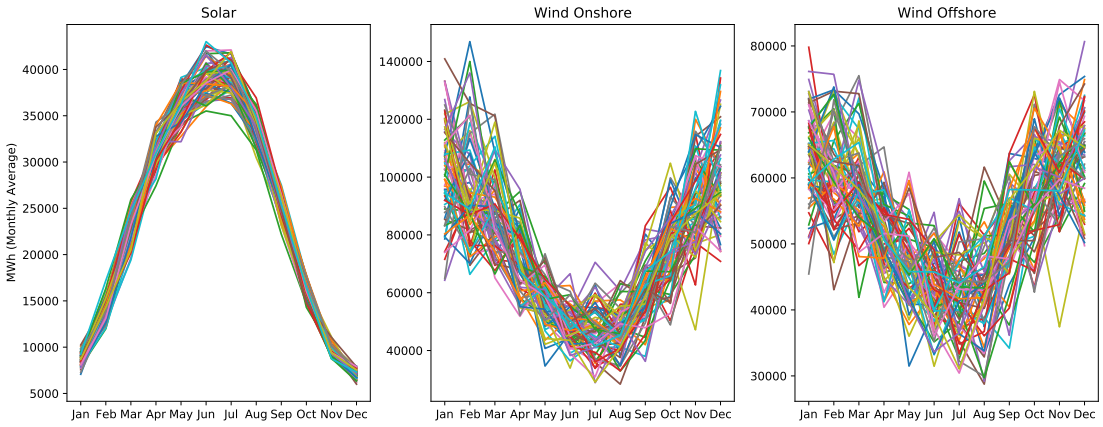


Figure 10: Average Energy Capacity per Month of the Sum over the entire spatial Region. Each line represents a year from 1950 to 2020

Both the data with and without seasonal effects might contain interesting outliers. In order to investigate if the MDI algorithm detects more interesting outliers without the seasonal effects, normalization via Z score was used. The Z score denotes the number of standard deviations a value diverts from the mean. The Z score of a data point is calculated using equation 13:

$$Z = \frac{\mathbf{X} - \mu}{\sigma} \tag{13}$$

Here μ is the mean and σ the standard deviation.

The idea of using this score is that it is an indication of deviation from normal behaviour. Since the normal behaviour contains the seasonality, we can potentially discover outlying behaviour considering the time of year.

To calculate the Z score we used the unique hour of the year. For example, to calculate the Z score of a data point on August 16th - 14:00 the average and standard deviation of all August 16th -14:00 over 70 years were used. This was done to ensure there were no more seasonal effects. More conventional implementations such as the average over a week or month still showed seasonal behavior for Solar Photovoltaics. This is because in the summer SPV reaches higher values then in the winter. SPV however doesn't have a significantly higher standard deviation. Combined with the fact that there is no SPV at night, seasonal patterns emerged in the negative Z scores. The nights received lower Z scores during the summer. When using unique hours this problem doesn't exist, as the night receives its Z score based on normal behaviour of the night.

Because there exist a few cases where in 70 years of data, there is a single hour with a very small amount of generation rather than no generation at all, these will have blown up Z scores. These would then quickly get tagged as outliers. Since the magnitude of these outliers is extremely low these are of no interest for this research. This only happened for SPV, for some morning and evening hours. To prevent Z scores blowing up we have assigned all SPV data points with a $\sigma < 10$ MWh a Z score of 0. Since we have insufficient data points for the 29th of February, these data points have been removed from the Z score data.

The monthly average of the Z score per energy source can be seen in figure 11. No clear seasonal patterns can be discovered.

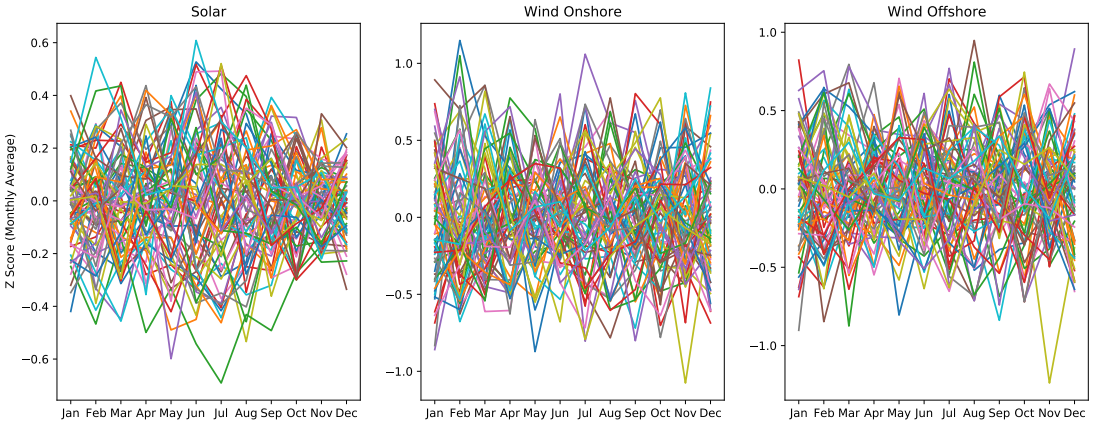


Figure 11: Average Energy Capacity Z Scores per Month of the Sum over the entire spatial Region. Each line represents a year from 1950 to 2020. February 29th has been removed from the data. Z score calculated per unique hour of the year.

5.2.2 Local and Global Outliers

The MDI algorithm finds global outliers. These outliers thus are unusual in comparison to the entire remaining distribution. In contrast, local outliers are unusual when considering the local distribution. For this research we are interested in both local

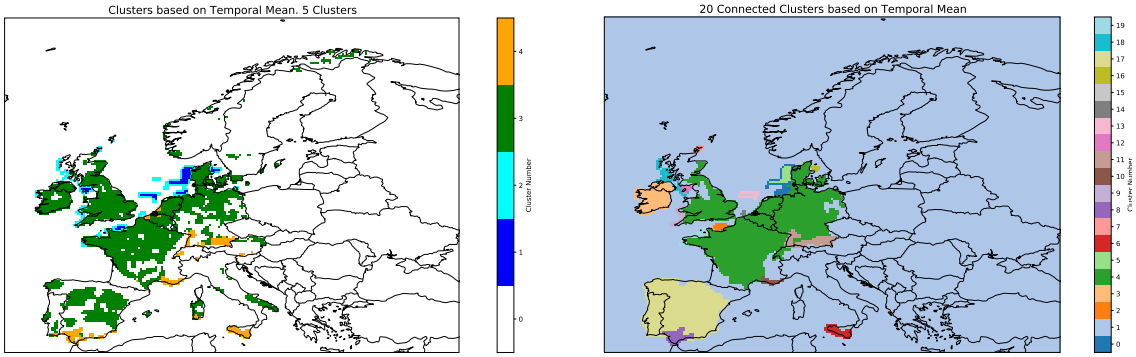
and global outliers. By only using specific regions as input for the MDI algorithm, it is possible to focus where the algorithm will detect outliers.

It should be noted that such outliers aren't truly local outliers. They are global outliers, in smaller regions. In order to be considered local outliers, intervals need to divert from their local regions. With the MDI algorithm, there might not be enough data points to properly estimate probability distributions to discover truly local outliers. These global outliers in smaller regions might still be interesting. In order to select such regions we use clustering.

In order to highlight related regions to investigate further the 70 year temporal mean of the data was clustered. For each grid cell the average energy generation over 70 years was calculated. These averages were then clustered. For the clustering we used two models, k-means clustering and agglomerative clustering.

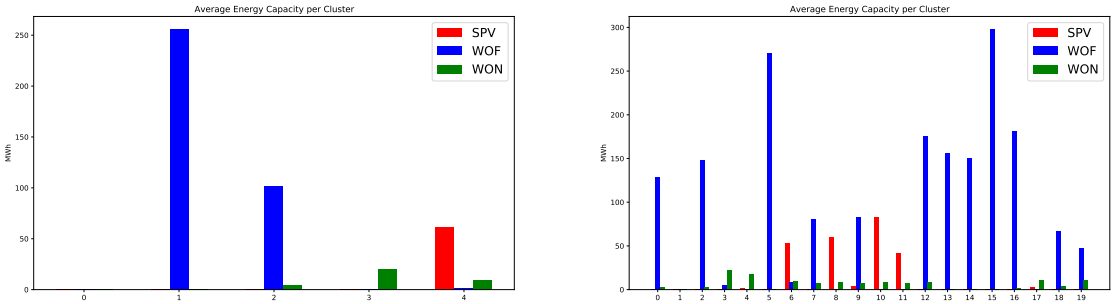
For k-means clustering we used Elkan's algorithm [79] using 5 clusters. The choice for 5 clusters was made after some initial testing with cluster numbers. We found that 5 clusters provides a good division of the energy sources in the clusters. It was implemented using the sklearn package for Python [80]. The resulting clusters can be seen in figure 12a. The properties of each cluster can be seen in figure 12c, where the average energy capacity per source per cluster is shown.

In order to ensure spatially coherent clustering we used agglomerative clustering. This clustering method can require regions to be connected. This clustering was performed using the agglomerative clustering algorithm, using ward linkage [81] and queen connectivity constraint. This means that data points that are diagonally adjacent are considered adjacent. As there are many similar regions, that are spatially distant a large number of clusters was required. We used 20 clusters. The resulting clustering can be seen in figure 12b. The per cluster averages are depicted in figure 12d. The agglomerative clustering was performed using the sklearn package for Python [80].



(a) K means clustering

(b) Agglomerative clustering



(c) K-means cluster attributes

(d) Agglomerative clustering attributes

Figure 12: Clusters and their attributes. Clusters based on temporal mean of SPV, WON and WOF data. Figure 12a shows the result of K means clustering with 5 clusters and figure 12b shows the result of agglomerative clustering using 20 clusters with ward linkage and queen connectivity. Figures 12c and 12d show the average energy capacity per source per cluster in MWh.

Based on both clustering models we make several observations. We see large connected offshore wind farms in the North Sea and smaller offshore wind farms on the French and Irish coasts. Large spatially coherent regions that are dominated by solar photovoltaic energy generation are present in southern Spain, Sicily, southern France and southern Germany. The remainder of mainland Europe and the United Kingdom are dominated by onshore wind.

Based on these observations, several regions were selected to perform further experiments with. These regions represent use cases for the algorithm. The selected regions can be seen in figure 13 and their properties can be seen in table 2. The selected regions covering Germany and The Netherlands, Spain and the North Sea represent wind-onshore, solar photovoltaic and wind-offshore use cases for the algorithm. The region covering France represents a smaller mixed region. It was selected as France is an important hub in the energy grid.

The MDI algorithm will be used to detect temporal outliers in the entire Western European region and the regions in table 2. The spatial-temporal data will be transformed into temporal data by summing over the entire region being investigated.

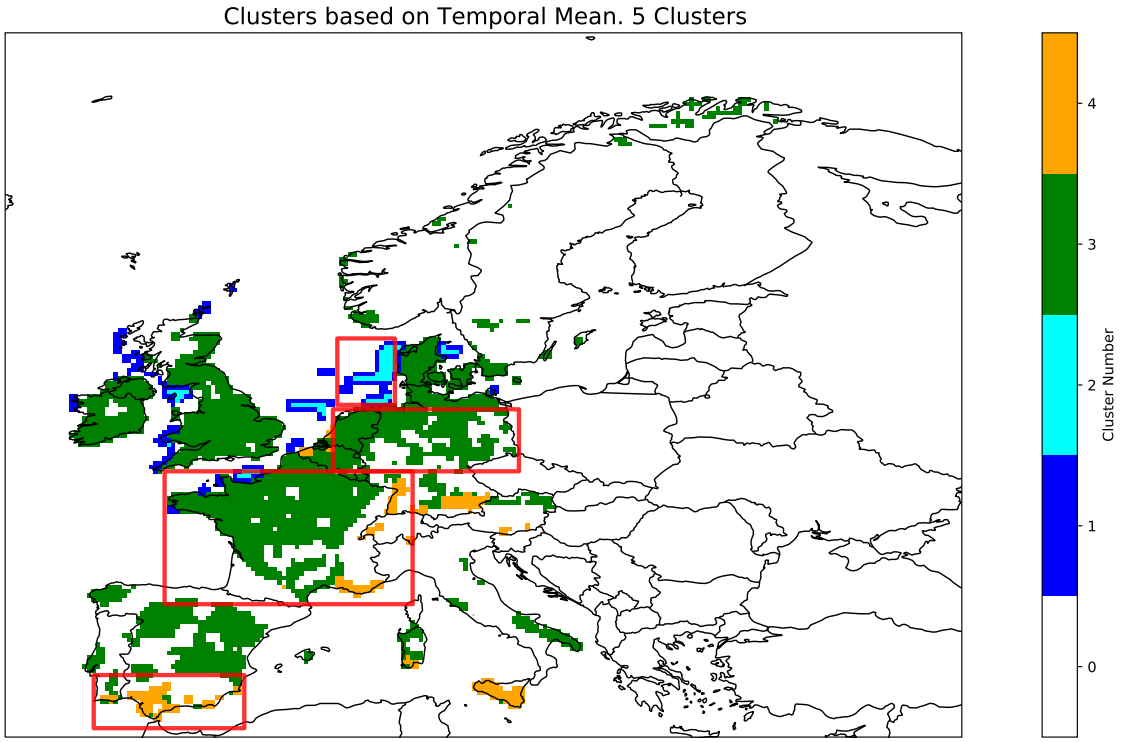


Figure 13: K means Clusters based on Temporal Means. In addition the geographic regions that represent use cases are highlighted.

Region Name	Energy Type	Latitude Bounds	Longitude Bounds
Mainland Onshore	Wind-Onshore	50 : 53.5	4.5 : 15
North Sea Offshore	Wind-Offshore	53.75 : 57.5	4.75 : 8
Spanish SPV	Solar Photovoltaic	35.5 : 38.5	-9 : -0.5
French Mixed	Mixed	42.5 : 50	-5 : 9

Table 2: Table containing information on use case regions selected based on clustering of the temporal mean of ERA5 energy capacity data.

5.3 SLOM

Since our data is derived from climate simulation data, which is both autocorrelated and heteroscedastic, the work by Chawla and Sun (2006) [16] is used. Their spatial local outlier measure (SLOM) compensates for both (spatial) autocorrelation in the data and (spatial) heteroscedasticity. Thus it is well suited for finding spatial outliers in our data. The goal of using the SLOM algorithm is to gain a better understanding of the temporal outliers discovered by the MDI algorithm. Even though the original paper by Chawla and Sun is from 2006, their algorithm is still considered a state-of-the-art spatial outlier detection algorithm [82, 83, 84].

First the algorithm as presented in the original paper is discussed. This is followed by a discussion on the post-processing steps we added to track SLOM outliers through time, and combine single SLOM outliers into SLOM events.

5.3.1 The SLOM algorithm

The original algorithm by Chawla and Sun (2006)[16] calculates a SLOM score for each data point by looking at its neighbors. The score of an object o is defined in equation 14. Before calculating the SLOM score, data is normalized such that non spatial attributes fall in the range [0,1].

$$SLOM(o) = \tilde{d}(o) * \beta(o) \tag{14}$$

Here $\tilde{d}(o)$ is the trimmed mean of the distance from object o to its spatial neighbors, and is calculated using equation 15.

$$\tilde{d}(o) = \frac{\sum_{p \in N(o)} dist(o,p) - maxd(o)}{|N(o)| - 1} \tag{15}$$

Here $N(o)$ is the set of neighboring data points of data point o, maxd is defined as $max\{dist(o,p)|p \in N(o)\}$. The largest distance in attribute values is removed, to compensate for the case where o is an outlier. If o is an outlier and a regular mean is used, then the distance from o to its neighbors will dominate the $\tilde{d}$ calculation of its neighbors.

Since $\tilde{d}(o)$ deals with spatial autocorrelation of outliers the goal of the β factor is to compensate for heteroscedasticity in the data. It does so by penalizing points that are surrounded by oscillating neighbors. The following pseudo-code shows how to calculate $\beta(o)$.

Algorithm 1: $\beta(o)$ calculation

```
   Result: β(o)
1  β(o) = 0
2  for each p ∈ N+(o) do
3  |  if d̃(p) > avg(N+(o)) then
4  |  |  β(o) = β(o) + 1
5  |  else if d̃(p) < avg(N+(o)) then
6  |  |  β(o) = β(o) − 1
7  end
8  β(o) = |β(o)|
9  β(o) = max(β(o),1) / (|N+(o)|−2)
10 β(o) = β(o) / (1+avg(d̃(p)|p∈N(o)))
```

Here $N_+(o)$ denotes the set of all objects in the neighborhood of o, with o itself included, $avg(N_+(o))$ is the average $\tilde{d}(p)$ of all $p \in N_+(o)$.

Intuitively the $\beta(o)$ factor can be seen as the number of times data points in $N_+(o)$ have a value larger than the average, minus the times values are lower than the average. Thus low β scores correspond with both high and low $\tilde{d}$ scores in the neighborhood. This indicates variability in the area. These high variability areas thus receive low β scores.

The steps at lines 9 and 10 of the algorithm aren't as intuitive as the others. Dividing by $|N(o)_+| - 2$ normalizes $\beta(o)$. The -2 in the denominator ensures that

an outlier o, surrounded by neighbors p with the same value for $\tilde{d}(p)$ has a β score of 1. This is desired, as it represent the case with the least variance and should receive the highest β score. The $max(\beta(o), 1)$ factor prevents β scores of 0, which would reduce the SLOM score to 0.

The last line of the β calculation allows for the differentiation of cases where the same number of $\tilde{d}(p)$ values are higher and lower than the average. For example in figure 14 the outliers would have the same β score without the last line of the algorithm. The last line ensures that figure 14 a) receives a higher β score, as it has a lower average trimmed mean in its neighborhood.

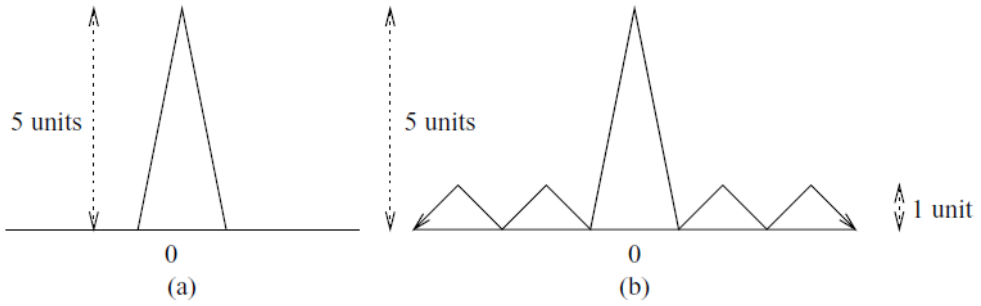


Figure 14: Example of a situation where (a) and (b) have the same outlier o and net count of values of $\tilde{d}(p)|p \in N_+(o)$ higher or lower than the average. The variance around (b) reduces its β value. Image taken from Chawla and Sun (2006) [16].

The resulting SLOM score, calculated using equation 14, is calculated for all data points. The SLOM score of an object o lies within the range $(0, \sqrt{n})$, where n is the number of non spatial variables. All data points above some user defined threshold, or the top-k SLOM score data points are considered to be potential outliers. The resulting high SLOM score outliers are local outliers, as data points are only compared with their neighbors.

In order to calculate the SLOM score a distance measure is required. There are multiple viable possibilities such as the Manhattan distance, Mahalanobis distance and Euclidean distance. We will use the Euclidean distance, as the original authors did.

The SLOM algorithm will be used to investigate the temporal ranges that are outliers according to the MDI algorithm. The SLOM algorithm will be used on the original spatial-temporal energy generation data. The SLOM scores are then merged to form SLOM events.

Computational Complexity

For the algorithm to work, the data needs to be normalized. This requires a pass over the data if information about the average, minimum and maximum value are available, or multiple passes if such information is not available. This costs $\mathcal{O}(N \cdot n)$, where N is the number of data points, and n the number of variables. To calculate $\tilde{d}(o)$ a pass over all data points is required, which costs $\mathcal{O}(N)$. The same holds for the

calculation of β and the resulting SLOM score. This means that the computational complexity of the complete algorithm is $\mathcal{O}(N \cdot n)$. In our case $n = 3$.

5.3.2 Post-processing

The SLOM algorithm will find single grid cells with high SLOM scores. Since we are working with climate model data however, we expect outliers to cover larger regions than single cells. Examples are a peak caused in wind energy capacity due to a storm or a dip in solar photovoltaic energy capacity due to heavy cloud cover. Such weather systems often span areas larger than a single grid cell of our data. In order to find larger spatial regions that are outlying, a post-processing procedure is used. It merges neighboring SLOM scores if they are larger than some user defined threshold value.

The procedure is inspired by Wu et al. (2008) [65]. In their paper they find spatial outliers, and track them through time using a tree based approach. The tree contains all potentially overlapping spatial outliers over time. Thus the tree contains all potential temporal evolutions of the spatial outliers.

The tree based tracking algorithm used by Wu et al [65]. is not suited for our purposes, as it is exponential in time. Since the tree contains all possible temporal evolutions of an outlier it can become quite large. The algorithm presented by Wu et al. to traverse the tree and look at all potential temporal evolutions has a computational complexity of $\mathcal{O}(n^y)$ where n is the number of outliers of temporal length one being analyzed, and y the number of time steps being analyzed. Since exponential in time isn't feasible for our dataset we use a simple greedy overlap check of outliers to create outlying events.

In order to find the temporal evolution of outliers we start with the largest found SLOM score, and look forward and backwards in time and check if they sufficiently overlap with another outlier. If so, we merge them to the same outlying event. In addition to just temporal overlap like performed by Wu et al. (2008) [65], we also merge spatially neighboring grid cells if their SLOM scores are high enough. This continues until we find no more outliers to add, and move to the next highest SLOM score. The pseudocode of the merging process can be seen in algorithm 2. Note that in this pseudocode a grid is an object with an associated latitude, longitude, time and SLOM score.

The algorithm recursively extends an outlier grid to form an event. In lines 1-4 we see that the algorithm has several variables and a user defined parameter. The threshold value, in line 1, is set by the user. This threshold determines if a SLOM value is high enough to be considered part of an outlier event. The Timestep variable indicates the temporal resolution of the data that the SLOM algorithm was used on. The AssignedEvents list is used to keep track of what data points are already part of an event, as a data point is only allowed to be part of a single outlying event. The TopSLOM list contains the data points with the highest SLOM scores. It should be noted that the merging algorithm also uses all the data for which the SLOM scores are calculated, and not just the TopSLOM. The neighboring data points of top slom

Algorithm 2: SLOM Outlier Merging

Result: Merged Outlier Events

```
1  Threshold = User Defined
2  Timestep = Temporal Resolution
3  AssignedEvents = Empty list
4  TopSLOM = Sorted List of grids with Top SLOM scores
5  for each outlier_grid ∈ TopSLOM do
6      if outlier_grid ∉ AssignedEvents then
7          AssignedEvents.append(outlier_grid)
8          return Extend([outlier_grid], outlier_grid)
9  end
10 Extend(event, outlier_grid)
11     past_grid = outlier_grid - Timestep
12     if past_grid["SLOM"] >= Threshold then
13         if past_grid ∉ AssignedEvents then
14             AssignedEvents.append(past_grid)
15             event.append(past_grid)
16             event = Extend(event, past_grid)
17     future_grid = outlier_grid + Timestep
18     if future_grid["SLOM"] >=Threshold then
19         if future_grid ∉ AssignedEvents then
20             AssignedEvents.append(future_grid)
21             event.append(future_grid)
22             event = Extend(event, future_grid)
23     for each neighbor_grid of outlier_grid do
24         if neighbor_grid["SLOM"] >=Threshold then
25             if neighbor_grid ∉AssignedEvents then
26                 AssignedEvents.append(neighbor)
27                 event.append(neighbor_grid)
28                 event =Extend(event, neighbor_grid)
29     end
30     return event
```

grids that are checked do not have to be in TopSLOM.

Each outlier in TopSLOM gets extended in lines 5-8. Line 6 checks if the top SLOM outlier is already assigned, in which case it is skipped. If it is not assigned yet, it is extended. The Extend method is used recursively, and as input takes the merged event so far, and the data point for which the neighbors are checked.

The *past_grid* in line 11 refers to the same location, but one time step earlier, in the original data with SLOM scores. The SLOM score of this *past_grid* is checked in line 12, if it surpasses the Threshold value it will potentially be added to the event. Before *past_grid* can be added, we check in line 13, that it does not already belong to another event. If it does not yet belong to another event, *past_grid* is added

to the event. After the *past_grid* has been added to the event, extend is called recursively to check, and potentially merge, all neighbors of *past_grid*. Lines 17-22 describe exactly the same process, but for $future\_grid$, which is the data point at the same location but one time step later.

For all potential spatial neighbors of the *outlier_grid*, lines 23-28 describe a similar processes. These *neighbor_grid* data points refer to data points that are the spatial neighbors of *outlier_grid* at the same moment in time. We consider diagonal data points to be neighbors. So each data point not at the edge has eight neighbors.

After the recursive process all potential neighbors, spatial and temporal, are checked to see if they belong to the SLOM event. The returned SLOM event can contain several time steps. The region of the event does not have to be same at every time step. It should be noted that the temporal range of the SLOM event does not have to be same as the temporal range of the MDI outlier. The temporal range of the SLOM event is a subset of the temporal range of the MDI outlier.

It is possible that during the merging process data points get merged to an event that are part of the top SLOM outliers. This means that the number of merged SLOM events does not have to be equal to the number of top SLOM outliers.

It is possible that when merging a SLOM outlier to form a SLOM event, all other data points are merged. All data points will be checked at most once. The complexity of the merging algorithm is thus $\mathcal{O}(N)$, where N is the number of data points.

The advantage of our post-processing, over the approach of Wu et al. is that we need not build a tree of all possible outlier events. By doing this we might miss some possible connection between events, but only if we picked a different time progression of an outlying event. This approach still captures all spatial outliers, and a possible temporal evolution, without Wu et al.'s exponential tree building algorithm.

This approach would find spatial outliers, their temporal location, and a possible evolution over time of the outlying event. It is possible that this procedure merges different outlier events into a large outlier event. This is of little concern, as the main use of these spatial outlier events is to highlight regions for further investigation by domain experts. Whether one region containing several outliers, or several outlier regions are highlighted should make no difference.

Process Pipeline

A figure representing the process pipeline can be seen in figure 15. A cylinder shape indicates a dataset, the hexagons indicate preprocessing steps, the squares contain algorithms and the rhombus shapes indicate (intermediate) results.

The initial data, shown in green, contains energy generation data based on ERA5 climate data and energy conversion models. There are three variables, wind-onshore, wind-offshore and solar photovoltaic energy generation. The yellow preprocessing steps are performed on the original data according to section 5.2. The spatial-temporal data is transformed to temporal data by summing over the region being investigated. These regions are potentially determined via clustering. The seasonal behavior of the temporal data can be removed using Z scores.

This time series data is then passed to the MDI algorithm. All steps performed by the MDI algorithm are shown in red. The time series gets normalized by sub-

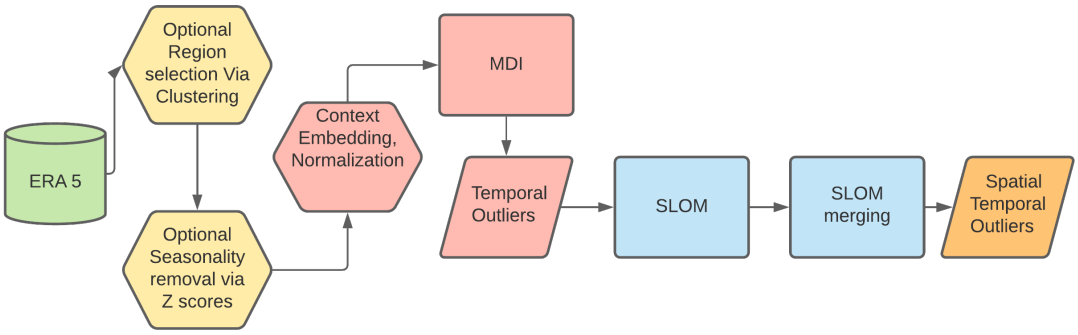


Figure 15: The process pipeline used in this research. Combines MDI and SLOM in a temporal first approach. Colors indicate to what the step relates. Data shown in green, our optional preprocessing is Yellow, red is MDI, SLOM is blue and the final result in orange.

tracting the average, and dividing by the maximum. The context embedding step incorporates context by adding previous data points to each data point. This normalized time series with context is then used to detect maximally divergent intervals. The resulting temporal outliers are then used by SLOM.

SLOM procedures are shown in blue. The temporal outliers resulting from MDI indicate the temporal ranges during which a temporal outlier occurs. These ranges are used as input for SLOM. The original ERA5 energy generation data, limited to those temporal ranges, is used to calculate SLOM scores. The highest SLOM scores are merged by the SLOM merging algorithm, resulting in SLOM events. These SLOM events have a temporal range that is a subset of the temporal range of the MDI outlier. The spatial range of a SLOM event can change per time step, but must be connected.

The final result contains SLOM events that occur during temporal outliers detected by the MDI algorithm, and is indicated by the orange rhombus. A temporal outlier detected by the MDI algorithm can contain multiple SLOM events.

5.4 Implementation

In this section we provide a brief discussion on the implementation details of the algorithms. The implementation can be found at `https://git.science.uu.nl/E.Duijm/thesis`.

Since the data is in the NetCDF-4 format, we use the xarray package for Python [85]. The package is well suited for working with large NetCDF-4 files and datasets. As stated in section 3, our dataset is rather large at 568GB. For data operations such as selecting regions, calculating averages and standard deviations, built in xarray operations were used.

The original MDI algorithm is implemented in an open source library and can be found at `https://github.com/cvjena/libmaxdiv`. It contains a Python implementation of the algorithm and a C++ implementation. The Python algorithm

isn't well suited for large data sets, thus the C++ implementation was used. In order to communicate with the C++ implementation we built a wrapper in Python. In addition changes were made to allow the setting of spatial context embedding parameters. Semantically our implementation of the algorithm is the same as the original implementation by the authors.

The SLOM algorithm was originally implemented in Python, using xarray. The built-in parallel computing provided by xarray via the DASK distributed package [86] was the main reason behind the choice to use Python. Unfortunately we ran into serious issues with the DASK distributed package, rendering the Python implementation useless.

Because of this another implementation was made in Cython [87]. Cython compiles C++ code based on Python code. Cython was chosen as much of the original implementation could easily be used. The C++ code compiled by Cython was significantly faster than the old Python code. So much so that we use a single threaded Cython implementation. This single threaded approach was fast enough for our purposes, but our approach can easily be extended to allow for multi-threading.

Our implementation combining Python and C style code is in line with the advice given by Singh et al. [88] for working with big NetCDF datasets.

6 Experiments

In order to investigate outliers that might impact the future energy grid we perform several experiments on energy generation data based on ERA5 reanalysis data, as described in chapter 3. Each experiment represents a potential use case where outlier detection might provide insight in critical systems that could influence the future energy system, and is thus suited as a testing ground for the algorithms. We performed three different experiments.

In section 6.1 we look at outliers in temporal data created by summing the total energy generation over potentially interesting regions. These regions are the entire Western European region and the clusters described in table 2. We refer to these experiments as the Regional Experiments. In addition to detecting outliers, another goal of these experiments is to investigate the effect of the preprocessing described in section 5.2.

In order to look at the potential effect of climate change, the top outliers per decade are looked at in section 6.2. This experiment is referred to as the Climate Change Experiment. The goal of this experiment is to look at the potential effect of climate change, but it also is a use-case for trend analysis in univariate systems.

Finally, in section 6.3 the capability of the MDI algorithm to find temporal outliers and their spatial location in spatial-temporal data is investigated. This experiment is called the Spatial Location Experiment. This is done in order to see if using the MDI algorithm on spatial-temporal data, without spatial context embedding, is a viable option, and provides additional insights over the strictly temporal approach of the Regional Experiments.

All experiments are performed on an Intel Xeon Gold 6130 CPU with 16 dual-cores at 2.1 GHz clock speed. Our setup has 125.6GB of available RAM memory. The multi-threading was limited to using 30 threads.

6.1 Regional Experiments

Since we are interested in the total renewable energy generation, we perform experiments on temporal data series constructed by summing over the entire spatial region being investigated. After summing we are left with a time series of 613,594 values, as we have 70 years of hourly data. At each time step the total energy generation per energy source is computed.

6.1.1 Full Region

In order to investigate potential shortfalls or surges that might affect the European energy system we look at the entire Western European region as defined in Chapter 3. For this experiment all variables: Wind-Onshore, Wind-Offshore and Solar Photovoltaic energy generation are used.

In order to select the correct distribution for the data, the data was fitted to both a Gaussian and a KDE using Gaussian Kernels. These fits can be seen in figure 16. These fits show us that the more complex KDE model is capable of better describing the data. This is not surprising, as any data that can be described by a

Gaussian Distribution can be described by a KDE using Gaussian kernels. Because of this we will be using the KDE using Gaussian kernels as distribution estimation model for all our Regional Experiments in section 6.1.

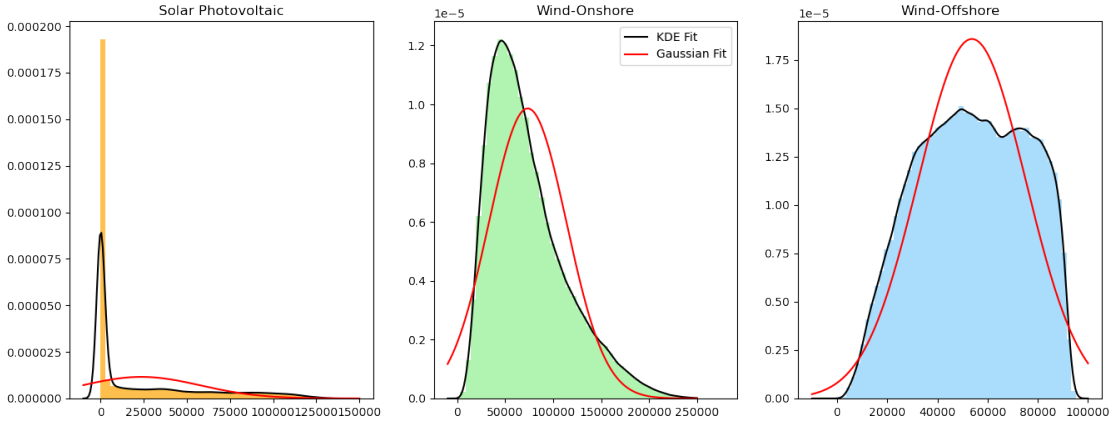


Figure 16: Histograms of the Solar Photovoltaic, Wind-Onshore and Wind-Offshore energy generation time series, plotted together with a fitted Normal Distribution and a fitted KDE using Gaussian Kernels

On the temporal series based on the entire Western European region, four experiments were performed. We looked at using the regular data, as well as Z scores in order to investigate the influence of seasonality. For each we used both Cross Entropy and the Unbiased Kullback-Leibler divergence measure to score outlierness. For temporal context embedding K=4 and T=8 parameter values were used, as discussed in section 5.1.4.

We used Kernel Density Estimation with Gaussian Kernels with a kernel width h of 1. Hotteling's T^2 proposal method is used and proposals starting and ending with a data point that has a gradient magnitude larger than the average gradient magnitude + 1.5 times the gradient magnitude standard deviation are proposed. The top 50 outliers with length between 2 days and 10 days are returned. We set the minimum length of the detected outlier to be 2 days, to ensure there is sufficient data to estimate a distribution.

For the experiments where the regular, non Z score, data was used, the built in normalization method of the MDI algorithm was used. This normalization subtracts the global average from each data point, and then normalizes by dividing by the maximum.

The top outliers for each experiment are presented below. The top outlier detecting using Cross Entropy divergence measure on regular data is depicted in figure 17 and using Unbiased Kullback-Leibler on regular data is shown in figure 18.

The top outlier based on Cross Entropy coincides with a period that is identified by Dawkins et al. [27] as an adverse weather system for the electricity system of the United Kingdom and Europe. The top outlier detected using the Unbiased Kullback-Leibler divergence is considered the worst storm in decades in The Netherlands [89, 90]. This storm was considered one of the worst of the century and 17 people lost their lives.

For Z scores the Cross Entropy and Unbiased Kullback-Leibler divergence results can be seen in figures 19 and 20 respectively. All top 20 outliers for all Regional Experiments are published online at `https://git.science.uu.nl/E.Duijm/thesis/-/tree/master/result_figures`.

Both the Cross Entropy and Unbiased Kullback-Leibler methods on the regular data took just over 29 hours wall clock time to calculate. For the Z score data just over 27 hours was needed for the MDI algorithm to run.

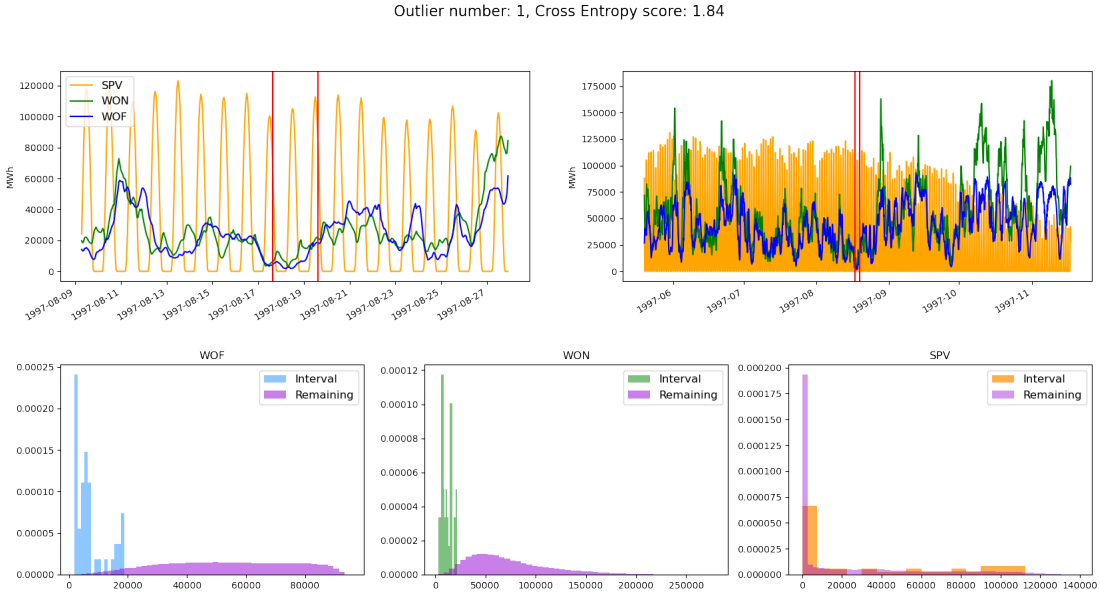


Figure 17: Figures depicting the outlier with the highest score using the Cross Entropy divergence measure. The vertical red lines indicate the start and finish of the detected interval. The top left image shows the outlier with several days before and after the outlier for context. The top right image provides larger context to the outlier. The bottom figures show histograms of the outlying interval per energy source combined with the histograms of the remaining data.

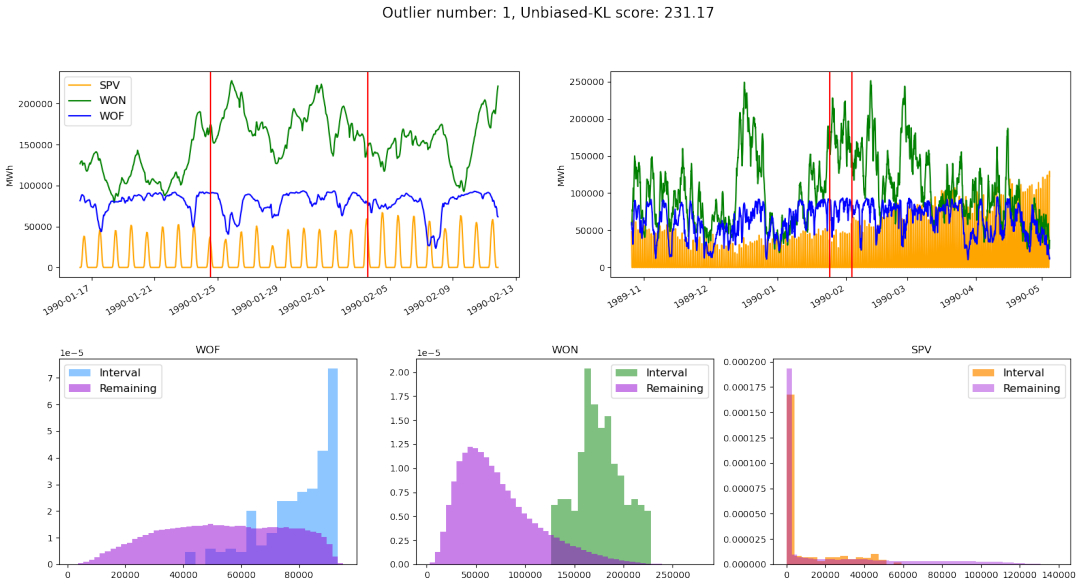


Figure 18: As Figure 17 but here for Unbiased Kullback-Leibler divergence.

Figure 19: As Figure 17 but here for Cross Entropy and Z scores.

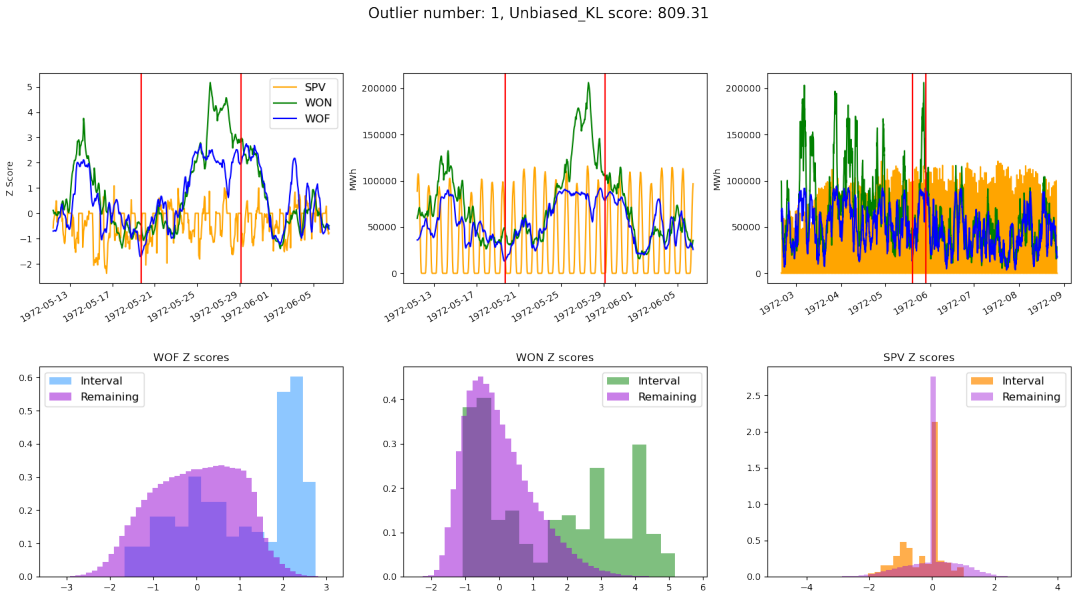


Figure 20: As Figure 17 but here for Unbiased Kullback-Leibler divergence and Z scores.

To learn more about the outliers detected by the algorithms, they were classified. For each of the four combinations of data type and divergence measures the top 20 outliers were analyzed. The grouping was based on the month of the year, the length of the outlier in hours, the outlying variables and whether they resembled a peak in energy generation, or a trough.

A Peak-and-Trough indicates that the outlier contains a variable that has peak, as well as one that has a trough and that it isn't clear if the overall energy generation experiences a peak or a shortage. The grouping based on these attributes for regular data can be seen in tables 3 and 4.

Outlier Type	Month	Length(h)	Outlying Variables	Outlier numbers
Trough	8	48-72	SPV High, WON Low, WOF Low	1,6,13,19
Peak	12	48-72	WON High, WOF High, SPV Low	2,11
Trough	6	48-72	SPV High, WON Low, WOF Low	3,5
Peak	4	48-72	WON High, WOF High	4
Trough	7	48-72	SPV High, WON Low, WOF Low	7,8,9,16,17
Trough	7	72-96	SPV High, WON Low, WOF Low	16
Trough	7	150-175	SPV High, WON Low, WOF Low	10,15
Peak	2	48-72	WON High, WOF High, SPV Low	12,18
Peak-and-Trough	2	48-72	WON High, WOF Low, SPV Low	14
Peak	1	48-72	WON High, WOF High, SPV Low	20

Table 3: Grouping of the top 20 outliers found by the MDI algorithm on regular data using Cross Entropy divergence measure.

Outlier Type	Month	Length(h)	Outlying Variables	Outlier numbers
Peak	1	216+	WON High, WOF High, SPV Low	1,5,7,10
Peak	12	216+	WON High, WOF High, SPV Low	2,8,12,17
Peak	2	216+	WON High, WOF High, SPV Low	3,4
Peak	2	216+	WON High, WOF High	6,13,16
Peak	11	216+	WON High, WOF High, SPV Low	11,18,19,20
Trough	8	216+	SPV High, WON Low, WOF Low	14
Trough	7	216+	SPV High, WON Low, WOF Low	15
Peak	1	192-216	WON High, WOF High, SPV Low	9

Table 4: Grouping of the top 20 outliers found by the MDI algorithm on regular data using the Unbiased Kullback-Leibler divergence measure.

Based on the grouping we defined classes for the outliers. For the Unbiased Kullback-Leibler divergence on regular data these classes are Winter Surplus and Summer Deficiency. We consider the outliers that show a peak in energy generation that are in the extended winter to be part of the class Winter Surplus. The extended winter is the period from November up to and including March. Outlier events that show a trough in overall energy generation in the extended summer, from May up to and including September, are part of the class Summer Deficiency.

For Cross Entropy on regular data we have similar classes: Winter Surplus, Long Term Summer Deficiency and Short Term Summer Deficiency. The only distinction is that for the Summer Deficiency we have sub classes based on the length of the event. Outliers that last between 48 and 72 hours are considered short term, and outlier events longer than 72 we consider long term.

We note that the outliers detected by the Cross Entropy divergence measure have in general a very short duration. The outlier events highlighted by using the Unbiased Kullback-Leibler divergence all have a rather long duration. The different methods appear to have some bias based on the length of the outlier interval, even though one method is created specifically to be unbiased.

For the Z scores the classification was infeasible, due to the lack of coherence in the discovered outliers. For the top 20 outliers using the Unbiased Kullback-Leibler divergence we found 20 groups. For the Cross Entropy top 20 we found 17 groups.

In addition we noted that when looking at the energy generations at the times of Z score outliers, the outliers were in general less intense and less informative. The outliers based on Z scores yielded no additional insight when compared to the regular data. Because of this the remainder of the experiments, described in sections 6.1.2, 6.2, 6.3 are performed on only the regular data.

Spatial Investigation of Temporal Outliers

The top outliers returned by the MDI algorithm based on regular data and both divergence measures were investigated further using the SLOM algorithm. For this we set the post processing threshold parameter at the top 5% of SLOM scores. If this top 5% threshold value is lower then 0.1, we use a SLOM score of 0.1 as threshold instead. The top SLOM outlier for the top MDI detection using the Unbiased Kullback-Leibler and Cross Entropy divergence measures are presented in figures 21 and 22 respectively.

For both SLOM outliers, the time step containing the highest SLOM score is shown. The SLOM score was calculated using all variables, but the wind-offshore variable is shown in the figures. This is because the offshore wind energy generation is the dominant variable in these calculations, as the highlighted regions contain almost exclusively offshore wind. The SLOM region spans three hours for the Cross Entropy outlier and four hours for the Unbiased Kullback-Leibler divergence.

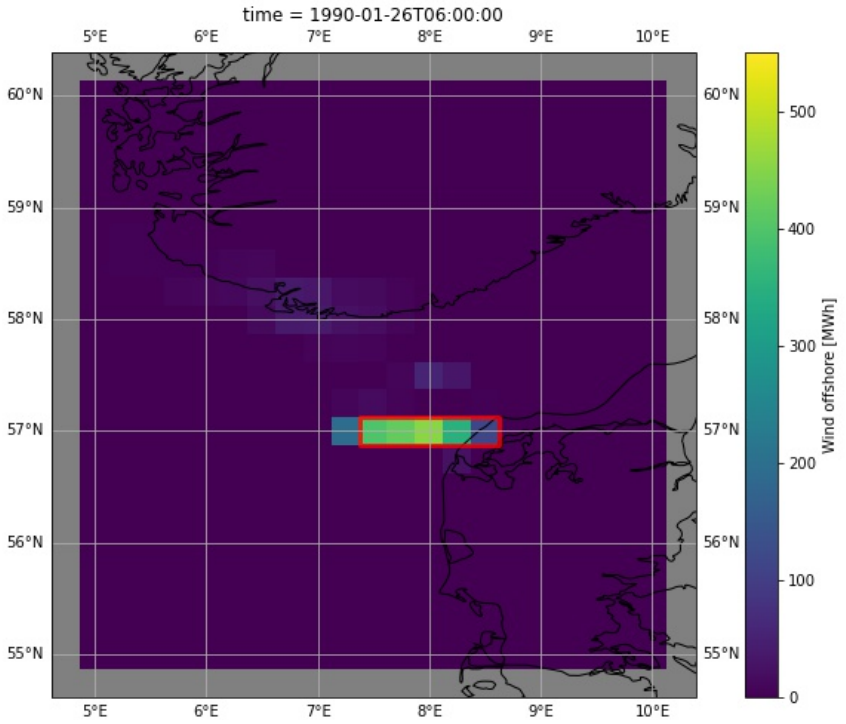


Figure 21: Region with the highest SLOM score in the period containing the largest temporal outlier detected by the MDI algorithm using the Unbiased Kullback-Leibler divergence. Offshore wind is shown as it is the main contributor to the SLOM score.

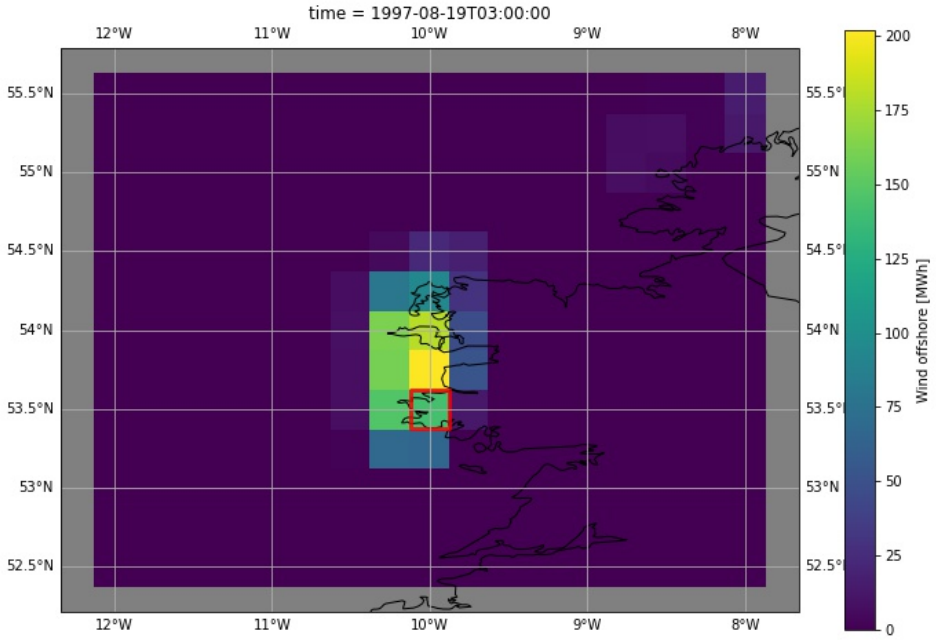


Figure 22: As figure 21, but for the outlier detected using the MDI algorithm with Cross Entropy as divergence measure.

Several SLOM outliers from different MDI outliers were studied, all with similar results as seen in figures 21 and 22 . The highest SLOM scores are assigned to areas with high wind-offshore energy generation. Even in situations where the temporal outlier is a trough, or a peak that is caused by wind-onshore energy generation.

The reason for this is that the wind-offshore energy generation is less spread out then the other energy generation forms. The wind-offshore energy generation is more clustered together. This causes grids that potentially have very high generation to be next to grids with no generation at all. The other energy sources appear to be better spread out and suffer less from this problem. The result is that the SLOM results are based on the offshore wind cluster, even if the temporal outlier has nothing to do with it. In addition, temporal outliers that affect the entire European energy grid are unlikely to be caused by a spatial outlier, as the energy generation is spread out across Europe.

Due to where offshore wind parks are located and how dense the energy generation there is, combined with the scale of the region being analyzed, we found that the SLOM algorithm provided no additional insight in the detected temporal outliers.

To summarize, the temporal outliers detected by the MDI algorithm are interesting as they contain two classes that are difficult to deal with. These classes, the Summer Deficiency and Winter Surplus, can be problematic as they influence the whole network. A long deficiency needs to be compensated with other methods of non-carbon generation that need to be flexible and can be controlled, as the current battery capacities aren't sufficient. Energy shortages are a huge risk, thus such Summer Deficiencies need to be taken into account by policy makers.

Shorter deficiencies during the summer are also problematic, as they require extensive use of battery capacity. During the day these charge on the available solar photovoltaic energy generation, but at night these need to be discharged to compensate for the lack of wind. This strain on the batteries causes them to wear. An increase in such short deficiencies represents an economic risk, as the batteries would need to be replaced more frequently. The Summer Deficiencies are clear examples of the low energy generation situations that pose a risk, as identified in section 2.2.

The Winter Surplus increases the energy generation of the grid, causing a surplus, which can be problematic if this isn't controlled. The surplus needs to be discharged somehow, as an energy surplus influences the frequency of the network. This discharge of unused energy represents an economic risk, as the wind turbines and solar panels are wearing down without the energy that is generated being used. These Winter Surplus events do not directly match one of the potentially risky situation identified in section 2.2. They represent a risky situation we did not anticipate.

We found that the SLOM algorithm adds no further insights into these temporal outliers. The specific properties of the data makes it difficult to detect spatial outliers. It is also possible that the detected outliers are purely temporal.

6.1.2 Clustered Regions

Based on the clustering performed in section 5.2.2 we have defined several potentially interesting regions. These regions represent different use cases for the algorithm. For ease of reading, the table containing information on the clustered regions is repeated in table 5. For the experiments on the clustered regions the same parameter values are used as with the experiment on the full region, described in section 6.1.1.

Region Name	Energy Type	Latitude Bounds	Longitude Bounds
Mainland Onshore	Wind-Onshore	50 : 53.5	4.5 : 15
North Sea Offshore	Wind-Offshore	53.75 : 57.5	4.75 : 8
Spanish SPV	Solar Photovoltaic	35.5 : 38.5	-9 : -0.5
French Mixed	Mixed	42.5 : 50	-5 : 9

Table 5: Table containing information on use case regions selected based on clustering of the temporal mean of ERA5 energy capacity data. Copy of table 2

For each of these regions time series are created by summing over the region. This results in 613,594 data points for 70 years of hourly data. Based on the results of the MDI algorithm on the full region, only the regular data containing seasonal behaviour was used for the experiments with the clustered regions.

Since we noted a bias in the length of detected outliers based on the selected divergence measure, we adjusted the parameter values in order to save time. Cross Entropy was used to detect outliers in intervals at least 2 days long and at most 7 days. The Unbiased Kullback-Leibler divergence measure was used to detect outliers with a duration of at least 8 days and at most 10.

Mainland Onshore Region

This experiment is performed on only the onshore wind variable. The goal is to investigate the use of our methods for onshore wind. The Cross Entropy results were calculated by the MDI algorithm in just under 18.5 hours and the Unbiased Kullback-Leibler results took just under 21.5 hours. The outliers with the highest divergence scores for Cross Entropy and Unbiased Kullback-Leibler are shown in figure 23.

For both scoring methods we found that the top 50 outliers contain exclusively peaks in energy generation. Taking a closer look at the top 20 outliers of both methods we found that almost all outliers occur within the extended winter period from November up to and including March. The only exception is the rank 15 outlier using Cross Entropy, as it occurred in October. All outliers are classified as Winter Surpluses, with the exception of the one surplus in autumn.

The length of the detected outlier events depends on the method used, as we looked for shorter interval outliers using Cross Entropy, and limited Unbiased Kullback-Leibler to longer intervals, due to the inherit bias we observed.

We used SLOM on the temporal ranges highlighted by the outliers. We only used the wind-onshore variable to calculate the SLOM scores. The merging threshold is set to the top 5% of SLOM scores. In most cases this resulted in a threshold of lower than 0.1, in which case 0.1 was used.

Figure 23: As Figure 17 but here for only wind-onshore on the Mainland Onshore Region. Cross Entropy on the left and Unbiased Kullback-Leibler on the right

The time step containing the top SLOM scores for the top temporal outliers detected using the MDI algorithm with Cross Entropy and Unbiased Kullback-Leibler divergence measures are shown in figures 24 and 25.

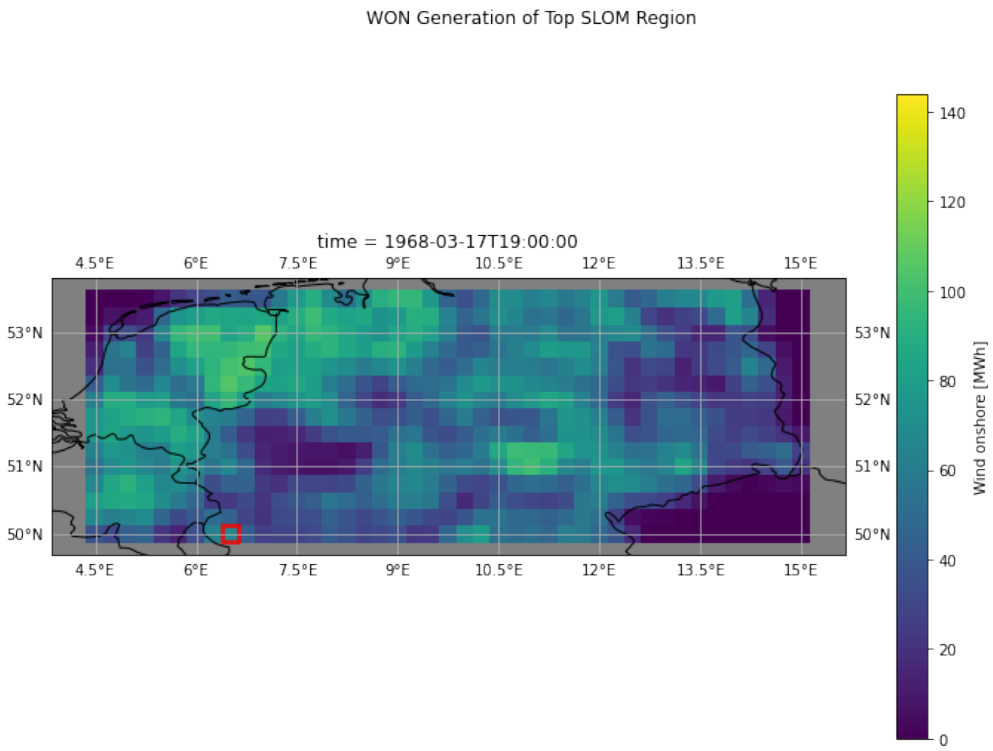


Figure 24: Region with the highest SLOM score in the period containing the largest temporal outlier detected by the MDI algorithm using the Cross Entropy divergence. Calculated using onshore wind energy generation.

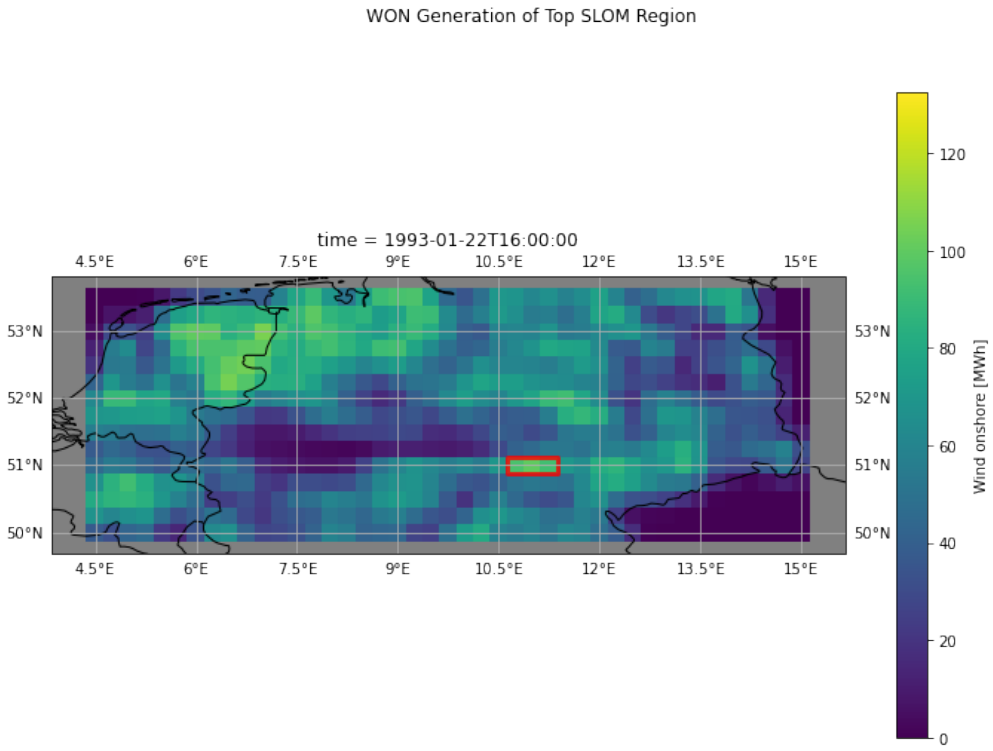


Figure 25: As figure 24, but MDI algorithm used Unbiased Kullback-Leibler divergence.

The regions with the highest SLOM scores are indicative of what we have seen with the highest SLOM regions of several top temporal outliers. The highlighted areas provide no new insights. They do not show any regions that allowed us to learn more about the outlier event.

It is possible that the outliers detected by MDI are purely temporal outliers. An event that increases the generation across the entire region would not be detected by a spatial outlier detection algorithm. Seeing how we are looking at the top temporal outliers over 70 years of data, summed over the entire region, it is very likely that the events that cause the outliers cover the entire region.

North Sea Offshore Region

For this experiment only the offshore wind variable is used. This experiment is done to see how well our methods are suited to investigate just offshore wind. The MDI algorithm took just over 18.5 hours to calculate the Cross Entropy results. The Unbiased Kullback-Leibler results took just under 21.5 hours to calculate.

The top temporal outliers detected using the MDI algorithm on wind-offshore data from the North Sea Offshore region are shown in figure 26.

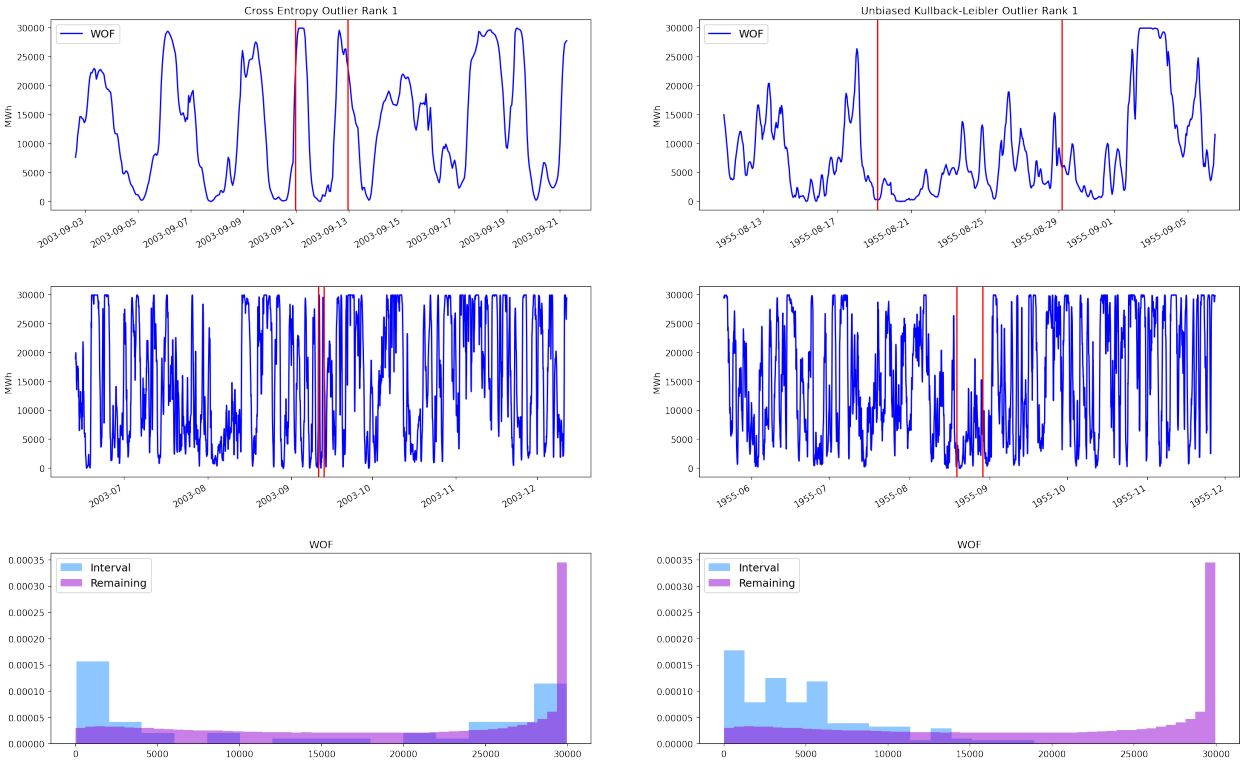


Figure 26: As Figure 17 but here for only wind-offshore on the North Sea Offshore Region. Cross Entropy on the left and Unbiased Kullback-Leibler on the right

The top 50 outliers for each method contain mainly troughs in offshore wind production. From the top 50 there are seven peaks in the Unbiased Kullback-Leibler outliers and one peak in the Cross Entropy outliers. This lack of peaks could be caused by the fact that offshore wind energy generation reaches a plateau. This

plateau value is reached quite often, and thus period of high generation aren't that exceptional.

Taking a closer look at the top 20, we note that for both methods these events occur throughout the year. This means that what we detect are not simply Summer Deficiencies as seen with the Western European Region. We will refer to these outliers as Wind Droughts. These Wind Droughts are an example of the low energy generation situation mentioned in section 2.2.

The investigation of these temporal outliers using SLOM was performed with a parameter value of top 5% for the post-processing threshold, with a fallback value of 0.1. The results we see from the SLOM algorithm on the temporal ranges indicated by the MDI outliers, are similar to those from the Western European Region. This makes sense, as we are looking at the same region, and the same energy generation variable, WOF, that dominated in the Western European SLOM results.

The region generates high SLOM scores for WOF due to the properties of the offshore wind energy generation in the North Sea. Offshore wind farms are highly condensed forms of energy generation, causing very high generation grids, to be right next to very low, or no generation grids. These contribute heavily to the SLOM score.

In addition we note that the investigated temporal outliers belong to the Wind Drought class. This means that the event causes an energy deficiency. SLOM does not deal with such a situation well. It simply finds those grids where there happens to be some energy generation, despite the overall lack of energy generation.

Spanish SPV Region

The Spanish Region experiment represents the capabilities of our method to analyze solar photovoltaic energy generation. Only the SPV variable was used for these experiments. The Cross Entropy part of the results took just over 15 hours to calculate using MDI, the Unbiased Kullback-Leibler part just over 17.25 hours. The top temporal outliers detected by the MDI algorithm are shown in figure 27.

Looking at the top 50 outliers we see that the method used to calculate the divergence determines what type of outliers we find. All 50 outliers detected by Cross Entropy have, on average, a higher SPV energy generation then the global SPV average.

The top 50 outliers found using Unbiased Kullback-Leibler all contain troughs in solar photovoltaic energy generation. Taking a closer look at the top 20, we see that most of them are around the period December to January, where solar photovoltaic energy generation is generally at its lowest.

We define the class of outliers detected by Cross Entropy to be Summer High SPV. The troughs detected by Unbiased Kullback-Leibler we will refer to as Winter Low SPV. Both classes can potentially be risky. The Winter Low represents a potential low energy generation situation as described in section 2.2. Especially if a region, depends heavily on solar photovoltaic energy generation.

The peaks in the Summer High SPV class can be problematic as, with the Winter Surplus, this potentially requires the discharging of generated electricity. This generating of electricity without using it represents an economic risk. As with

Figure 27: As Figure 17 but here for only solar photovoltaic on the Spanish SPV Region. Cross Entropy on the left and Unbiased Kullback-Leibler on the right

the Winter Surplus, these high energy generation events are situations not considered beforehand.

Again these temporal outliers were investigated further using SLOM. We found that the top SLOM outliers detected in the temporal ranges indicated by the temporal outliers barely exceed a value of 0.1. Seeing how the SLOM scores are limited to the range $(0, \sqrt{n})$, where n is the number of variables, we see nothing that can be considered a spatial outlier.

The temporal outliers appear to be strictly temporal. In addition the solar photovoltaic energy generators in the Southern Spanish Region are grouped rather tight in several grids. This means that these grids, right next to grids with no or very little generation, dominate the SLOM results. Even so the SLOM scores are very low. We see no new information that can be learned from these SLOM regions.

French Mixed Region

The French Mixed region is a use case that represents outlier detection on a national scale. France is an important part of the network, as it houses many connections and has diverse energy generation capabilities. This experiment uses multivariate data, using all energy generation variables. The MDI algorithm using Cross Entropy as divergence measure took just over 27.5 hours, and the Unbiased Kullback-Leibler part of the experiment took just over 31.5 hours to run.

The top temporal outliers detected by the MDI algorithm using Cross Entropy and Unbiased Kullback-Leibler divergence are shown in figures 28 and 29.

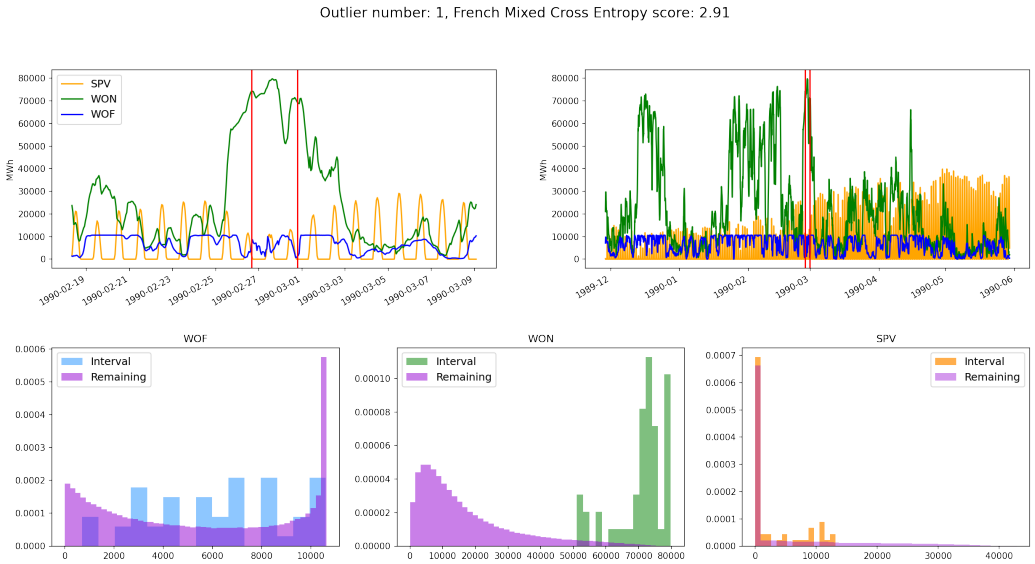


Figure 28: As Figure 17 but here for the French Mixed Region.

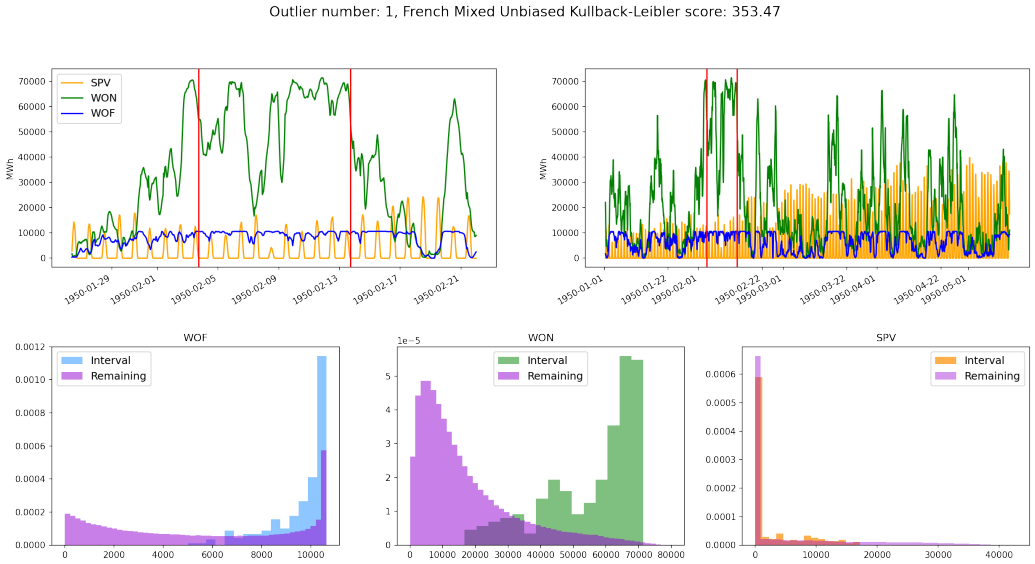


Figure 29: As Figure 17 but here for the French Mixed Region and Unbiased Kullback-Leibler divergence

Looking at the top 50 outliers for both methods, we found that nearly all detected outliers contain a peak in energy generation. A single outlier detected using Unbiased Kullback-Leibler represents a trough.

When taking a closer look at the top 20 outliers, wind-onshore appears to be the most important variable, followed by wind-offshore. solar photovoltaic energy generation appears to have little influence on the outliers. This isn't surprising, as approximately 20% of the total energy in the French Region in our data is generated by solar photovoltaic energy generation, and the remainder by Wind.

All outliers in the top 20, using both Cross Entropy and Unbiased Kullback-Leibler, are peaks within the extended winter period. All temporal outliers discovered are classified as Winter Surplus.

For the investigation of these temporal outliers using SLOM we again set the threshold value at the top 5% of SLOM scores. In practice this is often lower than 0.1, in which case 0.1 was used. Unfortunately, as we have seen in the other Regional Experiments, we note that the top SLOM outliers contain no additional insights into the detected temporal outliers.

Clustered Regions Results

To summarize the results of the experiments performed on the clustered regions, we have seen that depending on your use case, it is important to carefully think about the region, energy variables and method you use. Depending on the method, troughs or peaks are detected.

The energy variables also play an important role the type of outliers detected. When looking at the Western European Region, using all variables, Summer Deficiencies are detected. These are caused by a lack of both offshore and onshore wind energy generation. Yet when looking at a wind onshore use case, in the Mainland Onshore Region, only peaks are detected.

For all clustered regions we found that the SLOM algorithm provide no additional insights. This can be caused by a temporal outlier affecting the entire region, and thus is not considered a spatial outlier. Seeing how we are looking at the top few temporal outliers in 70 years of data, it is likely that these affect the entire region.

In addition the data has an underlying spatial bias, due to the fact that the energy generation has physical location and potential energy generation capacity per grid. This influences SLOM scores.

The fine grained nature of the data is potentially difficult for the SLOM algorithm to deal with. Consider there is some spatial phenomenon, such as a local storm, which is the cause of a temporal outlier. Due to the high resolution of the data, the transition from low energy generation to high due to the storm will be rather smooth. This lowers the difference between neighbors, lowering the overall SLOM scores.

6.2 Climate Change Experiment

In order to determine if there is an effect of climate change on outliers, we look at the top outliers per decade. The idea is that our model is based on historical data. This data thus contains any perceived effects due to climate change. By looking at the per decade top outliers we try to learn if there are changes in the intensity or duration of outlier events. The time of year of the outlier events is also investigated, to study if there are any shifts in when outlier events occur.

These experiments are performed on the sum of all variables. So we combine Wind-Offshore, Wind-Onshore and Solar Photovoltaic into a single variable we refer to as Total Renewable Energy Generation. This is done as the total energy generation is easier to interpret and provides good indication of shortages and surges in the energy system. A temporal series was constructed as in experiment 6.1.1, by summing over the Western European region.

Again we will use the more flexible Kernel Density Estimator using Gaussian Kernels for the distribution estimates. As discussed in section 6.1.1 we will only use the regular data containing seasonal effects, Z scores will not be used. The same parameter values for the MDI algorithm as in the Regional Experiments of section 6.1 are used for this experiment.

Outlier Intensity

The intensity of the outliers is investigated by looking at the average energy generation during the outlier. Boxplots of the average energy generation during the outlier for the top 5/20/50 outliers can be seen in figures 30, 31, 32 for Cross Entropy and figures 33, 34, 35 for the Unbiased Kullback-Leibler divergence measure.

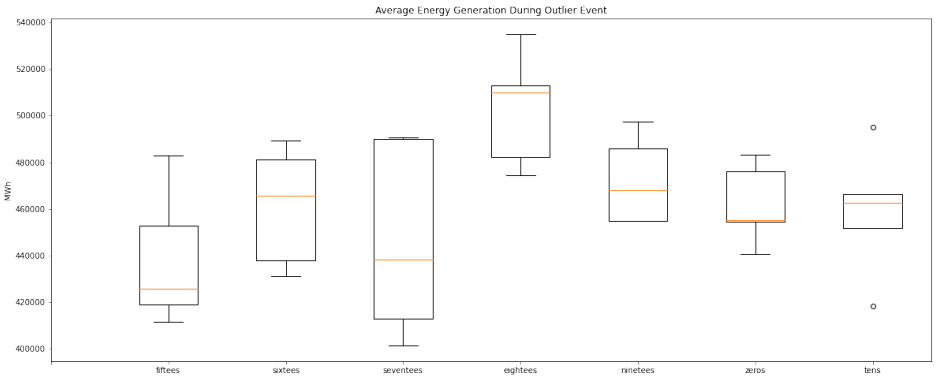


Figure 30: Boxplots of the average energy generation during an outlier event per decade. Using the Cross Entropy divergence measure. Top 5 outliers per decade were analyzed

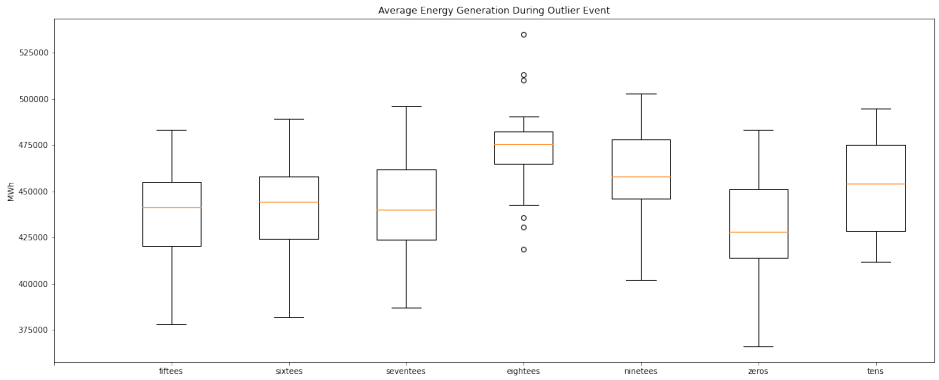


Figure 31: As figure 30. With top 20 outliers per decade

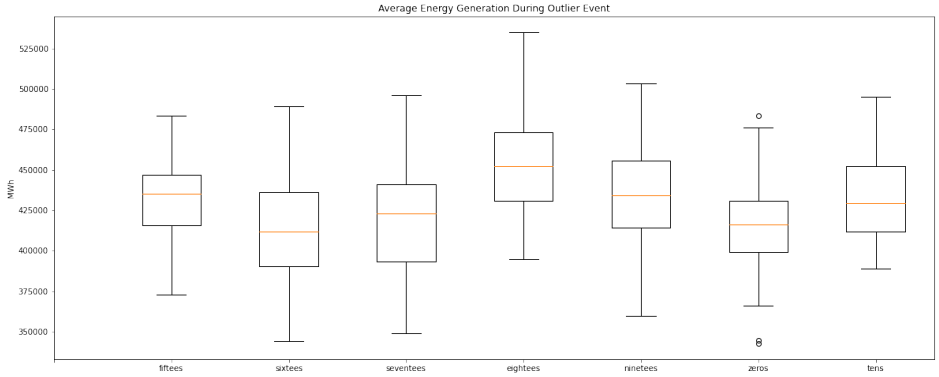


Figure 32: As figure 30. With top 50 outliers per decade

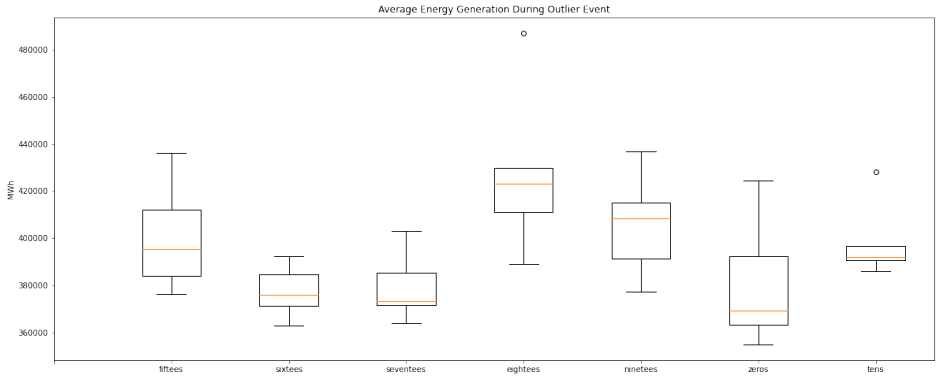


Figure 33: Boxplots of the average energy generation during an outlier event per decade. Using the Unbiased Kullback-Leibler divergence measure. Top 5 outliers per decade were analyzed

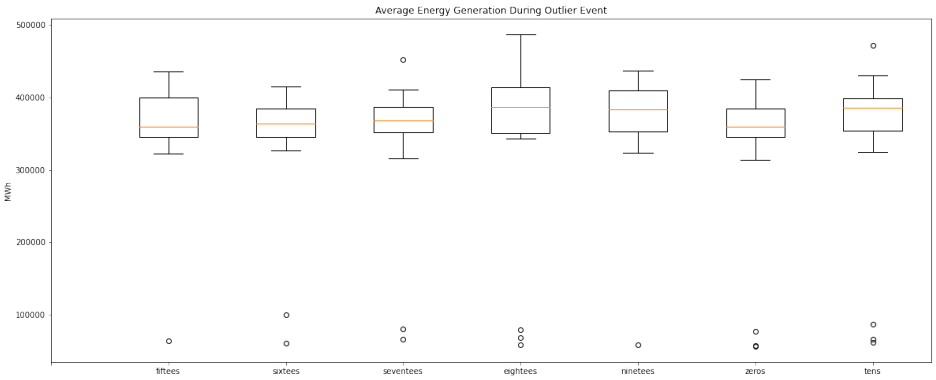


Figure 34: As figure 33. With top 20 outliers per decade

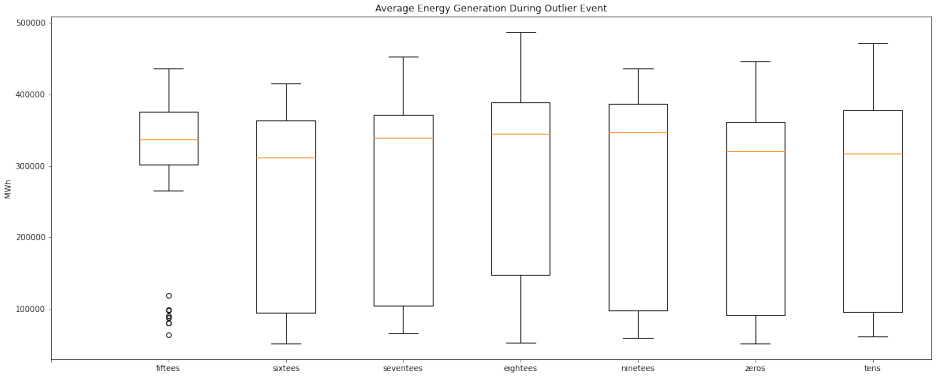


Figure 35: As figure 33. With top 50 outliers per decade.

While there is no linear trend visible there does appear to be some periodical behaviour influencing the outlier events. This periodic behaviour appears in all combinations of top number of outliers investigated and divergence measure used. In the top 50 for Unbiased Kullback-Leibler the effect is hard to see due to the spread of the outlier intensity. It is best observed in the averages of the top 50 Cross Entropy boxplot. Similar behavior of multidecadel variability in German wind energy generation was found by Wohland et al. [91].

These result shows that the multidecadel variability needs to be taken into account by policy makers as they influence the most extreme situations, allowing for even more intense outlier events.

In addition we see from these results that when looking at the univariate total energy generation we see mostly peaks. The number of outliers where the average generation during the outlier, was greater then the average of the overall energy generation is shown in table 6.

We see that troughs in the univariate total energy generation are difficult to

Decade	CE Peaks	CE Troughs	U-KL Peaks	U-KL Troughs
50-60	50	0	41	9
60-70	50	0	30	20
70-80	50	0	32	18
80-90	50	0	37	13
90-00	50	0	34	16
00-10	50	0	32	18
10-20	50	0	29	21

Table 6: Number of peaks and troughs in the top 50 outliers per decade.

detect using the MDI algorithm. Using the Cross Entropy divergence measure only peaks are detected and a potentially risky situation as in figure 17 remains undetected. This indicates the importance of selecting the correct divergence measure as well as what variables to investigate.

Outlier Duration

To determine if outlier events are lasting longer throughout the years, the average interval length of the top 50 outliers is investigated. These results are depicted in figures 36 and 37, where boxplots of the average length for the Cross Entropy and Unbiased Kullback-Leibler divergence measures are presented.

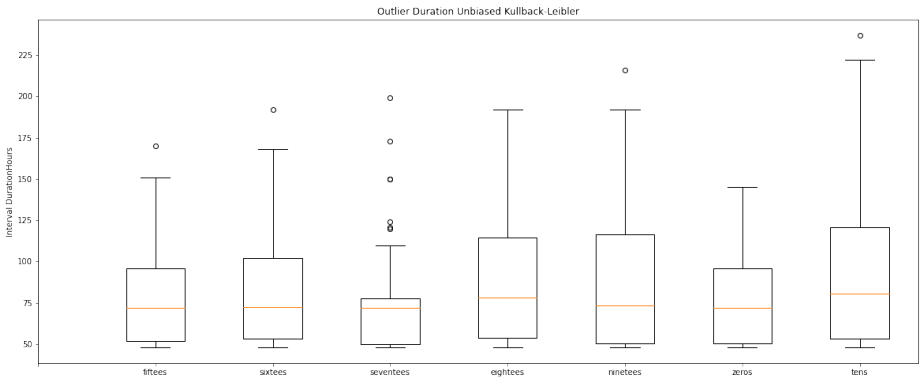


Figure 36: Boxplots of the average interval length in hours per decade. Results for the Cross Entropy divergence measure.

Figure 37: As figure 36 for Unbiased Kullback-Leibler divergence measure

We found no indication that the interval length changes significantly throughout the years. We do note that, as seen in the other experiments, the top outliers detected using Cross Entropy are relatively short and those detected using Unbiased Kullback-Leibler last longer. The method used has significant impact on the length of the detected outlier. This bias might make it more difficult to detect any change in the length of the exact outlying event.

Outlier Occurrence

To detects shifts in when outliers occur we investigated in what months the top outliers occur throughout the decades. For these results we counted how many of the top outliers occur during each month. If an outlier spans multiple months it was counted towards the month containing the majority of the outlier. Figures 38, 39 show these monthly counts per decade.

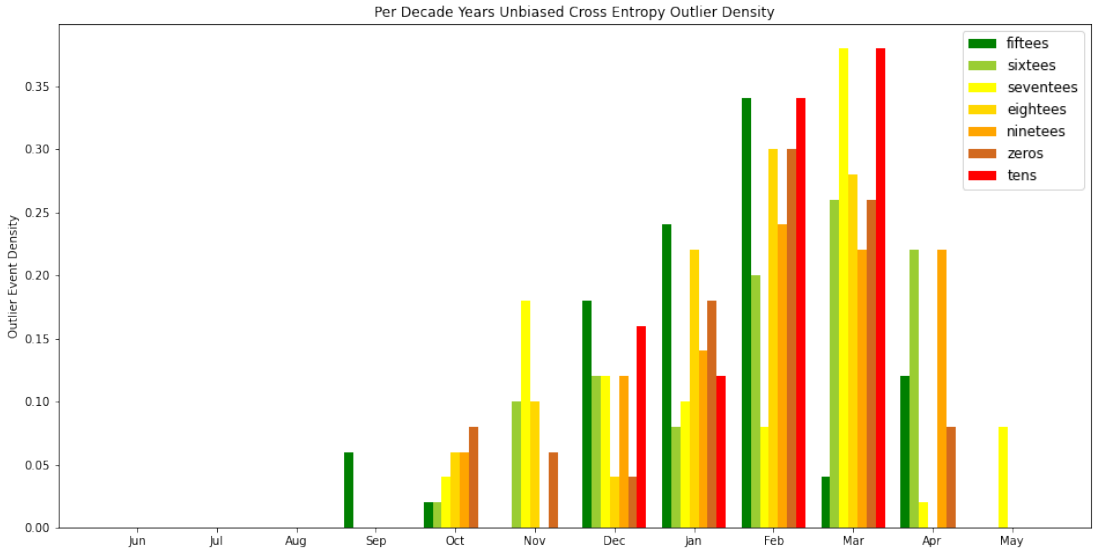


Figure 38: Event Density of the top 50 outlier events per decade. Calculated using the Cross Entropy divergence measure.

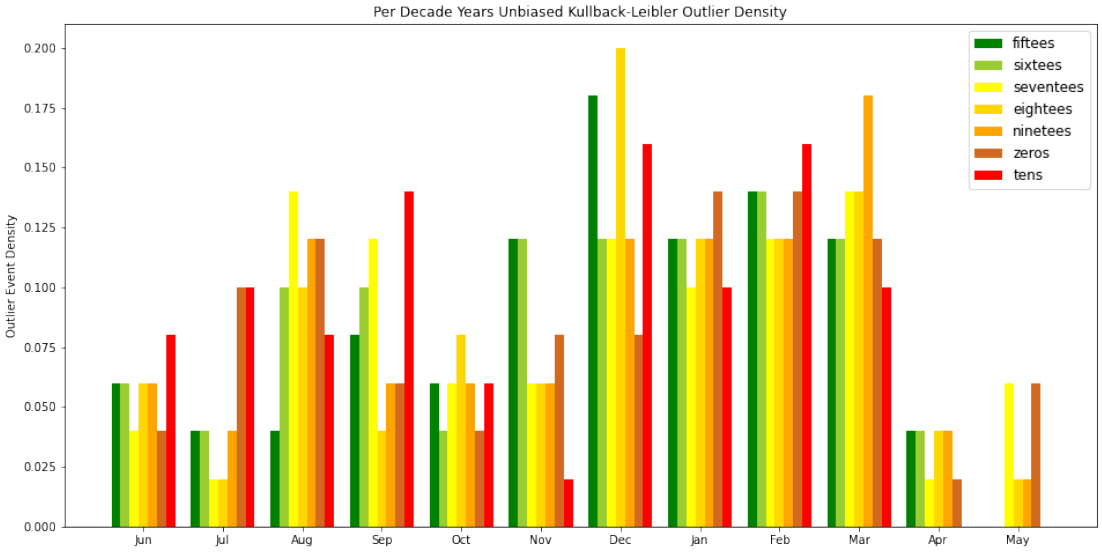


Figure 39: As figure 38 for Unbiased Kullback-Leibler divergence measure

The monthly count per 20 years can be seen in figures 40 and 41. The per 20 year aggregate was used to see if there were any trends that are more subtle that might not show up in the per decade analysis, or trends that take effect over a longer temporal period.

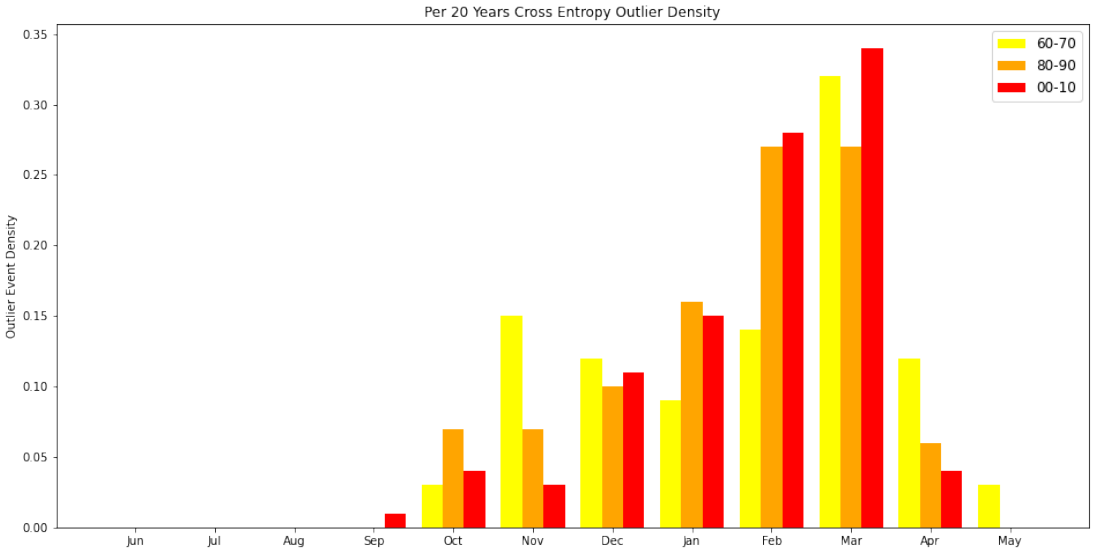


Figure 40: Event Density of the top 100 outlier events per 20 years. Calculated using the Cross Entropy divergence measure.

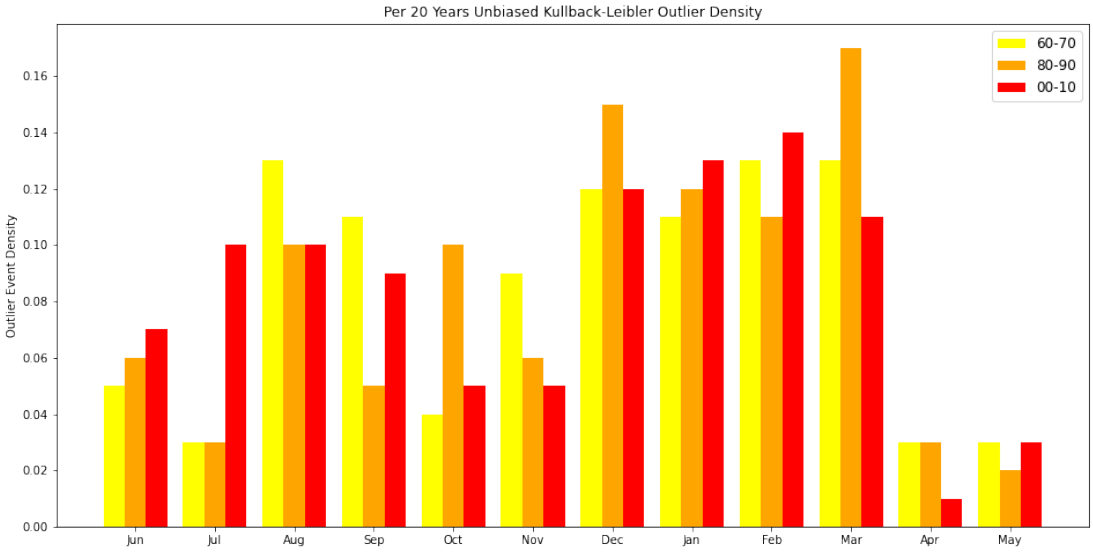


Figure 41: As figure 38 for Unbiased Kullback-Leibler divergence measure

The results do not contain any obvious trends or patterns that we could potentially attribute to climate change.

To recap, we have seen that multidecadel variability impacts the intensity of outlier events. We noted no significant change in the duration of outliers throughout the decades. This might be caused by biases in the methods used. There appears to be no shift in when outlier occur throughout the years.

6.3 Spatial Location Experiment

The goal of this experiment is to test the capabilities of the MDI algorithm to detect temporal outliers, and their spatial location, and to see if this provides any additional insight into the detected outlier.

For this experiment we used the Western Region with spatial and temporal data. We used the same parameter values for the MDI algorithm as in the previous experiments. No spatial context embedding was used. This means that the results of this experiment will be temporal outliers, and their spatial location.

Early checking of the usable parameter values shows that we need to leverage the closed form solutions of the Gaussian distribution here. Experiments that took hours using Kernel Density Estimation to estimate the distribution, took only seconds using the Gaussian distribution. As early tests for this experiment with the Gaussian distribution already took hours of wall clock time, the use of the KDE model for this experiment is not feasible.

We looked for outliers at least two days long, and at most seven days. We looked for events at least 250 KM across the latitude and longitude, and at most 750 KM. This is approximately 2.5 degree up to 7.5 degree. Since this experiment uses spatial-temporal data it is memory intensive. The spatial resolution was reduced to 0.5 degree grids, and the temporal resolution to two hours. In addition we used a single variable, the sum of WON, WOF and SPV. Only a single decade was analyzed, 2010-2019. This was done to save memory usage. This setup used just shy of 100 GB of RAM.

The MDI algorithm took approximately 20 minutes of wall clock time for both models, Cross Entropy and Unbiased Kullback-Leibler. For both divergence scoring methods the top outliers are shown in figures 42 and 43. For each outlier several time steps are shown.

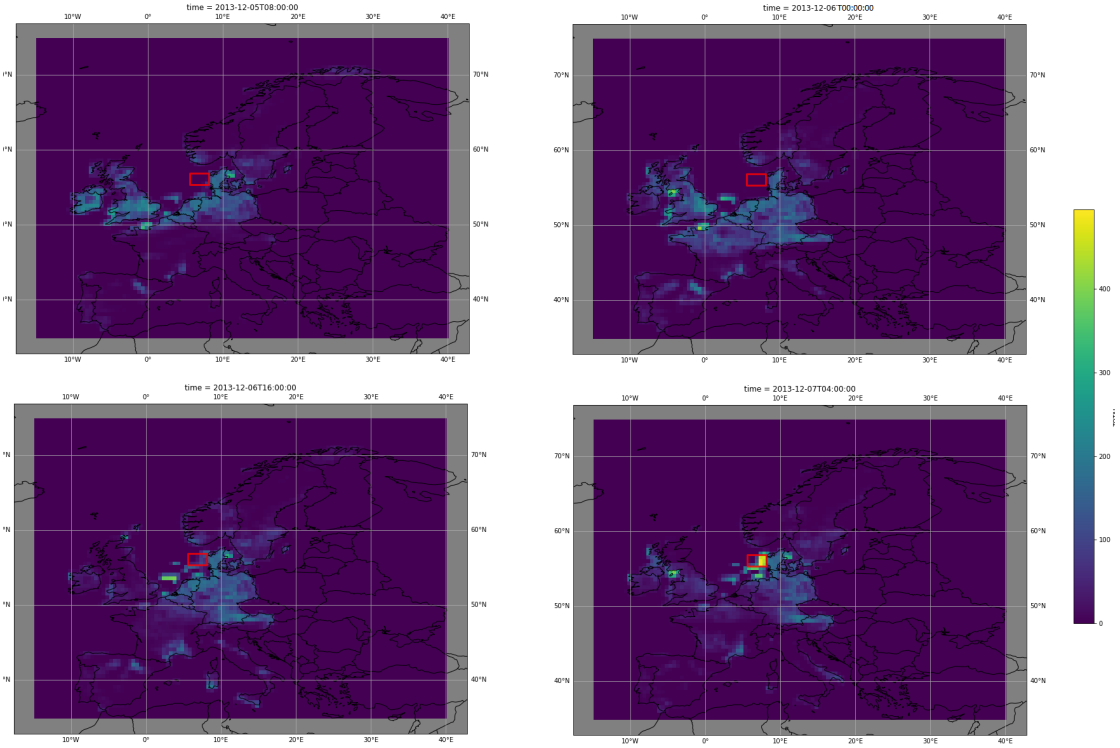


Figure 42: Four time steps of the top temporal outlier detected using Cross Entropy on spatial temporal data from 2010 to 2019. Spatial location of the outlier highlighted. Time steps are at the start (top left), one third (top right), two thirds (bottom left) and end (bottom right) of the interval.

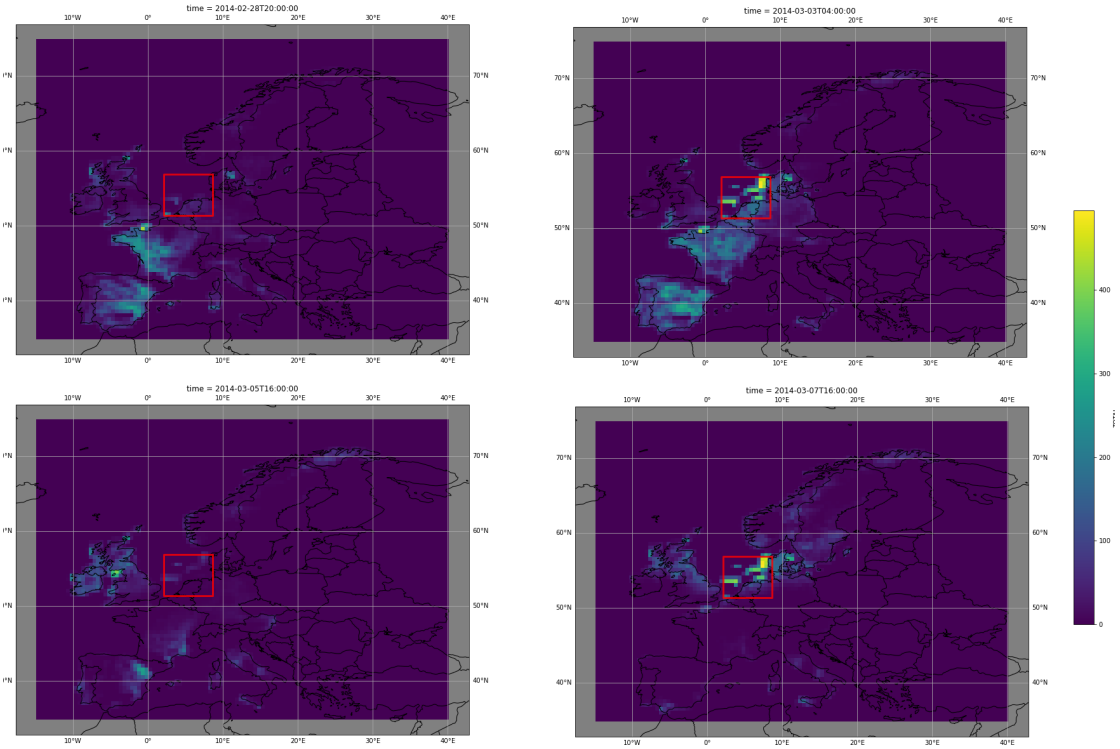


Figure 43: As figure 42, for the top Unbiased Kullback-Leibler outlier.

The regions highlighted in figures 42 and 43 indicate the region, whose estimated distribution diverges the most from the area outside of the highlighted rectangle. We note that all top outliers detected using this method occur in approximately the same area. All Cross Entropy outliers occur in the area highlighted in figure ref 42, with one exception where the outlier region is shifted 0.5 degrees east.

The vast majority of outliers detected using Unbiased Kullback-Leibler divergence occur in the area highlighted in figure 43. The outliers that do not occur in that exact area differ at most 1 degree in latitude and longitude. The spatial and temporal length of the outlying intervals is determined by the method used, as we have seen with the other experiments. All Cross Entropy outliers have a short duration and cover a small area, whereas the Unbiased Kullback-Leibler outliers have a longer duration, and cover a larger area.

We thus see that the area containing the offshore wind energy generation in the North Sea dominate the outliers detected in this experiment. Even when looking at the total energy generation per grid, it appears that the contribution of offshore wind energy is the most important factor when detecting outliers. This might be caused by the inherit spatial bias present in our data. Since energy generation is tied to physical generators this determines the potential a grid cell has. We note that offshore find energy generation is highly concentrated, inflating the total energy generation in specific grids. This is in line with what we have seen in the regional experiments.

These grids thus often contain high energy generation values, including the maximum present in our dataset. This combined with the fact that low energy generation values are very prominent in our dataset, as can be seen in figure 44, causes the areas in the North Sea containing offshore wind energy generation to consistently produce the top outliers.

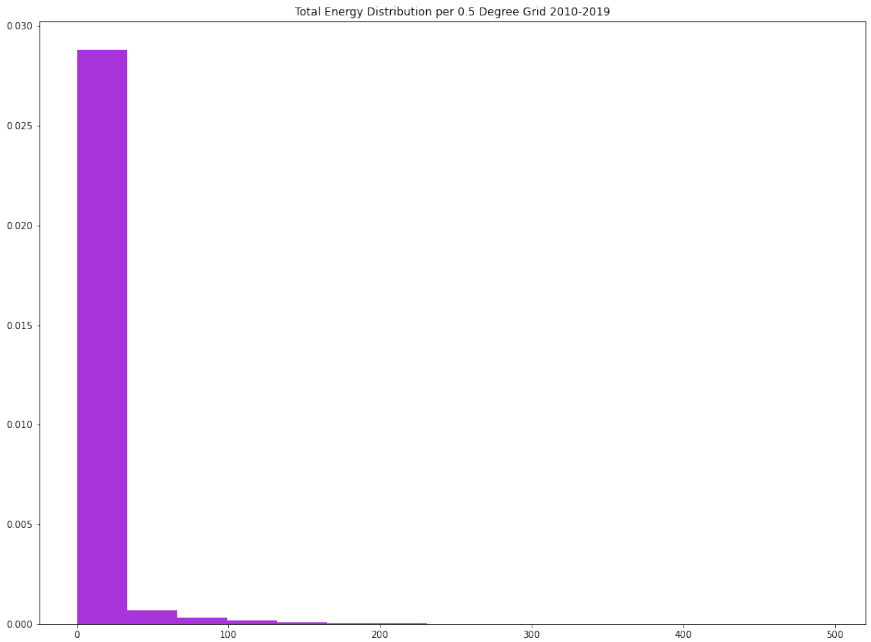


Figure 44: Distribution of the total energy per grid for 0.5 degree grids over the period 2010-2019

7 Conclusion

The goal of this research is to answer the question: "How can outlier detection algorithms be used to improve the assessment of future energy systems?". We subdivided this question into three sub-questions:

1. What adaptations to the algorithms need to be made to apply state-of-the-art spatial-temporal outlier detection algorithms to energy system data?
2. What are the discovered spatial-temporal outliers and do they pose a threat to energy security?
3. Do current levels of climate change have an effect on outlier events, and if so, what is this effect?

To answer the first sub-question, a literature study was performed and we found two promising algorithms. These algorithms couldn't directly be applied to our data. The scale of the problem and size of the data makes it so that only part of the results achieved by Barz et al. (2018) [15] are usable.

Notably the spatial context embedding is so memory intensive that it has no practical use for the data used for this research. Even without the spatial context embedding, the size of the data limited the possible divergence measures and distributions. For the experiments on spatial-temporal data we were forced to use the Gaussian distribution to leverage the closed form solutions.

The SLOM algorithm by Chawla and Sun (2006) [16] focuses on single grid points that divert from their neighbors. To find interesting outlying regions rather than single grid points a post-processing method is proposed. The post-processing method also tracks the regions through time.

In order to find spatial-temporal outliers, a temporal first approach was used. The MDI algorithm found temporal outliers, that were investigated further using the SLOM algorithm. To detect outliers 70 years of hourly historical reanalysis data, ERA5, was used.

The need for preprocessing of the data was investigated. We used clustering as a way to find related and spatially coherent regions to demonstrate the use of our approach for such use cases. The impact of seasonal behaviour was investigated by removing seasonality by way of hourly Z scores. We found that the removing of seasonality only reduced the quality of discovered outliers.

Our temporal first outlier detection approach was used to answer the second sub-question: "What are the discovered spatial-temporal outliers and do they pose a threat to energy security?".

We found that the spatial step in our approach, based on SLOM with additional post-processing, yielded no results. This could be caused by the high spatial resolution of the problem, smoothing out the differences a grid cell has with its neighbors. Additionally, it is likely that the temporal outliers discovered are strictly temporal. As these outliers are the top outliers over long temporal periods they are likely to affect the whole spatial range. The data also has an underlying spatial bias. All

energy generation happens at a physical location, and depends on the installed capacity. This means that there are certain areas, notably in the North Sea, where there are grids with a very high installed capacity, right next to areas with very little installed capacity. This means that those areas are likely to contain the biggest spatial outliers.

This means we only found interesting temporal outliers. We found that the choice of variables used and method had a significant impact on the type of detected outlier. Outliers detected using Cross Entropy had shorter intervals, and Unbiased Kullback-Leibler outlier intervals had a longer duration. There appears to be a bias in the divergence measure with regards to interval length. The choice of variables and divergence scoring method also had an impact on whether peaks in energy generation, or troughs were detected.

Several different classes of temporal outliers were found. These classes, Summer Deficiencies, Winter Surpluses, Wind Droughts, Summer High SPV and Winter Low SPV all have potential risks associated with them. The Summer Deficiencies, Wind Droughts and Winter Low SPV classes are all potential low energy generation risk situations, as discussed in section 2.2.

The other classes, Winter Surpluses and Summer High SPV, present potential economic risks. These high energy generation events could require the discharge of unused energy to ensure that the frequency of the European Grid remains stable. This discharging of unused energy causes wear on the generators without gaining any energy. These classes do not match any of the situations mentioned in section 2.2, and were not considered beforehand.

The second potentially risky situation identified in section 2.2, low energy generation with regards to demand, could not be detected due to the lack of available regional demand data.

The final situation that has potential risks mentioned in section 2.2, sufficient energy generation with local deficiencies that need to be compensated by spatially distant energy surpluses, could not be detected due to the strictly temporal nature of the discovered outliers. Since the top temporal outliers all appear to affect the entire region, no such spatial differences were detected. The top temporal outliers over 70 years are very likely to cover the entire region under investigation.

The Spatial Location experiment discussed in section 6.3, provided no additional insights in spatial-temporal outliers. The temporal outliers that were discovered, and their spatial location, were very similar for all top outliers. This is caused by an inherit spatial bias in the data caused by the underlying physical energy generators and their location.

An experiment was performed in order to detect if current levels of climate change have an effect on outlier events, which is the final sub-question. The top outliers per decade were investigated, looking at the total energy generation.

No shift in duration of the outliers was found, as there appears to be some bias in the used methods with regards to the outlier length.

The intensity of the outliers did not appear to have a trend that can be attributed to climate change. We did however see some periodic behaviour. This multidecadel variability appears to influence the intensity of outlier events, and needs to be taken

into account by policy makers when making decisions about the future energy grid.

When looking at when outlier events occur throughout the decades there did not appear to be a shift in when the top outliers occur.

Resulting from the Climate Change Experiment we found no trends or patterns that can be clearly attributed to climate change. This could be caused by the fact that the historical period analyzed only shows a 1 °C increase in temperature, while an increase of 3 °C or more is expected by the end of the century.

From both the literature study and the experiments we note that size of the data, and the scale of the problem are a limiting factor in both the approach and the execution of the algorithms.

With the sub-questions answered we now look back at the research question: "how can outlier detection algorithms be used to improve the assessment of future energy systems?". We conclude that outlier detection algorithms can be used to improve the assessment of future energy systems. The main contribution of outlier detection is highlighting of periods or events that need to be accounted for when making decisions on the future energy grid, and its assessment.

In addition, the detected outliers can be used to improve future energy scenarios, as outliers such as the ones we detected need to be incorporated in these scenarios. It is also possible to use the outliers to create synthetic data, to test future energy grid scenarios, and how they deal with extreme events.

8 Discussion

In this section we discuss the challenges when doing this type of research and the potentials for future research.

8.1 Challenges

One of the biggest challenges of this research is the size of our dataset. Due to the connectivity and complexity of the European energy system, it is difficult to reduce the resolution and scale of the data without losing too much information. The size severely limits possible methods and algorithms. Leveraging of multi-threading is also required for viable computation times. This added significant problems due to the fact that the built in multi-threading xarray uses based on DASK did not work on our system.

Since we used a temporal first approach, the majority of the computational work was performed by the MDI algorithm. One of the problems with this algorithm's current implementation is that it is not possible to accurately track its progress. The current implementation hands each thread a chunk of the data, and looks for outliers in that chunk. However per chunk it is unknown how many potentially interesting intervals there will be proposed. It only checks for the next interval that should be checked. Only approximate progress can be checked, but the proposed intervals don't have to be distributed uniformly. Such an approximate indication will thus be far from a constant progress. This was confirmed in correspondence with the original author, Björn Barz. We followed his advice to run the MDI algorithm on small instances of the data, and extrapolated the maximally viable parameter values.

Another major challenge was the evaluation of the results, since we are looking for the biggest outliers, without a priori knowing what these are in our dataset. This lack of ground truth makes it difficult to fine-tune the algorithm parameter values to find optimal results. This often required subjective decisions on what we presumed to be interesting.

Additionally, the evaluation of the results was rather labor intensive. For each experiment there are top MDI outliers. Each of these outliers has top SLOM outliers. Since we were not looking for very specific things, and were exploring what these methods offer in terms of new insights, there was no automatic classification. When applying this work to a real life scenario such an automatic classification is highly advised.

The spatial bias present in the data presented another challenge. As energy generation is tied to physical locations where the generators are installed, there is spatial bias in the high and low energy generation areas. The methods we used did not compensate for that. Any research performed on similar data needs to take this into account. Simply averaging per grid does not work, as this would make an area where 1 MWh of energy is produced as important as an area where 500 MWh is produced.

8.2 Future Research

During the research we came across several interesting questions and avenues of research that were left unexplored. One of them relates to one of the original ideas behind this research, the second situation posing risks, mentioned in section 2.2. Looking for outliers in energy generation data compared to demand is very interesting, but unfortunately no suitable regional energy demand dataset is publicly available. Once such a dataset is available research into outliers with regards to demand is interesting future research. The methods developed in this work would be suitable for such a research.

We noted during our experiments that the divergence scoring methods have some bias with regards to the length of the detected outlier interval. In the original paper by Barz et al. [15] the authors note that Jiang et al. [92] provide a similar derivation an Unbiased Kullback-Leibler divergence. The difference is that Barz et al. make the assumption that the derivation for infinite length intervals holds for finite intervals. Jiang et al. do not make this assumption, and find an Unbiased Kullback-Leibler divergence measure without the factor 2. Using the Unbiased Kullback-Leibler divergence as derived by Jiang et al. [92] could reduce the bias the method has on the detected interval length.

For this research we focused on a temporal first approach. The high spatial resolution makes it computationally difficult to compare the work with a spatial first approach. Comparing such approaches on data with a lower spatial resolution is potentially interesting work.

For the regional experiment on the full region we experimented with using Z scores to remove seasonal behaviour. We found that this added nothing interesting. We do however see that these outliers have very high scores. Perhaps the divergence scores are blown up due to the Z scores not being limited between -1 and 1. The results could potentially be improved by normalizing the Z scores so that they are limited to this range. This and other methods for removing seasonal behaviour are interesting to investigate further.

The Climate Change experiment could be investigated further. A potential trend might become visible when looking at the top outliers of certain classes. For example, looking at the top short peaks or long troughs might highlight trends not visible when looking at all top outliers. Looking at all outliers above some score threshold is also interesting, as it might indicate an increase or decrease in outliers throughout the decades. Looking at the top outlier events per year might also provide insight in any potential trends.

Investigating the predictability of the outliers detected by our methods could be an interesting continuation of this research. The prediction of such outliers would allow for time to prepare for the event. This could negate the impact such an outlier event would have on the European energy system.

Datasets Containing Future Projections

Seeing how we were able to detect outlier events that are concerning for energy security it is interesting to apply our method to datasets containing models of the future. These can then be used to gain better understanding of the extremes and

risks that the future energy grid faces.

The analysis of future energy security can be performed on either the EURO-CORDEX [93] or PRIMAVERA [94] data sets. These data sets contain scenarios that are based on that by the year 2100 the average temperature has increased by approximately $3°C$. The EURO-CORDEX data runs from 2006 to 2100 and has a spatial resolution of $\pm$ 50 km, and an hourly temporal resolution. The PRIMAVERA future projection data runs from 2015 to 2050. It has a spatial resolution of $\pm$ 25 km and a 3-hourly temporal resolution.

For a study of the potential impact of climate change both the EURO-CORDEX and PRIMAVERA data sets might be suitable. Since both are based on scenarios that include a temperature increase, these represent effects caused by climate change.

Additionally, if a study of steady state climate scenarios that differ on temperature is required, HiWAVES3 data [95] is available. This data set contains 2000 years of historical simulation data, and multiple future scenarios. These scenarios contain data up to 2000 years into the future, and represent different increases in average temperature due to climate change, such as $2°C$ and $3°C$. HiWAVES3 has a temporal resolution of six hours and spatial resolution of $\pm$80 kilometers.

Acknowledgements

The data used in the experiment contains modified Copernicus Climate Change Service information 2020. https://doi.org/10.24381/cds.adbb2d47.

This research received funding from the Netherlands Organisation for Scientific Research (NWO) under grant number 647.003.005.

References

[1] Joeri Rogelj, Malte Meinshausen, and Reto Knutti. "Global warming under old and new scenarios using IPCC climate sensitivity range estimates". In: *Nature climate change* 2.4 (2012), pp. 248–253.

[2] Rajendra K Pachauri, Myles R Allen, Vicente R Barros, John Broome, Wolfgang Cramer, Renate Christ, John A Church, Leon Clarke, Qin Dahe, Purnamita Dasgupta, et al. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC, 2014.

[3] United Nations Framework Convention on Climate Change. *Ratification Status Paris Agreement*. URL: https://unfccc.int/process-and-meetings/the-paris-agreement/what-is-the-paris-agreement (visited on 08/11/2020).

[4] United Nations Framework Convention on Climate Change. *Paris Agreement*. UN Treaty. United Nations, Dec. 2015. URL: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en (visited on 08/11/2020).

[5] European Commission. *European Green Deal*. European Union, Dec. 2019. URL: https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF (visited on 08/11/2020).

[6] Iain Staffell and Stefan Pfenninger. "The increasing impact of weather on electricity supply and demand". In: *Energy* 145 (2018), pp. 65–78.

[7] HE Thornton, Brian J Hoskins, and AA Scaife. "The role of temperature in the variability and extremes of electricity and gas demand in Great Britain". In: *Environmental Research Letters* 11.11 (2016), p. 114015.

[8] Marie Bessec and Julien Fouquau. "The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach". In: *Energy Economics* 30.5 (2008), pp. 2705–2721.

[9] Marianne Zeyringer, James Price, Birgit Fais, Pei-Hao Li, and Ed Sharp. "Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather". In: *Nature Energy* 3.5 (2018), pp. 395–403.

[10] K. van der Wiel, L.P. Stoop, B.R.H. van Zuijlen, R. Blackport, M.A. van den Broek, and F.M. Selten. "Meteorological conditions leading to extreme low variable renewable energy production and extreme high energy shortfall". In: *Renewable and Sustainable Energy Reviews* 111 (2019), pp. 261–275. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2019.04.065. URL: http://www.sciencedirect.com/science/article/pii/S1364032119302862.

[11] Charlotte Neubacher, Jan Wohland, and Dirk Witthaut. "Multi-decadal offshore wind power variability can be mitigated through optimized European allocation". In: *EGU General Assembly Conference Abstracts*. 2020, p. 9538.

[12] J. Zscheischler, B. Van Den Hurk, P. J. Ward, and S. Westra. *Multivariate extremes and compound events. In Climate Extremes and Their Implications for Impact and Risk Assessment*. Elsevier, 2029, pp. 59–76.

[13] Erik Rodner, Björn Barz, Yanira Guanche, Milan Flach, Miguel Mahecha, Paul Bodesheim, Markus Reichstein, and Joachim Denzler. "Maximally divergent intervals for anomaly detection". In: *arXiv preprint arXiv:1610.06761* (2016).

[14] Björn Barz, Yanira Guanche Garcia, Erik Rodner, and Joachim Denzler. "Maximally divergent intervals for extreme weather event detection". In: *OCEANS 2017-Aberdeen*. IEEE. 2017, pp. 1–9.

[15] Björn Barz, Erik Rodner, Yanira Guanche Garcia, and Joachim Denzler. "Detecting regions of maximal divergence for spatio-temporal anomaly detection". In: *IEEE transactions on pattern analysis and machine intelligence* 41.5 (2018), pp. 1088–1101.

[16] Sanjay Chawla and Pei Sun. "SLOM: a new measure for local spatial outliers". In: *Knowledge and Information Systems* 9.4 (2006), pp. 412–429.

[17] M Hekkenberg, RMJ Benders, HC Moll, and AJM Schoot Uiterkamp. "Indications for a changing electricity demand pattern: The temperature dependence of electricity demand in the Netherlands". In: *Energy Policy* 37.4 (2009), pp. 1542–1551.

[18] Morna Isaac and Detlef P Van Vuuren. "Modeling global residential sector energy demand for heating and air conditioning in the context of climate change". In: *Energy policy* 37.2 (2009), pp. 507–521.

[19] Jonas Hörsch, Fabian Hofmann, David Schlachtberger, and Tom Brown. "PyPSA-Eur: An open optimisation model of the European transmission system". In: *Energy Strategy Reviews* 22 (2018), pp. 207–215. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2018.08.012. URL: http://www.sciencedirect.com/science/article/pii/S2211467X18300804.

[20] Swissgrid. *Swiss Grid Frequency Monitoring*. URL: https://www.swissgrid.ch/en/home/operation/regulation/frequency.html (visited on 01/12/2021).

[21] Jonas Hörsch and Tom Brown. "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios". In: *2017 14th international conference on the European Energy Market (EEM)*. IEEE. 2017, pp. 1–7.

[22] Internation Energy Agency. *World Energy Outlook 2019*. URL: https://www.iea.org/reports/world-energy-outlook-2019 (visited on 08/21/2020).

[23] Thomas Petermann, Harald Bradke, Arne Lüllmann, Maik Poetzsch, and Ulrich Riehm. *What happens during a blackout: Consequences of a prolonged and wide-ranging power outage*. BoD–Books on Demand, 2014.

[24] Tom Brown, David Schlachtberger, Alexander Kies, Stefan Schramm, and Martin Greiner. "Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system". In: *Energy* 160 (2018), pp. 720–739.

[25] R.H. Wuijts, M.A. van den Broek, and J.M. van den Akker. "Effect of Modeling Choices in the Unit Commitment Problem". In: *Applied Energy* (2021), Submitted.

[26] Adriaan P Hilbers, David J Brayshaw, and Axel Gandy. "Importance subsampling: improving power system planning under climate-based uncertainty". In: *Applied Energy* 251 (2019), p. 113114.

[27] Laura Dawkins, Isabel Rushby, Andrew Dobbie, Emily Wallace, and Tom Butcher. "Characterising Adverse Weather for the UK Electricity System". In: (2020).

[28] Copernicus. *ERA5 documentation, Contains modified Copernicus Climate Change Service Information 2020*. URL: https://doi.org/10.24381/cds.adbb2d47 (visited on 08/31/2020).

[29] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), pp. 1999–2049. DOI: 10.1002/qj.3803. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803.

[30] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and Thépaut J-N. "ERA5 hourly data on single levels from 1979 to present." In: *Climate Data Store (CDS), (Accessed on 19-06-2019)* (2018). DOI: 10.24381/cds.adbb2d47.

[31] Waldo R Tobler. "A computer movie simulating urban growth in the Detroit region". In: *Economic geography* 46.sup1 (1970), pp. 234–240.

[32] James H Faghmous and Vipin Kumar. "Spatio-temporal data mining for climate data: Advances, challenges, and opportunities". In: *Data mining and knowledge discovery for big data*. Springer, 2014, pp. 83–116.

[33] B. Bell, H. Hersbach, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz Sabater, J. Nicolas, R. Radu, D. Schepers, A. Simmons, C. Soci, and J-N. Thépaut. "ERA5 monthly averaged data on single levels from 1950 to 1978 (preliminary version)." In: *Climate Data Store (CDS), (Accessed on 10-11-2020)* (2020). URL: https://cds.climate.copernicus-climate.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means-preliminary-back-extension?tab=overview.

[34] European Centre for Medium-Range Weather Forecasts. *ERA5 back-extension tropical storms are too intense*. URL: https://confluence.ecmwf.int/display/CKB/ERA5+back+extension+1950-1978+%28Preliminary+version%29%3A+tropical+cyclones+are+too+intense (visited on 08/31/2020).

[35] Bas van Zuijlen, William Zappa, Wim Turkenburg, Gerard van der Schrier, and Machteld van den Broek. "Cost-optimal reliable power generation in a deep decarbonisation future". In: *Applied Energy* 253.July (2019), p. 113587. ISSN: 03062619. DOI: 10.1016/j.apenergy.2019.113587. URL: https://doi.org/10.1016/j.apenergy.2019.113587.

[36] S. Jerez, F. Thais, I. Tobin, M. Wild, A. Colette, P. Yiou, and R. Vautard. "The CLIMIX model: A tool to create and evaluate spatially-resolved scenarios of photovoltaic and wind power development". In: *Renewable and Sustainable Energy Reviews* 42 (2015), pp. 1–15. ISSN: 13640321. DOI: 10.1016/j.rser.2014.09.041. URL: http://dx.doi.org/10.1016/j.rser.2014.09.041.

[37] Yves Marie Saint-Drenan, Romain Besseau, Malte Jansen, Iain Staffell, Alberto Troccoli, Laurent Dubus, Johannes Schmidt, Katharina Gruber, Sofia G. Simões, and Siegfried Heier. "A parametric model for wind turbine power curves incorporating environmental conditions". In: *Renewable Energy* 157 (2020), pp. 754–768. ISSN: 18790682. DOI: 10.1016/j.renene.2020.04.123.

[38] C. Carrillo, A. F. Obando Montaño, J. Cidrás, and E. Díaz-Dorado. "Review of power curve modelling for windturbines". In: *Renewable and Sustainable Energy Reviews* 21 (2013), pp. 572–581. ISSN: 13640321. DOI: 10.1016/j.rser.2013.01.012.

[39] Iratxe Gonzalez Aparicio, Andreas Zucker, Francesco Careri, Fabio Monforti, Thomas Huld, and Jake Badger. *EMHIRES dataset Part I: Wind power generation*. 2016. ISBN: 9789279631931. DOI: 10.2790/831549. URL: https://setis.ec.europa.eu/related-jrc-activities/jrc-setis-reports/emhires-dataset-part-i-wind-power-generation.

[40] P. Ruiz, W. Nijs, D. Tarvydas, A. Sgobbi, A. Zucker, R. Pilli, R. Jonsson, A. Camia, C. Thiel, C. Hoyer-Klick, F. Dalla Longa, T. Kober, J. Badger, P. Volker, B. S. Elbersen, A. Brosowski, and D. Thrän. "ENSPRESO - an open, EU-28 wide, transparent and coherent database of wind, solar and biomass energy potentials". In: *Energy Strategy Reviews* 26.June 2019 (2019), p. 100379. ISSN: 2211467X. DOI: 10.1016/j.esr.2019.100379. URL: https://doi.org/10.1016/j.esr.2019.100379.

[41] L.P. Stoop, M.A. Van den Broek, and A. Feelders. "Quasi-realistic energy conversion models for wind and solar energy resource". Unpublished Manuscript. 2021.

[42] Philip E. Bett and Hazel E. Thornton. "The climatological relationships between wind and solar energy supply in Britain". In: *Renewable Energy* 87 (2016), pp. 96–110. DOI: 10.1016/j.renene.2015.10.006. URL: http://dx.doi.org/10.1016/j.renene.2015.10.006.

[43] Victoria Hodge and Jim Austin. "A survey of outlier detection methodologies". In: *Artificial intelligence review* 22.2 (2004), pp. 85–126.

[44] Hongzhi Wang, Mohamed Jaward Bah, and Mohamed Hammad. "Progress in outlier detection techniques: A survey". In: *IEEE Access* 7 (2019), pp. 107964–108000.

[45] Manish Gupta, Jing Gao, Charu C Aggarwal, and Jiawei Han. "Outlier detection for temporal data: A survey". In: *IEEE Transactions on Knowledge and data Engineering* 26.9 (2013), pp. 2250–2267.

[46] Daniel B Neill and Andrew W Moore. "A fast multi-resolution method for detection of significant spatial disease clusters". In: *Advances in Neural Information Processing Systems*. 2004, pp. 651–658.

[47] Yan Shi, Min Deng, Xuexi Yang, and Jianya Gong. "Detecting anomalies in spatio-temporal flow data by constructing dynamic neighbourhoods". In: *Computers, Environment and Urban Systems* 67 (2018), pp. 80–96.

[48] Chang-Tien Lu and Lily R Liang. "Wavelet fuzzy classification for detecting and tracking region outliers in meteorological data". In: *Proceedings of the 12th annual ACM international workshop on Geographic information systems*. 2004, pp. 258–265.

[49] Nabil R Adam, Vandana Pursnani Janeja, and Vijayalakshmi Atluri. "Neighborhood based detection of anomalies in high dimensional spatio-temporal sensor datasets". In: *Proceedings of the 2004 ACM symposium on Applied computing*. 2004, pp. 576–583.

[50] Douglas M Hawkins. *Identification of outliers*. Vol. 11. Springer, 1980.

[51] Charu C Aggarwal. *Data mining: the textbook*. Springer, 2015.

[52] Jorma Laurikkala, Martti Juhola, Erna Kentala, N Lavrac, S Miksch, and B Kavsek. "Informal identification of outliers in medical data". In: *Fifth international workshop on intelligent data analysis in medicine and pharmacology*. Vol. 1. 2000, pp. 20–24.

[53] Martin Kulldorff. "A spatial scan statistic". In: *Communications in Statistics-Theory and methods* 26.6 (1997), pp. 1481–1496.

[54] Antonio Loureiro, Luis Torgo, and Carlos Soares. "Outlier detection using clustering methods: a data cleaning application". In: *Proceedings of KDNet Symposium on Knowledge-based systems for the Public Sector*. Springer Bonn. 2004.

[55] Rik Warren, Robert F Smith, and Anne K Cybenko. *Use of Mahalanobis distance for detecting outliers and outlier clusters in markedly non-normal data: a vehicular traffic example*. Tech. rep. SRA INTERNATIONAL INC DAYTON OH, 2011.

[56] Richard A Bauder and Taghi M Khoshgoftaar. "Multivariate outlier detection in medicare claims payments applying probabilistic programming methods". In: *Health Services and Outcomes Research Methodology* 17.3-4 (2017), pp. 256–289.

[57] Koen Smets and Jilles Vreeken. "The odd one out: Identifying and characterising anomalies". In: *Proceedings of the 2011 SIAM international conference on data mining*. SIAM. 2011, pp. 804–815.

[58] Arno Siebes, Jilles Vreeken, and Matthijs van Leeuwen. "Item sets that compress". In: *Proceedings of the 2006 SIAM International Conference on Data Mining*. SIAM. 2006, pp. 395–406.

[59] Ville Hautamaki, Ismo Karkkainen, and Pasi Franti. "Outlier detection using k-nearest neighbour graph". In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* Vol. 3. IEEE. 2004, pp. 430–433.

[60] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. "LOF: identifying density-based local outliers". In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000, pp. 93–104.

[61] Spiros Papadimitriou, Hiroyuki Kitagawa, Phillip B Gibbons, and Christos Faloutsos. "Loci: Fast outlier detection using the local correlation integral". In: *Proceedings 19th international conference on data engineering (Cat. No. 03CH37405)*. IEEE. 2003, pp. 315–326.

[62] Jian Tang, Zhixiang Chen, Ada Wai-Chee Fu, and David W Cheung. "Enhancing effectiveness of outlier detections for low density patterns". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer. 2002, pp. 535–548.

[63] Ke Zhang, Marcus Hutter, and Huidong Jin. "A new local distance-based outlier detection approach for scattered real-world data". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer. 2009, pp. 813–822.

[64] Auroop R Ganguly and Karsten Steinhaeuser. "Data mining for climate change and impacts". In: *2008 IEEE international conference on data mining workshops*. IEEE. 2008, pp. 385–394.

[65] Elizabeth Wu, Wei Liu, and Sanjay Chawla. "Spatio-temporal outlier detection in precipitation data". In: *International Workshop on Knowledge Discovery from Sensor Data*. Springer. 2008, pp. 115–133.

[66] Alp Kut and Derya Birant. "Spatio-temporal outlier detection in large databases". In: *Journal of computing and information technology* 14.4 (2006), pp. 291–297.

[67] Maria Bala Duggimpudi, Shaaban Abbady, Jian Chen, and Vijay V Raghavan. "Spatio-temporal outlier detection algorithms based on computing behavioral outlierness factor". In: *Data & Knowledge Engineering* 122 (2019), pp. 1–24.

[68] James P Rogers, Daniel Barbara, and Carlotta Domeniconi. "Detecting spatio-temporal outliers with kernels and statistical testing". In: *2009 17th International Conference on Geoinformatics*. IEEE. 2009, pp. 1–6.

[69] Alex Gammerman and Volodya Vovk. "Prediction algorithms and confidence measures based on algorithmic randomness theory". In: *Theoretical Computer Science* 287.1 (2002), pp. 209–217.

[70] Tao Cheng and Zhilin Li. "A multiscale approach for spatio-temporal outlier detection". In: *Transactions in GIS* 10.2 (2006), pp. 253–263.

[71] Qi Liu, Rudy Klucik, Chao Chen, Glenn Grant, David Gallaher, Qin Lv, and Li Shang. "Unsupervised detection of contextual anomaly in remotely sensed data". In: *Remote Sensing of Environment* 202 (2017), pp. 75–87.

[72] Carolina Centeio Jorge, Martin Atzmueller, Behzad M Heravi, Jenny L Gibson, Cláudio Rebelo de Sá, and Rosaldo JF Rossetti. "Mining Exceptional Social Behaviour". In: *EPIA Conference on Artificial Intelligence*. Springer. 2019, pp. 460–472.

[73] Paul Viola and Michael J Jones. "Robust real-time face detection". In: *International journal of computer vision* 57.2 (2004), pp. 137–154.

[74] Jiusun Zeng, Uwe Kruger, Jaap Geluk, Xun Wang, and Lei Xie. "Detecting abnormal situations using the Kullback–Leibler divergence". In: *Automatica* 50.11 (2014), pp. 2777–2786.

[75] Holger Kantz and Thomas Schreiber. *Nonlinear time series analysis*. Vol. 7. Cambridge university press, 2004.

[76] Skipper Seabold and Josef Perktold. "statsmodels: Econometric and statistical modeling with python". In: *9th Python in Science Conference*. 2010.

[77] Harold Hotelling. "The generalization of Student's ratio". In: *Breakthroughs in statistics*. Springer, 1992, pp. 54–65.

[78] Karin van der Wiel, Laurens P Stoop, BRH Van Zuijlen, Russell Blackport, MA Van den Broek, and FM Selten. "Meteorological conditions leading to extreme low variable renewable energy production and extreme high energy shortfall". In: *Renewable and Sustainable Energy Reviews* 111 (2019), pp. 261–275.

[79] Charles Elkan. "Using the triangle inequality to accelerate k-means". In: *Proceedings of the 20th international conference on Machine Learning (ICML-03)*. 2003, pp. 147–153.

[80] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[81] Fionn Murtagh and Pierre Legendre. "Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?" In: *Journal of classification* 31.3 (2014), pp. 274–295.

[82] Ayman Taha and Ali S Hadi. "Anomaly detection methods for categorical data: A review". In: *ACM Computing Surveys (CSUR)* 52.2 (2019), pp. 1–35.

[83] Tomilayo Komolafe, A Valeria Quevedo, Srijan Sengupta, and William H Woodall. "Statistical evaluation of spectral methods for anomaly detection in static networks". In: *Network Science* 7.3 (2019), pp. 319–352.

[84] Erich Schubert, Arthur Zimek, and Hans-Peter Kriegel. "Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection". In: *Data mining and knowledge discovery* 28.1 (2014), pp. 190–237.

[85] S. Hoyer and J. Hamman. "xarray: N-D labeled arrays and datasets in Python". In: *Journal of Open Research Software* 5.1 (2017). DOI: 10.5334/jors.148. URL: http://doi.org/10.5334/jors.148.

[86] Dask Development Team. *Dask: Library for dynamic task scheduling*. 2016. URL: https://dask.org.

[87] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D.S. Seljebotn, and K. Smith. "Cython: The Best of Both Worlds". In: *Computing in Science Engineering* 13.2 (Mar. 2011), pp. 31–39. ISSN: 1521-9615. DOI: 10.1109/MCSE.2010.118.

[88] RM Singh, J Yu, and G Podger. "Efficient NetCDF processing for big datasets". In: *22nd International Congress on Modelling and Simulation* ().

[89] KNMI. *Heaviest Storm in Decades*. URL: https://www.knmi.nl/kennis-en-datacentrum/uitleg/zwaarste-storm-in-decennia (visited on 01/22/2021).

[90] Rijnmond. *Dag van Toen: De storm van 25 januari 1990*. URL: https://www.rijnmond.nl/nieuws/124973/Dag-van-Toen-De-storm-van-25-januari-1990 (visited on 01/22/2021).

[91] Jan Wohland, Nour Eddine Omrani, Noel Keenlyside, and Dirk Witthaut. "Significant multidecadal variability in German wind energy generation". In: *Wind Energy Science* 4.3 (2019), pp. 515–526.

[92] Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi, Shiqiang Yang, and Christos Faloutsos. "A general suspiciousness metric for dense blocks in multimodal data". In: *2015 IEEE International Conference on Data Mining*. IEEE. 2015, pp. 781–786.

[93] Daniela Jacob, Juliane Petersen, Bastian Eggert, Antoinette Alias, Ole Bøssing Christensen, Laurens M Bouwer, Alain Braun, Augustin Colette, Michel Déqué, Goran Georgievski, et al. "EURO-CORDEX: new high-resolution climate change projections for European impact research". In: *Regional environmental change* 14.2 (2014), pp. 563–578.

[94] R. J. Haarsma, M. J. Roberts, P. L. Vidale, C. A. Senior, A. Bellucci, Q. Bao, P. Chang, S. Corti, N. S. Fučkar, V. Guemas, J. von Hardenberg, W. Hazeleger, C. Kodama, T. Koenigk, L. R. Leung, J. Lu, J.-J. Luo, J. Mao, M. S. Mizielinski, R. Mizuta, P. Nobre, M. Satoh, E. Scoccimarro, T. Semmler, J. Small, and J.-S. von Storch. "High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6". In: *Geoscientific Model Development* 9.11 (2016), pp. 4185–4208. DOI: 10.5194/gmd-9-4185-2016. URL: https://gmd.copernicus.org/articles/9/4185/2016/.

[95] Karin van der Wiel, Frank M Selten, Richard Bintanja, Russell Blackport, and James A Screen. "Ensemble climate-impact modelling: extreme impacts from moderate meteorological conditions". In: *Environmental Research Letters* 15.3 (Mar. 2020), p. 034050. DOI: 10.1088/1748-9326/ab7668. URL: https://doi.org/10.1088%2F1748-9326%2Fab7668.