

Maximally Divergent Intervals for Extreme Weather Event Detection

Björn Barz, Yanira Guanache Garcia, Erik Rodner and Joachim Denzler
Computer Vision Group
Friedrich Schiller University Jena
Jena, Germany
Email: bjoern.barz@uni-jena.de

Abstract—We approach the task of detecting anomalous or extreme events in multivariate spatio-temporal climate data using an unsupervised machine learning algorithm for detection of anomalous intervals in time-series. In contrast to many existing algorithms for outlier and anomaly detection, our method does not search for point-wise anomalies, but for contiguous anomalous intervals. We demonstrate the suitability of our approach through numerous experiments on climate data, including detection of hurricanes, North Sea storms, and low-pressure fields.

I. INTRODUCTION

The identification of anomalous intervals in time-series data is an important task in a variety of domains and of particular interest when dealing with climate data, since such anomalies usually relate to extreme weather events like heat waves, droughts, or cyclones, which often entail significant destruction, insurance costs or even deaths. Such events are relevant from a retrospect as well, e.g., for studying regularities or long-term trends of weather phenomena. A well-known example for this is the discovery of the correlation between the *El Niño* weather phenomenon and extreme surface pressures over the equator by Gilbert Walker [1] during the early 20th century through the analysis of extreme events in time-series of climate data.

However, manually locating anomalies and rare events in large-scale multivariate and perhaps even spatio-temporal data is like looking for a needle in a haystack and an even more difficult task for humans if the data consist of multiple different variables.

In order to automate the localization of extreme events, we propose to apply a novel machine learning algorithm for detection of anomalous intervals to the task of detecting storms and hurricanes using measurements of marine climate variables. Previous works in the field of anomaly detection have primarily focused on detecting single isolated points in the data as outliers, i.e. measurement errors or noise [2]–[4]. Anomalies driven by natural processes, however, are more likely to occur during a longer period of time and may even be unidentifiable by looking at a single measurement, but only by a contiguous collection of samples that is anomalous as a whole (a so-called *collective anomaly*). Thus, analysts will intuitively be searching for anomalous *intervals* in the data instead of anomalous points and the algorithm assisting them

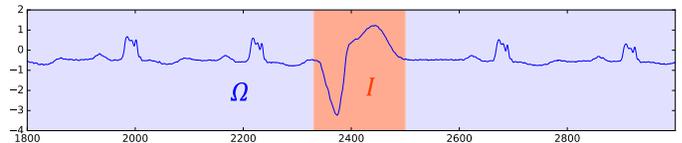


Fig. 1. Schematic illustration of the principle of the MDI algorithm: The distribution of the data in the inner interval I is compared with the distribution of the remaining time-series in the outer interval Ω .

should do so as well. In addition, applying point-wise anomaly detection methods to large data sets will most likely lead to an unmanageable amount of detections. Grouping nearby detections is not a good approach either, since the anomaly scores of the samples may vary widely, which will lead to disrupted detections. However, research on the detection of anomalous intervals in time-series has been very limited.

In this work, we propose to use the novel Maximally Divergent Intervals (MDI) algorithm for detection of extreme events in climate data. This algorithm leverages machine learning techniques to search for anomalous intervals instead of isolated points in multivariate time-series data and can work in a completely unsupervised scenario, i.e., without involving any domain-specific prior knowledge.

The remainder of this paper is organized as follows: We briefly review related work on detection of anomalous intervals in spatio-temporal data in section II and introduce the basics of the Maximally Divergent Intervals (MDI) algorithm in section III, along with some adaptations which were necessary to make it suitable for the task of detecting extreme weather events. Application examples on several datasets are demonstrated in section IV. Section V concludes this paper.

II. RELATED WORK

Despite the large amount of work on detection of outliers and single anomalous points, for which an extensive survey has been conducted by Chandola et al. [5], the task of *collective* anomaly detection, however, has only received marginal attention in literature. This is surprising, since anomalies driven by a natural process are unlikely to occur at a single isolated point of time only. Some approaches dealing with climate data try to mitigate this issue by downsampling the data to a coarse resolution and/or manually merging nearby detections which seem to belong to the same event [6]. Clearly, such approaches

that use point-wise detectors, but are actually interested in intervals, are very error-prone and unsatisfactory.

Despite the sparsity of suitable approaches, the problem of detecting anomalous intervals has been known for a couple of years. Keogh et al. [7] have already tackled this task in 2005 with a method they called ‘‘HOT SAX’’. They try to find anomalous sub-sequences (‘‘discords’’) of time-series by representing all possible sub-sequences of length d as a d -dimensional vector and using the Euclidean distance to the nearest neighbor in that space as anomaly score. However, this method is limited to univariate data and a fixed length of the intervals must be specified in advance, making it a clearly sub-optimal approach for applications dealing with often multivariate climate data and events of varying length.

More similar to our objective, a recent work of Jiang et al. [8] measures the divergence between probability distributions to search for anomalous blocks in multi-modal tensors (an analogy to spatio-temporal data and our method). However, their approach is designed for discrete data only (e.g., relations in social networks) and uses a Poisson distribution to model the data. Since their search strategy for anomalous blocks is very specific to applications dealing with graph data, it is not applicable in the general case for multivariate continuous data dealt with in our work.

With respect to our pursued application, i.e., detecting extreme events in climate data, Mínguez et al. [9], for example, investigate how regression models can be used to detect outliers corresponding to hurricanes in hindcast datasets based on the residuals of the data or the model parameters. However, such an approach is strictly limited to the detection of single points.

In contrast, a recent approach of Liu et al. [10] localizes hurricanes spatially using convolutional neural networks. Being a supervised approach, this method requires annotated training data which is often difficult to obtain. This is in opposition to our approach, which does not need any training data and is, in consequence, not restricted to certain kinds of anomalies defined by the annotations. This allows us to also detect anomalies that have not been expected. Moreover, the approach of Liu et al. [10] requires the size of the regions to be fixed in advance and does not consider the temporal extent of the anomalies at all.

In spatio-temporal data, however, anomalies may also move over time, which cannot be captured by regular sub-blocks. Wu et al. [11] follow a sequential approach for detecting such anomalies in precipitation data first spatially, then temporally and apply a merge-strategy afterwards. For the detection of spatial outliers, they scan over all possible regions at each time-step and search for the top k regions with the highest value of the *Kulldorff Spatial Scan Statistic (KSS Statistic)*. They then associate the outliers across contiguous time-steps using the following algorithm: Expand each anomalous region at time t by a given amount of cells and search for ‘‘children’’ at time $t + 1$ that are completely contained in the expanded region. This way, an anomaly tree is built, associating each anomaly with multiple children at the next time. In a final

step they extract each possible sequence of moving anomalous spatio-temporal regions from that tree.

Two major drawbacks of this approach are, that the complexity of the extraction of all possible sequences of spatial outliers from the tree is exponential in the time dimension and that the KSS Statistic is limited to binary-valued data. Wu et al. [11] use a simple threshold operation to binarize climate data, which can be seen as a very simple outlier detection technique itself. In addition, their approach only finds anomalies that are both spatial and temporal anomalies, whereas we are more interested in determining the spatial location of temporal anomalies. Our approach, in contrast, is able to deal with multivariate real-valued data and does not search for temporal and spatial anomalies separately, but treats time and space jointly.

III. THE MAXIMALLY DIVERGENT INTERVALS ALGORITHM FOR ANOMALY DETECTION

In this section, we briefly revisit the recently proposed Maximally Divergent Intervals (MDI) algorithm [12] for detection of anomalous intervals in time-series and propose some crucial adaptations and extensions needed for a successful application to extreme weather event detection. In particular, we propose a modification to the divergence measure used by the algorithm to make it independent of the size of the intervals being compared. Moreover, we extend the algorithm, which was initially designed to deal with non-spatial time-series, to spatio-temporal data and investigate how to deal with missing values in the data, which is a common problem when dealing with real data obtained from multiple error-prone sensors.

A. Detecting Intervals of Maximal Divergence

Given a series $(x_t)_{t=1}^n$, $x_t \in \mathbb{R}^d$, of n timesteps, the MDI algorithm searches for intervals $I = \{t \in \mathbb{N} \mid a \leq t < b\} = [a, b)$ whose data distribution is most different from the distribution of the data in the remainder of the time-series $\Omega = [1, n] \setminus I$ (cf. fig. 1). The Kullback-Leibler (KL) divergence is employed for quantifying the degree of deviation between the two distributions p_I and p_Ω :

$$\text{KL}(p_I, p_\Omega) = \int p_I(x_t) \cdot \log \left(\frac{p_I(x_t)}{p_\Omega(x_t)} \right) dx_t . \quad (1)$$

The probability density functions $p_I = \mathcal{N}(\mu_I, S_I)$ and $p_\Omega = \mathcal{N}(\mu_\Omega, S_\Omega)$ are approximated by fitting a multivariate normal distribution (a ‘‘Gaussian’’) to the data in the two intervals I and Ω , respectively. This model comes along with the advantage that there is a known closed-form solution for the KL divergence of two Gaussians [13], allowing for explicit computation of (1):

$$\text{KL}(p_I, p_\Omega) = \frac{1}{2} \left((\mu_\Omega - \mu_I)^\top S_\Omega^{-1} (\mu_\Omega - \mu_I) + \text{trace} \left(S_\Omega^{-1} S_I \right) + \log \frac{|S_\Omega|}{|S_I|} - d \right) . \quad (2)$$

To mitigate the inadequate assumption of this model that all points in the time-series are independent from each other,

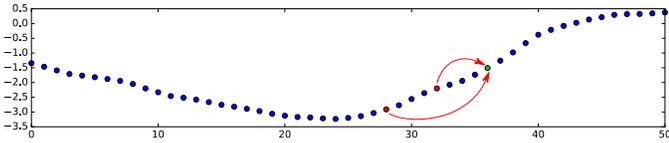


Fig. 2. Illustration of time-delay embedding with $\kappa = 3, \tau = 4$. The attribute vector of each sample is augmented with the attributes of the samples 4 and 8 time steps earlier.

time-delay embedding [14] is applied as a pre-processing step in order to obtain a modified time-series $(x'_t)_{t=1+(\kappa-1)\tau}^n$, $x'_t \in \mathbb{R}^{\kappa d}$, where each sample x'_t incorporates attributes from some previous time-steps as context:

$$x'_t = \left(x_t^\top \quad x_{t-\tau}^\top \quad x_{t-2\tau}^\top \quad \cdots \quad x_{t-(\kappa-1)\tau}^\top \right)^\top. \quad (3)$$

The number of aggregated samples κ is called *embedding dimension* and the distance between two consecutive context samples τ is called the *time lag*. An illustrative example is given in fig. 2.

After having computed the KL divergence for every pair of intervals I, Ω with user-specified constraints on the minimum and maximum size of I , the algorithm returns a desired number of top k detections with the highest KL divergence. Non-maximum suppression is applied to obtain non-overlapping intervals only.

The MDI algorithm is, in principle, an unsupervised technique, i.e., it does not require any training data and is, hence, a very versatile method that can be applied to different data from various domains. However, some prior knowledge about the task at hand is at least useful to achieve optimal performance. In particular, restricting the size of the anomalous intervals to be searched for to reasonable limits in advance will lead to a significant reduction of the search space and, thus, faster detection.

Furthermore, the quality of the resulting detections is often quite sensitive to the parameters of time-delay embedding: A larger embedding dimension κ will provide the MDI algorithm with more context information, but also increases the dimensionality of the data, requiring more data per interval for a robust estimation of the distribution parameters. The time-lag τ can be used to take more context into account without increasing the embedding dimension by skipping some time-steps between context samples, but a too large time-lag will hence increase the risk of missing important information. Thus, a suitable trade-off has to be found. However, a satisfactory approach for automatic determination of these parameters does not yet exist, since their optimal values vary widely depending on both the data and the application [14]. Therefore, application-specific prior knowledge is beneficial in this regard as well.

B. Dealing with Spatio-Temporal Data

In order to detect anomalies in climate data, which usually do not only possess a temporal, but also spatial contextual attributes, the MDI algorithm needs to be extended to be

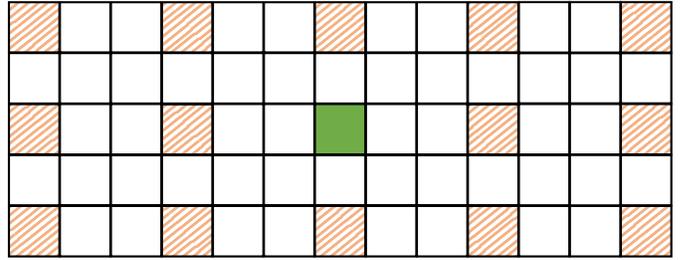


Fig. 3. Exemplary illustration of spatial-neighbor embedding with $\kappa_x = 3, \kappa_y = 2, \tau_x = 3, \tau_y = 2$. The attribute vector at the location with a solid fill color is augmented with the attributes of the samples with a striped pattern.

capable of handling spatio-temporal data. Instead of searching for anomalous intervals along the time axis only, it has to search for anomalous *blocks* in a data tensor with both temporal and spatial context.

This extension is straight-forward, since the probability density model employed by the MDI algorithm already assumed independence of the data across different time-steps and fixed this assumption by incorporating temporal context into each sample as a pre-processing step. Analogously, we extend this independence assumption to the spatial axes and augment each sample not only with the attributes at previous points of time, but also with the values at neighboring locations (cf. fig. 3). This pre-processing step, which we refer to as *spatial-neighbor embedding*, is parametrized with 3 parameters $\kappa_x, \kappa_y, \kappa_z$ for the embedding dimension along each spatial axis and 3 parameters τ_x, τ_y, τ_z for the lag along each axis.

Note that, in contrast to time-delay embedding, neighbors from both directions are aggregated, since spatial context is bilinear. Thus, $\kappa_x = 3$, for example, would mean to consider 4 neighboring coordinates along the x -axis, 2 in each direction.

Spatial-neighbor embedding can either be applied before or after time-delay embedding. As opposed to many spatio-temporal anomaly detection approaches that perform temporal and spatial anomaly detection sequentially (e.g., [11], [15], [16]), the MDI algorithm in combination with the two embeddings allows for a joint optimization. However, it implies a much more drastic multiplication of the data size.

C. An unbiased KL divergence

During initial experiments with the MDI algorithm on real data, we discovered that the Kullback-Leibler divergence proposed by [12] usually leads to detections of the minimum allowed size, so that larger anomalies are split up across multiple consecutive detections (see fig. 4a for an example).

This bias towards smaller intervals is due to the different amount of samples involved in the estimation of the distribution of the data in the intervals. Under the assumption of data being sampled from a Gaussian, the KL divergence becomes a random variable as well, but its mean depends on the length of the interval I .

To remedy this issue we leverage theory about statistical tests, since the objective of assessing the “divergence” between

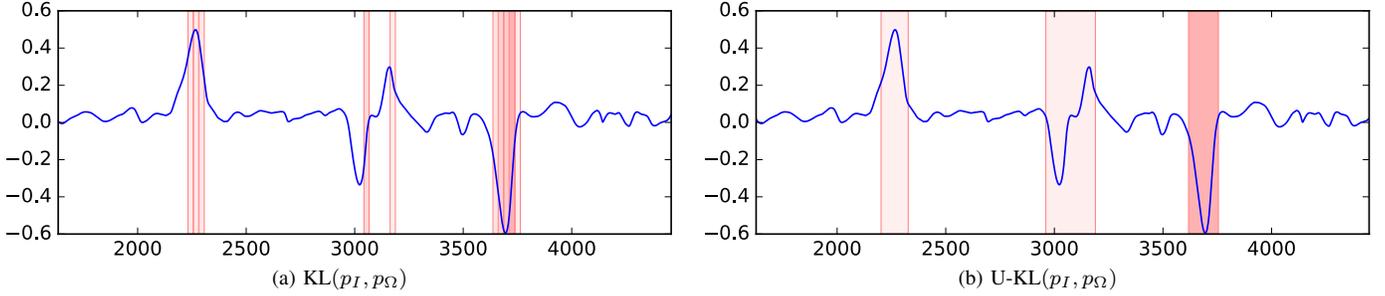


Fig. 4. Top 10 detections obtained from the KL divergence on a real time-series (left) and top 3 detections obtained from the unbiased KL divergence on the same time-series (right). This example illustrates the phenomenon of several contiguous minimum-size detections when using the original KL divergence (note the thin lines between the single detections in the left plot). The MDI algorithm has been applied with a time-delay embedding of $\kappa = 3, \tau = 1$ and the size of the intervals to analyze has been limited to be between 25 and 250 time-steps.

the data distribution in an inner interval I and an outer interval Ω can also be interpreted as a statistical test: Under the asymptotic assumption of very long time-series, the estimated distribution p_Ω converges towards the true distribution of the data in Ω , which can, thus, be assumed to be known. The null hypothesis of the test would be that the data in I has been sampled from the same distribution. A certain test statistic may then be used as a measure for how well the data in the interval I fit the model established based on the data in the remainder of the time-series [17].

We define an *unbiased KL divergence* that is equivalent to this test statistic and corrects the bias of the KL divergence through a multiplication with the number of samples involved in estimating p_I (i.e., the size of the interval):

$$\text{U-KL}(p_I, p_\Omega) := 2 \cdot |I| \cdot \text{KL}(p_I, p_\Omega) . \quad (4)$$

The mean value of this divergence can be shown to depend on the number of attributes d only [18] and is independent of the size of I and Ω . Though that is only true under the asymptotic assumption of infinitely long time-series, this simple correction may also be useful for time-series of finite length. An example of actual detections resulting from the use of the unbiased KL divergence compared with the original one can be seen in fig. 4.

D. Missing Values

One restriction of the MDI algorithm is that its embedding mechanisms assume regular spacing between the values of the contextual attributes. That means, time-steps have to be equidistant and spatial locations have to be organized in a regular grid. However, real datasets often contain “missing values”, i.e., samples for which a value is not available, e.g., due to sensor failures.

Since the probability density models employed by the MDI algorithm do not take the contextual attributes into account at all, samples with missing values for some attributes can just be ignored during probability density estimation under certain conditions. This procedure of ignoring entire samples with missing values for parameter estimation is known as *listwise deletion* or *complete case analysis* and is only valid under the assumption, that the data is missing at random (MAR) [19].

That means, whether a value is missing or not is independent of the value itself (though it may depend on the values of other, observed variables). This condition is, for example, not met if a value is missing from the data because it was out of the range that can be measured by the sensor. In this case, parameter estimation ignoring missing values is likely to be extremely biased and the missing data mechanism has to be modeled as part of the estimation process [19].

Though listwise deletion is considered to be one of the safest methods for dealing with missing values in the MAR scenario [19], it discards a potentially large amount of usable data and, thus, often leads to a severe loss of statistical power. In case of the MDI algorithm, this may become particularly problematic due to the application of time-delay and spatial-neighbor embedding, so that a single missing value can spread to multiple samples. Entire samples with missing values in their context would, therefore, be ignored, though their remaining attributes may be valid.

However, approaches for parameter estimation exploiting all available data in spite of missing values often involve iterative procedures for data imputation [20]. Such methods are too time-consuming for being used as part of the MDI algorithm. Therefore, we rely on the simple listwise deletion scheme, which is not problematic as long as missing values are scarce.

E. Software and Graphical User Interface

Since the detection of anomalies in time-series is an important task in many fields, the software performing this task should be employable by non-computer scientists as well. Of particular importance for a good usability is the availability of a graphical user interface (GUI) that does not only assist the user in loading the data and setting the parameters, but also in the analysis of the results.

In order to make the MDI algorithm available to a larger audience, we provide an efficient implementation as C++ library that can be accessed via a convenient Python interface.

For non-spatial time-series, a graphical user interface (GUI) is available, allowing for comfortable experimentation with custom time-series data and the various parameters of the MDI algorithm (see fig. 5). After running the algorithm, the resulting detections are directly visualized in the GUI and their

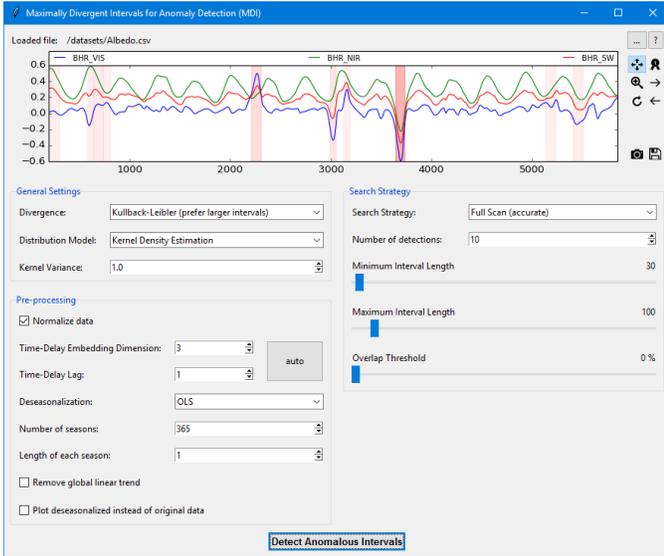


Fig. 5. Graphical User Interface

confidence is indicated by varying intensities of the fill color. The user may zoom and pan that interactive graph as well as traverse through the detections in their respective order using the corresponding tool buttons. Of course, not only the graph, but also the detections themselves may be exported for further analysis with other software.

The software is made available as open source and can be obtained at <https://cvjena.github.io/libmaxdiv/>.

IV. DETECTION OF EXTREME WEATHER EVENTS

In this section we explore how the MDI algorithm can be used for detection of extreme weather events in climate data, beginning with hurricane detection based on a plain time-series measured at a single location, continuing with storm detection in a spatio-temporal dataset covering the southern North Sea, and finally showing an application to detection of low-pressure areas in data of a much wider spatial extent.

In all our experiments, we use the unbiased KL divergence proposed in section III-C.

A. Hurricanes

First, we try to detect extreme events in a purely temporal time-series without spatial information. We use meteocean data (significant wave height, H_s , wind speed, W , and sea level pressure, SLP) measured at a location near the Bahamas in the Atlantic Sea (23.866° N, 68.481° W). Six months of hourly data, from June 2012 until November 2012, were extracted from the National Data Buoy Center from the NOAA¹. This period corresponds to the Atlantic hurricane season, which was particularly active in that year with 19 tropical cyclones (winds above 52 km/h), where 10 of them became hurricanes (winds above 64 km/h).

Since this data is non-spatial, it can be conveniently processed with our GUI without having to write a single line of

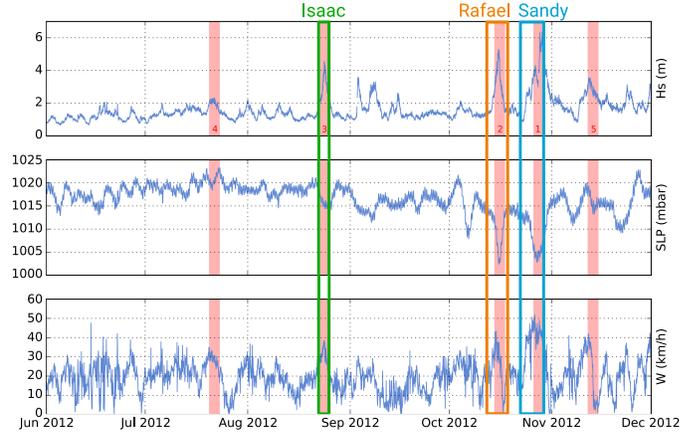


Fig. 6. Hurricane detections. Colored outlines represent historical hurricanes and filled red areas are the top 5 intervals detected by the MDI algorithm. Small red numbers indicate the ranking of the detections.

code. Since the data contains missing values at 6 time-steps, this experiment also demonstrates that such a scenario can be handled by the MDI algorithm as well.

We have restricted the size of the intervals to be searched for to be between 12 and 72 hours and applied time-delay embedding with parameters $\kappa = 3$, $\tau = 1$. All variables in the time-series have been normalized individually before running the MDI algorithm by subtracting their mean and dividing by the maximum value.

Figure 6 shows the top 5 detections returned by the algorithm as filled red areas on top of the time-series. The colored outlines represent the official duration of the three main events of that season that passed by near our location, i.e., the hurricanes *Isaac*, *Rafael*, and *Sandy*. The top 3 detections returned by the MDI algorithm correspond quite accurately to those three events. Note that in general the ground-truth areas may be slightly larger than the detections, because they span the entire lifetime of the hurricanes and not just their presence at the Bahamas.

B. North Sea Storms

The time-series used in the previous experiment was rather short. In the following, we apply the MDI algorithm to a much longer time-series for detecting storms over the southern North Sea. For this purpose, the coastDat-1 reanalysis database, provided by the Helmholtz-Zentrum Geesthacht [21], has been used. This dataset comprises marine climate parameters at an hourly resolution over 50 years, from 1958 to 2007 (~450.000 observations). We have selected the area between 53.9° N, 0° E and 56° N, 7.7° E as a subset for this experiment, since it is located entirely over sea. Because cyclones and other storms usually have a large spatial extent and move over the region covered by the measurements, we reduce the spatio-temporal data to purely temporal data in this experiment by averaging over all spatial locations. The variables used for this experiment are significant wave height, mean wave period and wind speed.

¹<http://www.ndbc.noaa.gov/>

Since North Sea storms lasting longer than 3 days are usually considered two independent storms, the maximum length of the possible intervals is set to 72 hours, while the minimum length is set to 12 hours. As before, the parameters of time-delay embedding are fixed to $\kappa = 3$, $\tau = 1$.

We were able to associate 28 out of the top 50 and 7 out of the top 10 detections returned by the algorithm with known historic storms. The highest scoring detection is the well-known ‘‘Hamburg-Flut’’, which flooded one fifth of Hamburg in February 1962 and caused 340 deaths. Also among the top 5 is the so-called ‘‘North Frisian Flood’’, which was a severe surge in November 1981 and led to several dike breaches in Denmark.

Figure 7 shows a heatmap of the three variables under consideration during the middle of the detected timeframes for the top 7 detections. Animated heat maps for the entire time-frame covered by the top detections can be found on our web page: http://www.inf-cv.uni-jena.de/libmaxdiv_applications.html.

A visual inspection of the remaining 22 detections that could not be matched against our database of historic storms revealed, that almost all of them are North Sea storms as well. Only 4 of the top 50 detections are not storms, but the opposite: they span times of extremely calm sea conditions with nearly no wind and very low waves, which is some kind of anomaly as well. Two examples of such detections are shown in fig. 8.

We have also repeated this experiment after applying Z-Score deseasonalization [22], but found it to be neither needed nor useful for the detection of North Sea storms in this scenario. Since such storms are quite common during the winter, but very uncommon during the summer months, deseasonalization would emphasize summer storms. Thus, whether deseasonalization is useful or not does not only depend on the presence of seasonal patterns in the data, but also on the application.

C. Low Pressure Areas

As a genuine spatio-temporal use-case, we have also applied the MDI algorithm to a time-series with daily sea level pressure (SLP) measurements over the North Atlantic Sea with a much wider spatial coverage than in the previous experiment. For this purpose, we selected a subset of the NCEP/NCAR reanalysis [23] covering the years from 1957 to 2011. This results in a comparatively short time-series of about 20,000 days. The spatial resolution of 2.5° degrees is rather coarse and the locations are organized in a regular grid of size 28×17 covering the area between 25° N, 52.5° W and 65° N, 15° E.

Regarding the time dimension, we apply time-delay embedding as usual with $\kappa = 3$, $\tau = 1$ and search for intervals of size between 3 and 10 days. Concerning space, we do not apply any contextual embedding for now and set a minimum size of $7.5^\circ \times 7.5^\circ$, but no maximum. 7 out of the top 20 detections could be associated with known historic storms.

A visual inspection of the results shows that the MDI algorithm is not only capable of detecting occurrences of anomalous low-pressure fields over time, but also their

spatial location. This can be seen in the animations on our web page: http://www.inf-cv.uni-jena.de/libmaxdiv_applications.html. Some snapshots are shown in fig. 9.

It is not necessary to apply spatial-neighbor embedding in this scenario, since we are not interested in spatial outliers, but only in the location of temporal outliers. We have also experimented with applying spatial-neighbor embedding in this scenario and it led to the detection of some high-pressure fields surrounded by low-pressure fields, i.e., spatial outliers, but not necessarily temporal ones. Since high-pressure fields are both larger and more common in this time-series, they are not detected as temporal anomalies. Thus, if one is only interested in the localization of temporal anomalies, but not necessarily in the detection of spatial or spatio-temporal anomalies, spatial-neighbor embedding does not seem to be necessary.

V. CONCLUSION

We have shown how the Maximally Divergent Intervals (MDI) algorithm for anomaly detection can be used to find extreme weather events in climate data in an unsupervised fashion, i.e., without any need for training data or hand-crafted rules. To enable this, we extended the algorithm to be capable of handling spatio-temporal data and proposed an unbiased Kullback-Leibler divergence for anomaly detection, which is not biased with respect to the size of the anomalous intervals.

Experiments on climate data demonstrated that this method can be a useful tool for automatic analysis of environmental data, including detection of hurricanes, storms and low-pressure fields.

However, especially the latter experiment on a spatio-temporal dataset has also exposed some limitations of the current version of the MDI algorithm. It is not able to track anomalies that move in space over time, since it considers rectangular blocks in the data tensor only. For the same reason, it currently cannot handle data with spatial locations not organized in a regular grid. Moreover, scalability is an issue that has to be tackled in future work, since the number of possible intervals to be analyzed grows very quickly in spatio-temporal data.

ACKNOWLEDGMENT

The support of the project EU H2020-EO-2014 project BACI ‘‘Detecting changes in essential ecosystem and biodiversity properties-towards a Biosphere Atmosphere Change Index’’, contract 640176, is gratefully acknowledged.

REFERENCES

- [1] G. Walker, ‘‘World weather,’’ *Quarterly Journal of the Royal Meteorological Society*, vol. 54, no. 226, pp. 79–87, 1928.
- [2] J. F. MacGregor and T. Kourti, ‘‘Statistical process control of multivariate processes,’’ *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995.
- [3] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, ‘‘Lof: identifying density-based local outliers,’’ in *ACM sigmod record*, vol. 29, no. 2. ACM, 2000, pp. 93–104.
- [4] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, ‘‘Estimating the support of a high-dimensional distribution,’’ *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

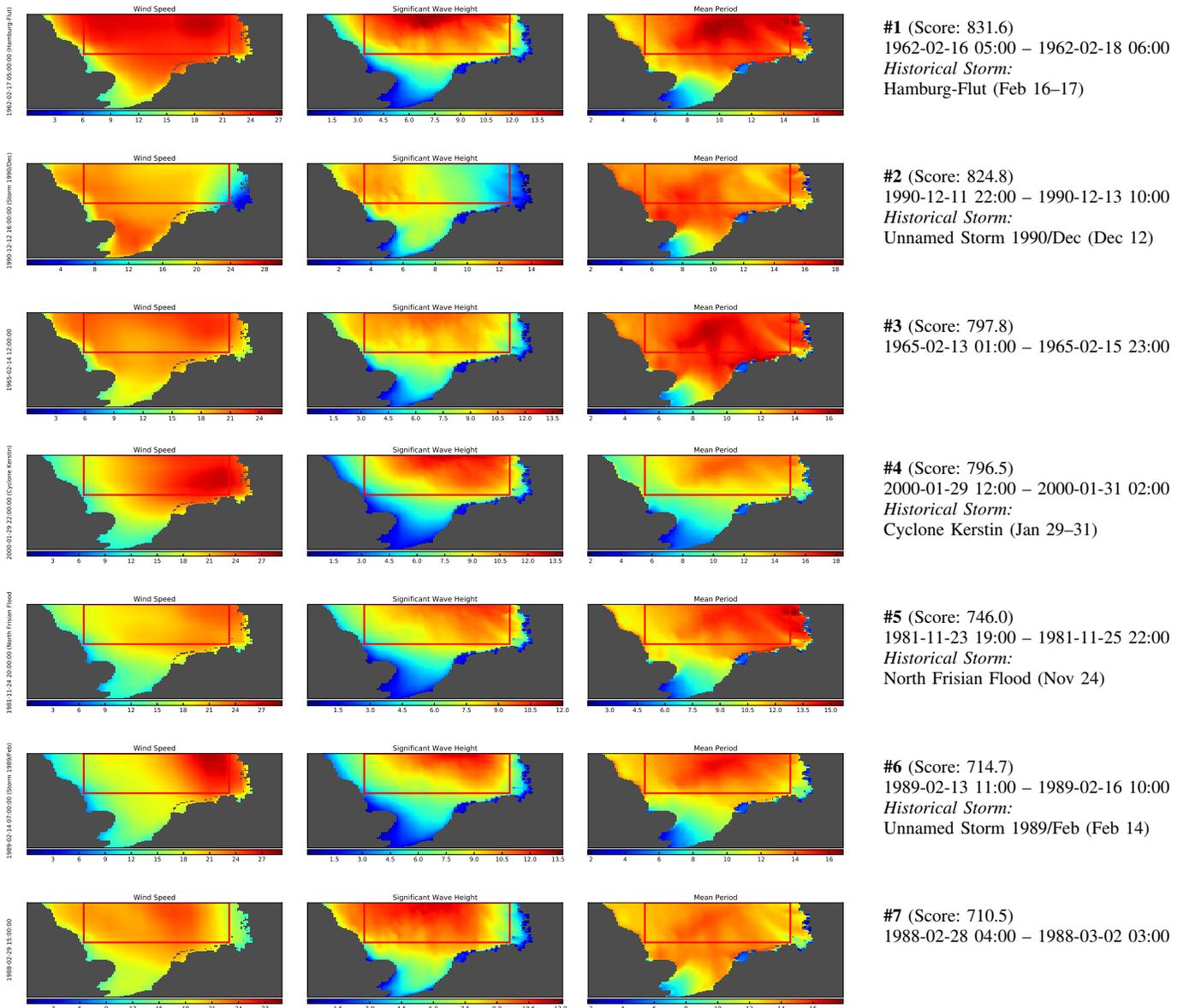


Fig. 7. Top 7 detected time frames on the coastDat-1 hindcast [21]. Heatmaps show the state of the three variables during the middle of the detected time frame. The red frame highlights the area whose data have been used for this experiment. Heatmaps are best viewed in color.

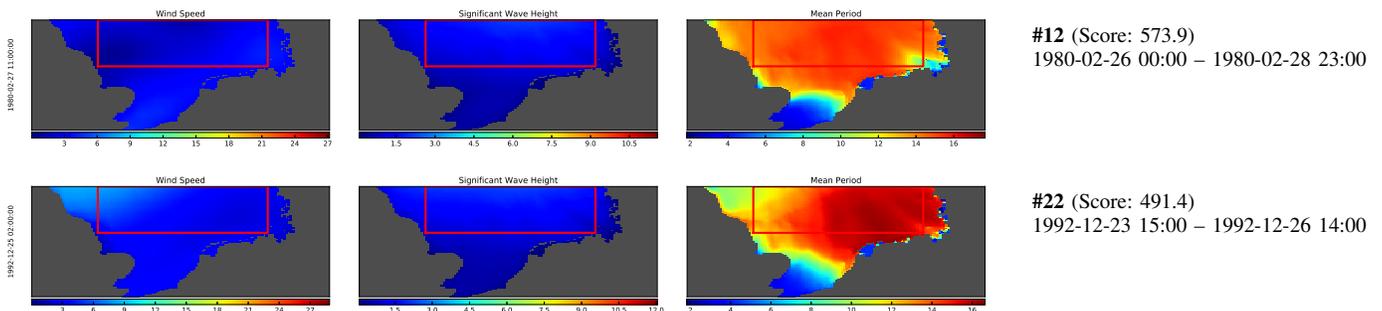


Fig. 8. Exemplary detections of time frames with unusually calm sea conditions on the coastDat-1 hindcast [21]. Only the middle of the time frame is shown.

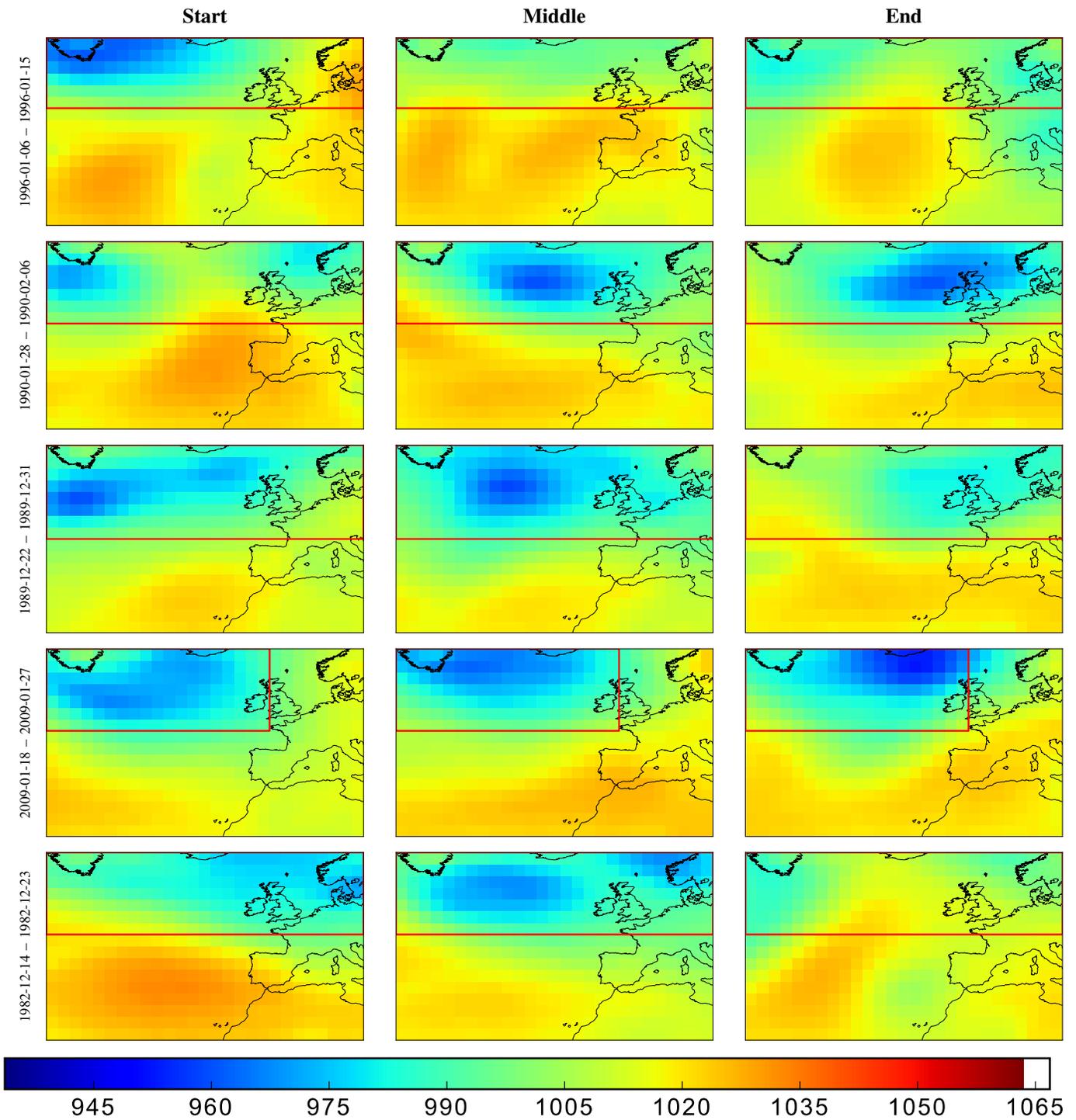


Fig. 9. Heatmaps showing sea level pressure at the beginning, the middle and the end of the top 5 detections on the SLP dataset. The red box marks the detected area. Heatmaps are best viewed in color.

- [5] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.
- [6] M. Das and S. Parthasarathy, "Anomaly detection and spatio-temporal analysis of global climate system," in *Proceedings of the third international workshop on knowledge discovery from sensor data*. ACM, 2009, pp. 142–150.
- [7] E. Keogh, J. Lin, and A. Fu, "Hot sax: Efficiently finding the most unusual time series subsequence," in *IEEE International Conference on Data Mining (ICDM)*. IEEE, 2005, pp. 8–pp.
- [8] M. Jiang, A. Beutel, P. Cui, B. Hooi, S. Yang, and C. Faloutsos, "A general suspiciousness metric for dense blocks in multimodal data," in *IEEE International Conference on Data Mining (ICDM)*. IEEE, 2015, pp. 781–786.
- [9] R. Mínguez, B. Reguero, A. Luceño, and F. Méndez, "Regression models for outlier identification (hurricanes and typhoons) in wave hindcast databases," *Journal of Atmospheric and Oceanic Technology*, vol. 29, no. 2, pp. 267–285, 2012.
- [10] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins *et al.*, "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," *arXiv preprint arXiv:1605.01156*, 2016.
- [11] E. Wu, W. Liu, and S. Chawla, "Spatio-temporal outlier detection in precipitation data," in *Knowledge discovery from sensor data*. Springer, 2010, pp. 115–133.
- [12] E. Rodner, B. Barz, Y. Guaniche, M. Flach, M. Mahecha, P. Bodesheim, M. Reichstein, and J. Denzler, "Maximally divergent intervals for anomaly detection," in *ICML Workshop on Anomaly Detection (ICML-WS)*, 2016.
- [13] J. Duchi, "Derivations for linear algebra and optimization," *Berkeley, California*, 2007.
- [14] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical review letters*, vol. 45, no. 9, p. 712, 1980.
- [15] A. Kut and D. Birant, "Spatio-temporal outlier detection in large databases," *CIT. Journal of computing and information technology*, vol. 14, no. 4, pp. 291–297, 2006.
- [16] T. Cheng and Z. Li, "A multiscale approach for spatio-temporal outlier detection," *T. GIS*, vol. 10, no. 2, pp. 253–263, 2006.
- [17] T. Kanungo and R. M. Haralick, "Multivariate hypothesis testing for gaussian data: Theory and software," Citeseer, Tech. Rep., 1995.
- [18] T. W. Anderson, "An introduction to multivariate statistical analysis," Wiley New York, Tech. Rep., 1962.
- [19] P. D. Allison, *Missing data*. Sage publications, 2001.
- [20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
- [21] Helmholtz-Zentrum Geesthacht, Zentrum für Material- und Küstenforschung GmbH. (2012) coastdat-1 waves north sea wave spectra hindcast (1948–2007). World Data Center for Climate (WDCC). [Online]. Available: https://doi.org/10.1594/WDCC/coastDat-1_Waves
- [22] P. Tan, M. Steinbach, V. Kumar, C. Potter, S. Klooster, and A. Torregrosa, "Finding spatio-temporal patterns in earth science data," in *KDD 2001 Workshop on Temporal Data Mining*, vol. 19, 2001.
- [23] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen *et al.*, "The ncep/ncar 40-year reanalysis project," *Bulletin of the American meteorological Society*, vol. 77, no. 3, pp. 437–471, 1996. [Online]. Available: <https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>