



Detecting changes in essential ecosystem and biodiversity properties - towards a Biosphere Atmosphere Change Index: BACI

Deliverable 5.2: Methods for Multivariate Novelty Detection for Synthetic Change Index



Project title:	Detecting changes in essential ecosystem and biodiversity properties- towards a Biosphere Atmosphere Change Index
Project Acronym	BACI
Grant Agreement Number:	640176
Main pillar:	Industrial Leadership
Topic:	EO-1-2014: New ideas for Earth-relevant space applications
Start date of the project:	1st April 2015
Duration of the project:	48 months
Dissemination level:	Public
Responsible of the deliverable:	Joachim Denzler Phone: +49 3641 9 46420 Email: joachim.denzler@uni-jena.de
Contributors:	Erik Rodner, Yanira Guanache, Björn Barz, Maxim Chernet-skiy, Milan Flach, Miguel Mahecha
Date of submission:	September 8, 2016

Contents

Summary	3
1 Multivariate Novelty Detection with Autoregressive Models	4
1.1 Methodology	4
1.1.1 Data Acquisition	5
1.1.2 Preprocessing: deseasonalization and standarization	5
1.1.3 Feature Extraction: ARMA models	5
1.1.4 Event Detection	8
1.2 Global Application	11
1.3 Future Work: Attribution scheme	13
1.4 Conclusions	14
2 Advances in the Maximally Divergent Intervals method for Anomaly Detection	16
2.1 Improvements and extensions to the MDI algorithm	16
2.1.1 Efficient distribution parameter estimation using cumulative sums	16
2.1.2 Constant-time evaluation of the KL divergence for Gaussian distributions	17
2.1.3 Interval Proposals	17
2.1.4 Time-Delay Embedding	18
2.1.5 An unbiased variant of the KL divergence	19
2.1.6 Deseasonalization	20
2.1.7 Implementation details	21
2.2 Experiments	22
2.2.1 Synthetic Data	22
2.2.2 Hurricane Detection	24
2.2.3 Detection of North Sea Storms	25
2.3 Preliminary tests with BACI data	26
2.3.1 Albedo and reflectance	26
2.4 Conclusions	29
3 Conclusions	31
References	32

Summary

This deliverable is dedicated to the works done within the *WP5 - Synthetic Index and Attribution Scheme: the BACIndex*. WP5 is divided into 4 main tasks and this second report refers to the second one: *Task 5.2 - Near-real-time processing of non-normalized time series and large-scale adaptations*.

Our main focus during this task was twofold: on one hand we have developed a methodology to identify abnormal events in multivariate biosphere time series and on the other hand we have been working on the extension and improvement of the methods developed in Task 5.1 in order to cope with large-scale data and speed up computation time.

These two main objectives correspond to the two sections of this deliverable:

1. **Novelty Detection with Multivariate Autoregressive Models.** In this section, we show a methodology to detect extreme events in biosphere data based on autoregressive models fitted locally. Two complementary approaches were tested, one based on coexceedances above a threshold and another one based on a distance metric of the joint distribution of the variables. The methods have been applied locally and globally.
2. **Advances in the Maximally Divergent Intervals method for Anomaly Detection.** In this second section, we present the improvements implemented for our method developed during Task 5.1 and detailed in Deliverable 5.1. These advances include a more efficient evaluation of the Kullback-Leibler divergence that allows us to speed up the computation time and to use significantly larger data bases than before. Several experiments have been carried out along these months with both artificial and real data. Additionally, we have started to test the algorithm with real data developed within the BACI project by other partners of the consortium.

1 Multivariate Novelty Detection with Autoregressive Models

During the first months of the project, we have been testing different methods and techniques to detect abnormal events and applying them mainly to artificial data specifically created for this task. Now, we propose a preliminary approach and we test it with a real data application. The development of this approach would help us to later define a BACIndex.

This intuitive approach combines multiple techniques and has been applied to assess the occurrences of multivariate tail events of terrestrial biosphere variables. A correct assessment of extreme events will allow us to detect sensitive regions as well as more severe phases in the historical records.

In addition, the way our method is implemented will allow the definition of an attribution scheme in the near future (*Task 5.4 - Attribution Scheme*). This would help us to understand the processes underlying behind these anomalies.

The methodology proposed as well as the results obtained have been accepted for presentation at the peer-reviewed conference: 6th International Workshop on Climate Informatics, to be hosted by the National Center for Atmospheric Research next September in Colorado, USA, [12].

1.1 Methodology

We have followed the same scheme proposed in Deliverable 5.1 to achieve a Change Index, dividing the process into: 1. Data Acquisition, 2. Preprocessing, 3. Feature Extraction, 4. Event Detection and 5. Change Index, see Figure 1.

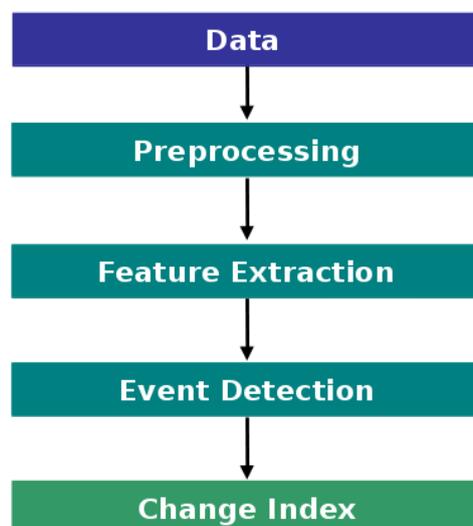


Figure 1: Steps to develop a Change Index, BACIndex.

In the following sections, the approach used at each step will be detailed. We explain and illustrate the method with the help of real data acquired at 50.5°N, 12.5°E, in middle-Europe. Afterwards, an extension to a global scale will be presented.

1.1.1 Data Acquisition

In the tests carried out during the last months, a preliminary version of the Earth System Data Cube developed within the CAB-LAB project¹ has been used. The Earth System Data Cube is a practical way of storing spatio-temporal data consisting of: time, latitude, longitude, and multivariate EOs. The initial global version covers from January 2001 to December 2012 with one observation every 8 days and a spatial grid with resolution of 1°. Up to 14 atmospheric and biospheric variables are included in this database. From these variables, we have used those that measure terrestrial biosphere parameters:

- Fraction of photosynthetic active radiation (*Fpar*)
- Leaf surface temperature (*LSTD*)
- Global primary productivity (*GPP*)
- Terrestrial ecosystem respiration (*TER*)
- Heat flux (*H*)
- Latent heat flux (*LE*)

1.1.2 Preprocessing: deseasonalization and standarization

To avoid later inconsistencies, the data need to be pretreated. We have applied two common techniques in environmental sciences; initially, we have removed the seasonal pattern usually present in environmental variables. To do so we have just subtracted the mean value of the time series corresponding to each day of the year. Then we have standarized the remaining variables by subtracting its mean, μ , and dividing by its variance, σ .

Figure 2 shows the deseasonalization and standarization of the variables at 50.5°N, 12.5°E. In the left plot, the 6 original variables are represented in blue, black lines show their seasonal cycle and the remaining data once the seasonal cycle has been subtracted is plotted in red. The seasonal cycle has been simply estimated as the mean value of the signal for each day of the year. The right plot shows the resulting time series after removing the seasonality and standarizing them ($\mu=0$, $\sigma=1$). The transformed time series presented in the right plot are the ones to be used in the following steps.

1.1.3 Feature Extraction: ARMA models

An abnormal event can be defined as a point within the time series not well represented by a previously fitted statistical model, [4]. Following this intuitive point of view, we have

¹see: earthsystemdatacube.net/

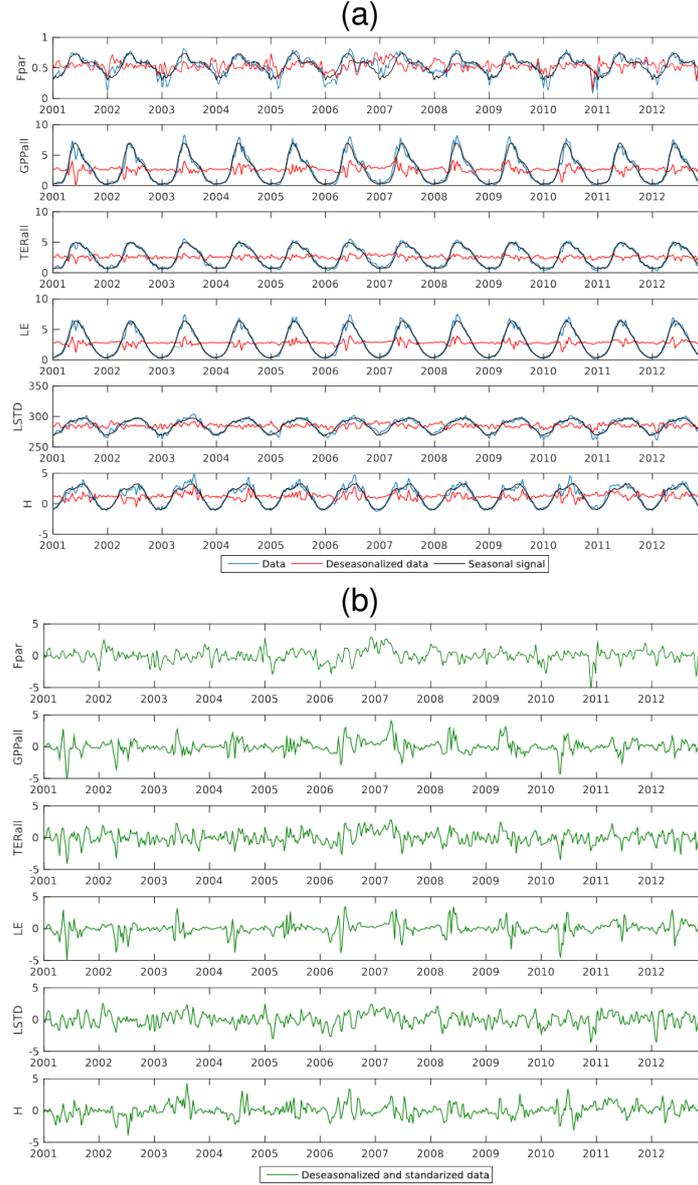


Figure 2: Deseasonalization (a) and standardization (b) of the 6 variables of interest.

applied an Autoregressive-moving-average model (ARMA model) to afterwards compute the residuals between the model and the data. Those points where the residuals are significantly high can be considered as extreme events that the model is not able to represent. An $ARMA(p, q)$ model consists of two parts, an autoregressive part (AR) and a moving average part (MA). The coefficients p and q refer to the order of each part:

$$X_t = \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (1)$$

where $\varphi_1, \dots, \varphi_p$ and $\theta_1, \dots, \theta_q$ are parameters of the model and $\varepsilon_t, \varepsilon_{t-1}, \dots$ are white noise error terms.

Table 1: Cros-correlations coefficients

	Fpar	GPPall	TERall	LE	LSTD	H
Fpar	1.00	0.451	0.575	0.403	0.716	0.203
GPPall	-	1.00	0.809	0.806	0.447	0.519
TERall	-	-	1.00	0.668	0.725	0.400
LE	-	-	-	1.00	0.485	0.596
LSTD	-	-	-	-	1.00	0.430
H	-	-	-	-	-	1.00

Although there are more complex approaches available, e.g. [3, 23], we are working with univariate ARMA models independently fitted to each variable. The work of [10] compared the performance of univariate and multivariate models and as a result of their work they recommended the use of univariate models, specially in those cases where the cross-correlations between variables are not particularly strong. Multivariate models involve a greater number of parameters, which might be a disadvantage, while their performance is comparable to univariate approaches. Univariate simpler structure and easier interpretation are additional valuable features. The cross-correlations within our pretreated variables do not present very strong correlations (only two values above 0.8), see Table 1. This encourages us to follow an univariate approach. Nonetheless, we plan to test other more complicated methods like multivariate ARMA models, Generalized Autoregressive Conditional Heteroscedasticity - GARCH models or Support Vector Data Description (SVDD) techniques [25].

Following the procedure detailed by [2], we have plotted the autocorrelation and partial autocorrelation functions of each pretreated variable, see Figure 3. These two statistical measures indicate how a time series is related to itself over time. For example, the autocorrelation function at lag k is the correlation between the original series and the same series moved forward k timesteps (X_t and X_{t-k}). On the other hand, the partial autocorrelation at lag k is the autocorrelation between X_t and X_{t-k} that is not accounted for lags between 1 and $k - 1$. Applied to our time series, both plots (Figure 3) present an exponentially decreasing height of the red bars depicted: this indicates that we need to consider a model with both parts (AR and MA) represented. Taking a look at Figure 3, counting the number of bars exceeding the significance bands in the partial autocorrelation gives us an idea of the autoregressive order to be considered. In this case, it seems that 3 autoregressive terms represent sufficiently most of the variables.

Considering this initial visual inspection we have fixed $p = 3$ (3 first autoregressive terms) and we tested different values of q . To compare different fitted models, we have used the Ljung-Box test applied to the residuals of the model [16]. The residuals of a well fitted ARMA model should look like white noise [2]. By means of this test we are able to analyze whether the residuals fulfill this assumption or not. In Table 2 the p -values of the Ljung-box test for models from (3,0) to (3,3) and all the variables are summarized. Models with p, q values of (3,1), (3,2) and (3,3) guarantee residuals without autocorrelation for all the variables while with the (3,0) model this is not guaranteed for Fpar and LSTD. Additionally the p -values for GPPall, LE and H are pretty

low. Therefore we discard model (3,0). From the three remaining models, although the (3,2) presented slightly higher p -values of the Ljung-Box test we selected the simplest one: (3,1).

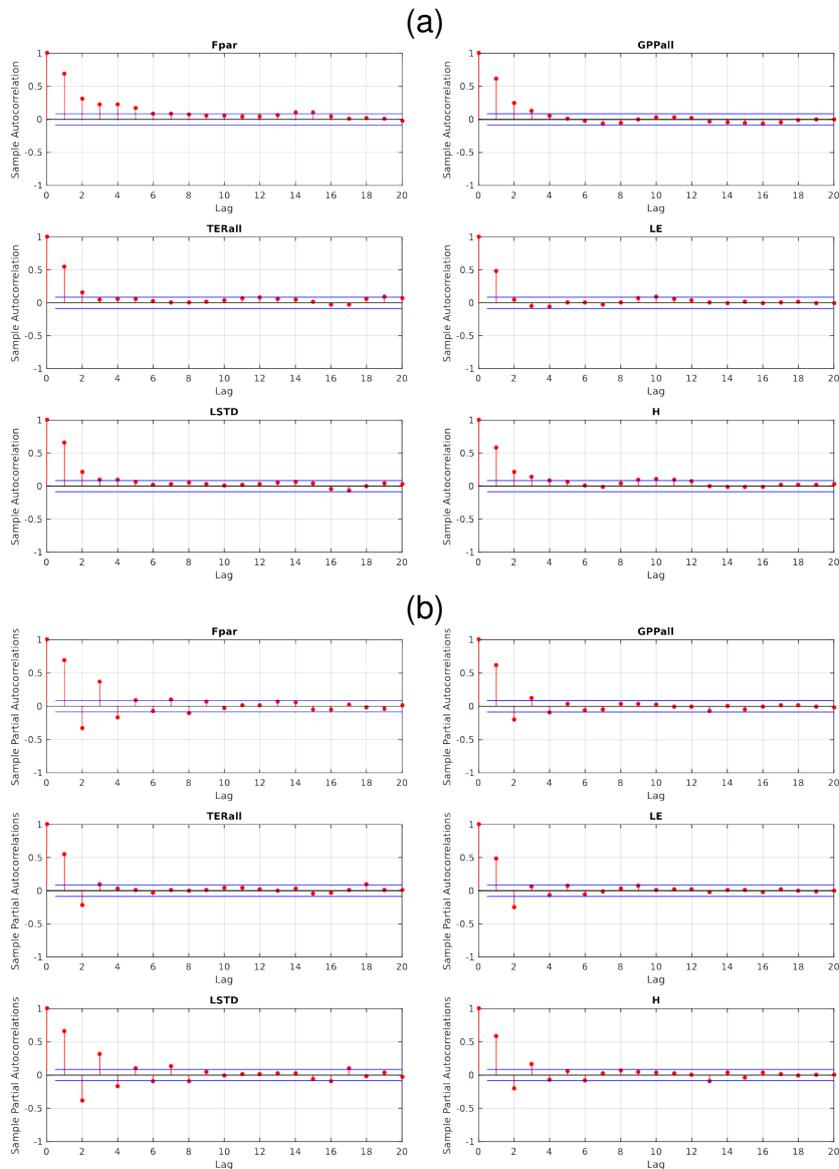


Figure 3: Autocorrelation plots (a) and Partial autocorrelation plots (b) of the 6 variables pretreated.

1.1.4 Event Detection

Once we have selected the ARMA model to be applied, we extract the residuals between the model and the data. By means of these residuals, we identify the extreme events within the multivariate time series. To do so, we have applied two different methods: a) estimating the coexceedances over a certain threshold and b) estimating the Mahalanobis distance of the residuals to their mean.

	(p, q)			
	(3,0)	(3,1)	(3,2)	(3,3)
Fpar	—	0.665	0.990	0.989
GPPall	0.309	0.999	0.999	0.922
TERall	0.968	0.998	0.995	0.908
LE	0.242	0.982	0.991	0.999
LSTD	—	0.849	0.987	0.572
H	0.212	0.989	0.953	0.945

Table 2: p -values of the Ljung-Box test for autocorrelation of the model residuals

Extreme residuals coexceedances We have defined an extreme residual as one for which its absolute value lies above the 85% percentile of the residuals distribution. This is done for each of our 6 variables of interest. From our 12-years database, the 15% highest residuals count 83 exceedances for each variable. Next, we count the number of variables simultaneously presenting an extreme residual for each timestep. Figure 4 shows the time series of coexceedances, also known as coincidences [1, 19, 22, 9, 29]. This is a very intuitive method and easy to understand, which, in addition, allows a direct interpretation of the extremes detected.

Usually, extreme events are defined over higher/lower thresholds, for example 90% for top-tail events or 10% for bottom-tail events. But we have considered at the same time positive and negative extremes, therefore, we are not dividing between top-tail coexceedances and bottom-tail coexceedances. A 10% cut-off to the absolute value would give us few observations for a meaningful analysis, which is the reason why we adopted a threshold at the 85% percentile.

Mahalanobis distance The main disadvantage of the coexceedances method is that it does not take into account the shape of the joint distribution of the residuals. To try to solve this problem, the Mahalanobis distance would be a good alternative [17]. This metric, also known as Hotellings T^2 [13], compared to the Euclidean distance does not only takes the mean but also the covariance matrix of the distribution into account (Figure 5).

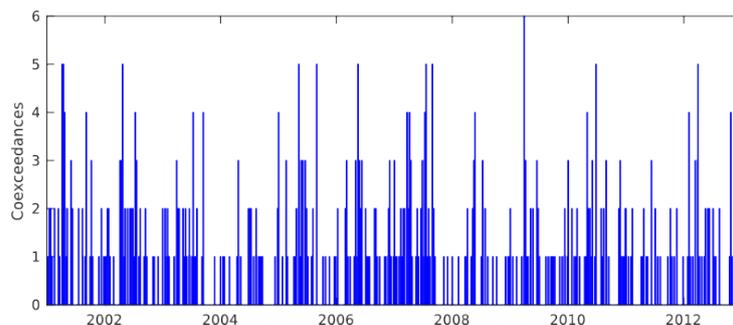


Figure 4: Coexceedances.

Although the residuals joint distribution does not follow a multivariate Gaussian distribution, it does not present a clear multimodality, see Figure 6. Therefore, Mahalanobis is still a robust approach as argued by [27]. Another more complex option would be the use of Support Vector Data Description (SVDD) techniques [25].

A direct comparison between both methods would not be fair as both methods present different characteristics: the coexceedances method requires an initial threshold for each variable while the Mahalanobis distance is computed to all the residuals without considering a threshold. In addition, both methods are measuring different aspects: a peak or strong coexceedance event requires that almost all the variables at the same time exceed the threshold imposed; but on the other hand, a peak in the Mahalanobis distance time series is only pointing to an observation that differs significantly from the rest of joint distribution without informing if this large distance is caused in 1,

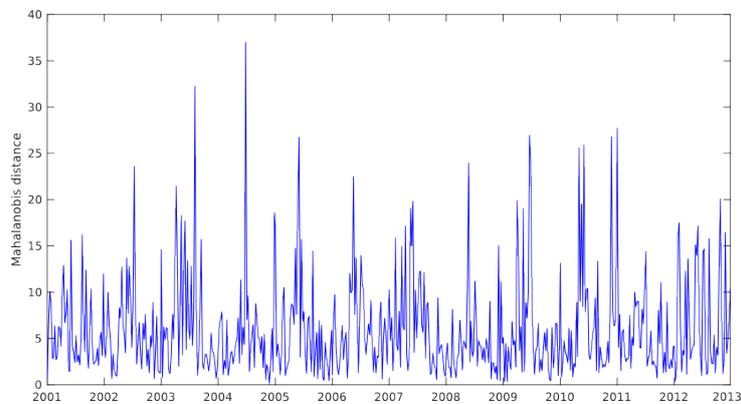


Figure 5: Mahalanobis distance.

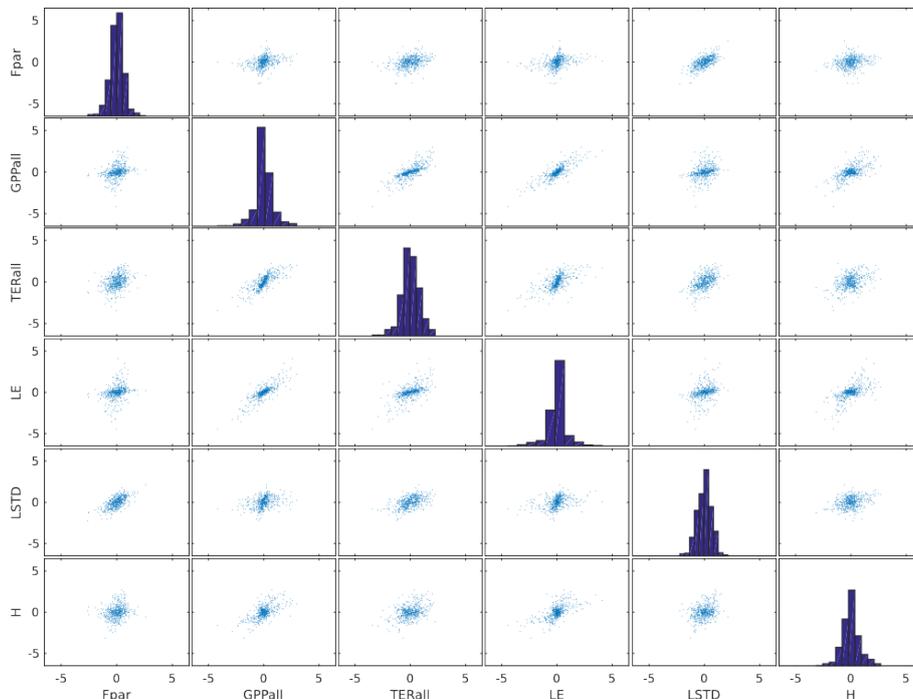


Figure 6: Histograms and scatter-plot matrix of the ARMA (3,1) residuals.

2 or the 6 dimensions at the same time. In other words, coexceedances only inform about those time steps where more than one variable exceed the threshold but without informing about the intensity of this exceedance and Mahalanobis distance only informs about far points in the joint distribution but without informing about in which or how many dimensions.

For benchmarking, we have focused on one important historical event, the heat-wave of 2003. In the summer of 2003, Europe experienced the warmest summer record so far with many casualties, economic and ecologic damages [7, 5, 20, 21, 24]. From a simple visual inspection, it can be seen in Figures 4 and 5 that the Mahalanobis distance seems to detect the event during summer 2003 with a period of high distance values. On the other hand, within this period the number of variables coexceeding does not seem to be more than 3 out of 6 variables.

1.2 Global Application

Following the same procedure explained before, we have applied it to all the locations encompassed in the Earth System Data Cube. At each punctual location, we have de-seasonalized and normalized the 6 variables ($\mu=0$, $\sigma=1$) and then fitted an ARMA(3,1) model. We have assumed the same kind of ARMA model for all the locations for comparison and simplicity reasons. However, we are aware this might be a strong assumption and we keep on investigating alternative approaches.

Comparing the results provided by the model with the 6 variables itself, we obtain the residuals at each location. With these residuals we have: a) counted the coexceedances over a local threshold at the 85% percentile and b) estimated the Mahalanobis distances for all the timesteps considering the local covariance matrix and a critical distance for which an extreme is significant. This critical distance has been estimated as the 99% quantile of the χ^2 with 6 degrees of freedom [6]. Figure 7 presents the percentage of observations with *strong* coexceedances (upper plot) and Mahalanobis distance scores above the critical value (lower plot). Strong coexceedances are those timesteps where at least 4 variables present values above its 85% percentile simultaneously.

This general overview allows us to detect regions where to focus. However, it also presents some open issues: in the northern latitudes (*e.g.* Russia, Canada, and Alaska), the amount of extreme residuals detected by both methods is higher to other areas in the globe. This can be related to the fact that in northern latitudes, biosphere variables present big changes in its variance along the year. This heteroscedastic behaviour overestimates the number of extremes. Additional error sources might be the globally fixed (p, q) parameters of the ARMA model, trends in the data or issues related to the projection error from the satellites. These issues need further investigations and we are currently working on them.

Coming back to the historical event of 2003 in Europe, Figure 8 represents the results obtained through both approaches over Europe for a certain timestep: 5th of August 2003. At this time, Europe was under the extreme conditions of the hottest

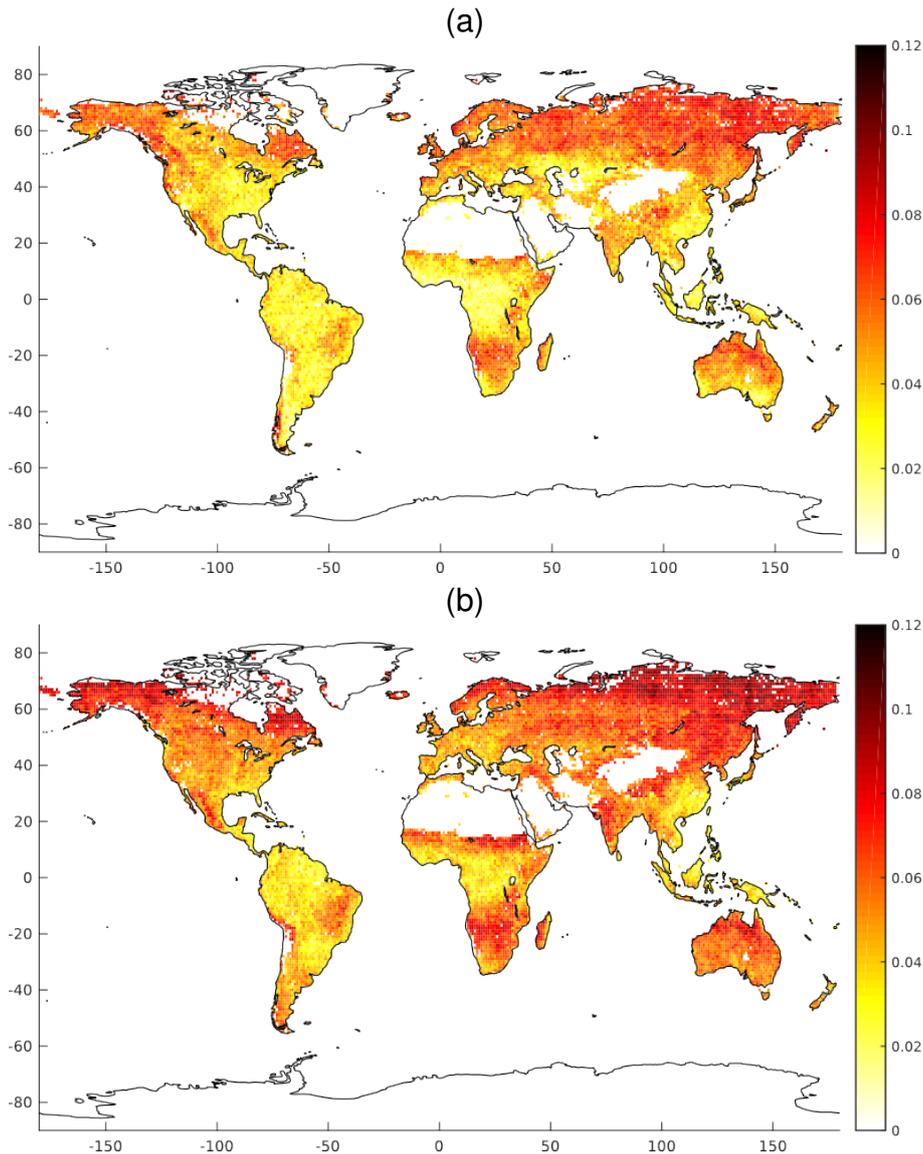


Figure 7: (a) Percentage of strong coexceedance events (4-6 residuals extremes). (b) Percentage of Mahalanobis distance above a critical value.

summer record since the XVI century [24]. The Mahalanobis distance represented in Figure 8 clearly shows the pattern of an extreme event in central Europe. On the other hand, the coexceedances are detecting 2-3 of 6 variables being extreme. However, with the coexceedances method it can be easily seen that LSTD and H are extreme along mid-Europe while the 2003 heat wave (Figure 9); showing that variables closer to the atmosphere (LSTD and H) present distinctly different behavior than the biosphere one (GPP, TER, LE).

Moving to another historical event; during March 2012 one of the greatest heat waves in North America was recorded [8]. Over the Great Lakes region all temperature records were broken with extremely warm temperatures for that area and month of the year. Figure 10 shows the results of our approach by the middle of March in North America. Both methods clearly detect this event with strong coexceedances and high Mahalanobis distance scores.

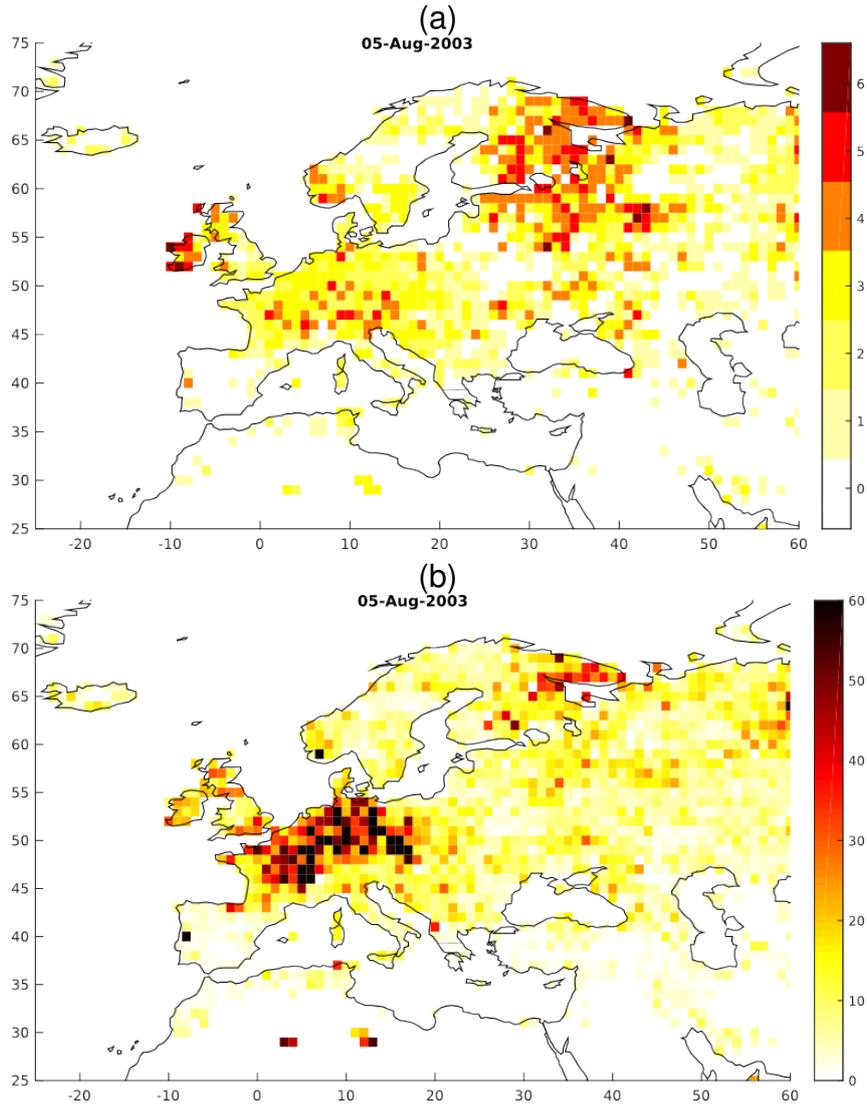


Figure 8: (a) Coexceedances on 5th of August 2003. (b) Mahalanobis distance on 5th of August 2003.

1.3 Future Work: Attribution scheme

The presented methodology has been implemented in such a manner that it is easy to establish an attribution scheme later on. Although there are some open issues, we already have a pretty well defined idea on how we want to proceed.

Once we have defined the extreme events, either by means of coexceedances, Mahalanobis distance, or even a combination of both, we plan to implement a multinomial logistic regression model to estimate their occurrence. This kind of models allow us to introduce explanatory variables as covariates. We plan to use atmospheric and temporal variables as explanatory variables to model multivariate biosphere extremes. By means of statistical tests such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) we could decide which covariates are worth to be added in

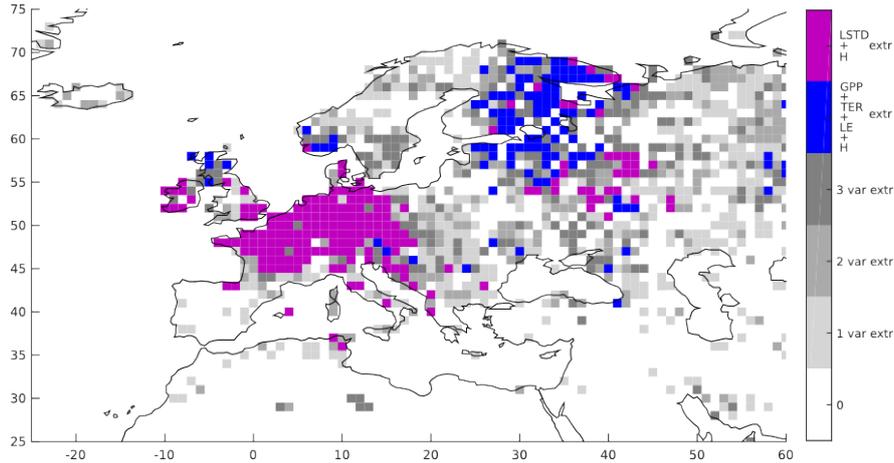


Figure 9: Variables coexceeding on 5th of August 2003. Grey scale-up to 3 unclassified variables being extreme, purple pixels-extreme LSTD and H and blue pixels- extreme GPP, LE, TER and H.

the model. Additionally we will also test whether the inclusion of widely known climate indices such as ENSO, NAO, SOI, etc is significant.

1.4 Conclusions

We have presented a method to detect abnormal events in multivariate time series. The basic idea behind our approach is that an anomaly is a point in the time series which cannot be well represented by a statistical model that describes the entire time series. This assumption allowed us to investigate the occurrence of extreme observations in biospheric variables with two approaches: coexceedances and Mahalanobis distance of extremal residuals. Both methods present advantages and disadvantages that make them useful and complementary. The coexceedances approach is a method very easy to interpret and allows a direct explanation of the detected extremes. The Mahalanobis distance is able to provide more information about the multivariate joint distribution by losing the direct information about which are the variables being extreme.

Both methods tested are able to detect historical events like the European Heat-wave from 2003 or the so-called *Meteorological March Madness* from 2012 in North America.

Multivariate methods allow detections not appreciable from univariate methods; like the different behavior of variables closer to atmosphere compared to those closer to biosphere shown in the 2003 heat wave.

During the development of the proposed methodology and the consequent results, we faced several open questions that needed further investigation. We keep on working on these aspects to improve our methods and make them more robust and general. Therefore, our near future work in this direction will be focused on the following points:

- *i*) Resolving the overestimation of extremes in the northern latitudes;

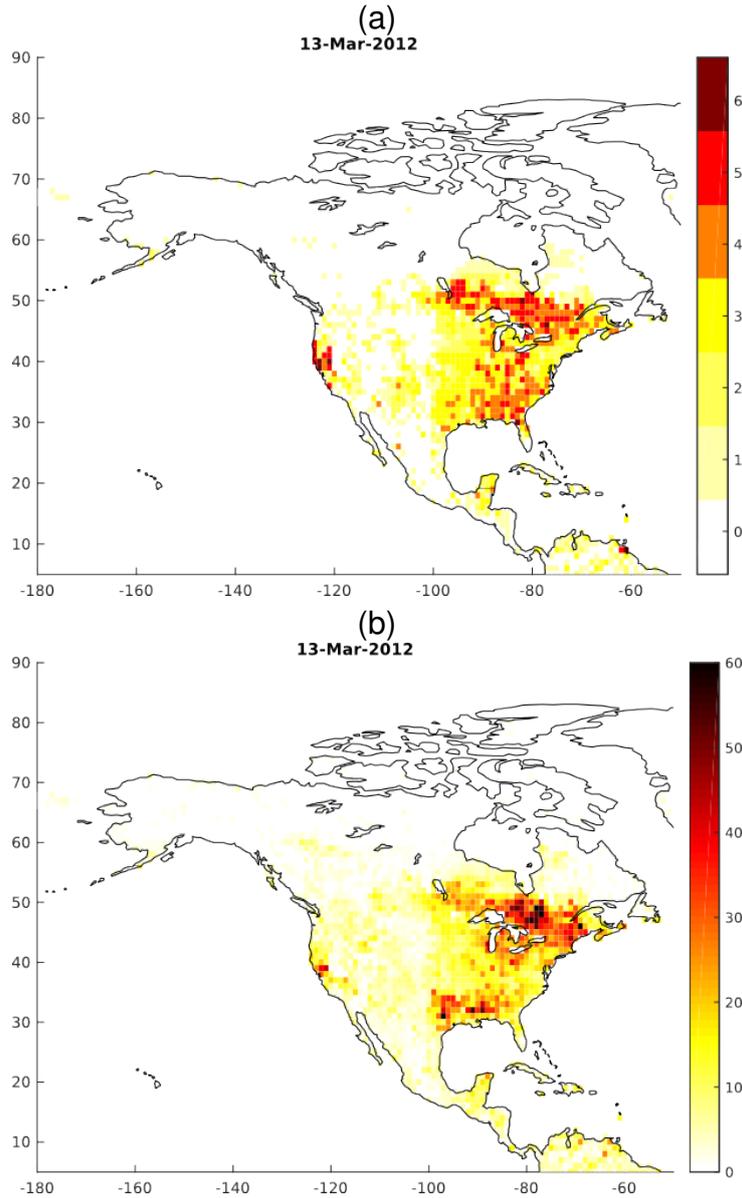


Figure 10: (a) Coexceedances on 13th of March 2012. (b) Mahalanobis distance on 13th of March 2012.

- *ii) A better threshold selection needs to be done to make both methods comparable (Task 5.3 - Incremental novelty detection, automatic dataset cleanup and going from novelty scores to direct detections);*
- *iii) Regionalization into areas with similar behavior. Consequently, adapting specific ARMA models for each similar region (Task 5.3 - Incremental novelty detection, automatic dataset cleanup and going from novelty scores to direct detections);*
- *iv) Attribution scheme. This step is crucial to understand the processes causing abnormal events (Task 5.4 - Attribution Scheme).*

2 Advances in the Maximally Divergent Intervals method for Anomaly Detection

The early prototype of our novel *Maximally Divergent Intervals Algorithm for Anomaly Detection* described in Deliverable 5.1 yielded promising results, but still suffered from some major shortcomings:

- Estimating the distribution of the data within each possible interval and computing the Kullback-Leibler (KL) divergence was computationally demanding and prevented the method from being used on large-scale data sets.
- The KL scores were systematically higher for smaller intervals, particularly in combination with the assumption of Gaussian distributions, leading to a bias towards smaller detections.
- Both distribution models proposed were based on the assumption of independent samples, which does almost never hold when working with real data.

We have tackled all those problems and implemented some improvements to the algorithm which will be described in the following sections. We have also tested the performance of the method on both synthetic data and real climate data, obtaining very promising results.

2.1 Improvements and extensions to the MDI algorithm

2.1.1 Efficient distribution parameter estimation using cumulative sums

In Deliverable 5.1, we have proposed two different approaches for modelling the distribution of the data: Kernel Density Estimation (KDE) and multivariate Gaussian distributions.

For estimating the value of the probability density function $p_I(x_t)$ at time t in the time series $(x_t)_{t=1}^n, x_t \in \mathbb{R}^D$ with Kernel Density Estimation a sum over all samples in the time series has to be evaluated:

$$p_I(x_t) = \frac{1}{n} \sum_{t'=1}^n K(x_t, x_{t'}), \quad (2)$$

where $K(x, x')$ is a kernel function. Evaluating this sum for every time step in every interval being scored is both time-consuming and redundant. We were able to speed up this computation significantly by using cumulative sums [26]:

$$C_{t,t'} = \sum_{1 \leq t'' \leq t'} K(x_t, x_{t''}). \quad (3)$$

This allows us to evaluate the empirical KL divergence for a given interval $I = \{t | a \leq t < b\}$ in linear time:

$$\begin{aligned}
KL_{I,\Omega} &= \frac{1}{|I|} \sum_{t \in I} \log \left(\frac{p_I(x_t)}{p_\Omega(x_t)} \right) \\
&= \frac{1}{|I|} \sum_{t \in I} \log (C_{t,b-1} - C_{t,a-1}) - \log (C_{t,n} - C_{t,b-1} + C_{t,a-1}).
\end{aligned} \tag{4}$$

Similarly, cumulative sums can also be used to speed up the estimation of the mean vector and covariance matrix of a Gaussian distribution.

2.1.2 Constant-time evaluation of the KL divergence for Gaussian distributions

Using the assumption of normally distributed data, additional speed can be gained by using a closed-form solution for the KL divergence given by [11]:

$$\begin{aligned}
KL(p_I(x_t), p_\Omega(x_t)) &= \frac{1}{2} (\text{trace}(S_\Omega^{-1} S_I) + (\mu_I - \mu_\Omega)^T S_\Omega^{-1} (\mu_I - \mu_\Omega) \\
&\quad + \log |S_\Omega| - \log |S_I| - D).
\end{aligned} \tag{5}$$

In combination with the cumulative sums approach described above for estimating the parameters of the distributions this allows us to compute the anomaly score of a given interval in constant time.

2.1.3 Interval Proposals

Despite the improvements described above, evaluating the KL divergence for all intervals with size between a given minimum and maximum may still take a long time if the time series contains many measurements and the range of possible interval sizes is large. To achieve a further significant speed-up, we propose an interval proposal method which proposes interesting intervals to the algorithm for in-depth analysis. This way no time is wasted for checking a large number of uninteresting intervals.

The proposals are based on a fast point-wise anomaly detection method, for example Hotelling's T^2 [13] or point-wise Kernel Density Estimation. Both methods return an anomaly score for every time step. We then approximate the gradient of those scores using a simple convolution to eliminate regions of a constant degree of anomalousness. The peaks of these point-wise gradient scores are extracted by a simple threshold operation and all intervals whose start and end point is one of those peaks and whose size is between the given limits are proposed to the algorithm for scoring by the KL divergence.

In our experiments, which will be described later, this led to a speed-up of several orders of magnitude while not jeopardizing the accuracy, since the proposal method acts as a high-recall retrieval system.

2.1.4 Time-Delay Embedding

In order to mitigate the assumption of independent time-steps we apply *Time-Delay Embedding* [18, 15] as a preprocessing step to integrate the attributes of some previous time steps into each single time step as context:

$$x'_t = (x_t, x_{t-T}, x_{t-2T}, \dots, x_{t-(k-1) \cdot T}), \quad (6)$$

where k is called the *embedding dimension*, *i.e.* the number of previous time-steps considered, and T is called the *time lag*, *i.e.* the distance between each two consecutive time steps. This idea is illustrated in Figure 11.

The optimal value for the parameters k, T highly depends on the data itself. Both parameters are not independent, since their product can be seen as the size of the window of relevant context. But a high value of k results in a high-dimensional time series and, thus, increases computational complexity. On the other hand, a high value of T is more likely to skip some important context information. After obtaining a good estimate for the size of that window, one can choose k and T to make a trade-off between computational complexity and accuracy.

We have investigated various methods to find a good estimate for the context window size based on experiments with our synthetic data set described in section 2.2.1:

Rank Aggregation We fix the value of T and apply the MDI algorithm for several different values of k . We then aggregate the resulting rankings of intervals into a single ranking using *Approximate Kemeny Aggregation* which prefers one interval over another one if the majority of rankers (here: embedding dimensions) does so.

Length Scale We fit a Gaussian Process with a squared-exponential correlation function to the data and determine the size of the context window based on the *characteristic length scale* of that process (*i.e.* the standard deviation of the correlation function). The characteristic length scale can be seen as a measure for how distant two samples have to be to not influence each other significantly.

Mutual Information According to Tobler's First Law of Geography, "everything is related to everything else, but near things are more related than distant things." Thus, mutual information between a sample and another sample T steps away usually decreases continuously, except for seasonal effects. Hence, instead of searching for local minima of mutual information, we determine an appropriate size of the context window based on the speed of the decrease (*i.e.* the gradient)

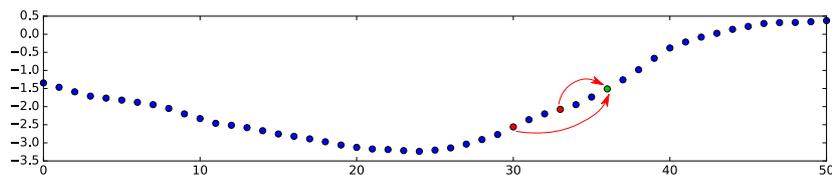


Figure 11: Illustration of Time-Delay Embedding with $k = T = 3$

of mutual information with increasing distance. We select the first distance where the gradient is below a pre-defined threshold as the context window size.

In our experiments, all three methods performed better than setting the parameters to a fixed value for all time series. Rank Aggregation performed best, but is infeasible in practice since it involves multiple expensive runs of the entire algorithm and is not compatible to our interval proposal method described above. Determining the context window size based on the decrease of Mutual Information gave comparable results and is faster than the Length Scale method, which performed worse in addition. But unfortunately, the threshold involved with the Mutual Information method again seems to be application-dependent.

Thus, optimization of the time-delay embedding parameters will be an area of ongoing research.

2.1.5 An unbiased variant of the KL divergence

We have theoretically analyzed the distribution of the KL divergence under the assumption of Gaussian distributions applied to time series of white noise and found a systematic bias towards intervals of smaller size, as can be seen in Figure 12.

This bias can be corrected by a simple multiplication of the divergence with the length of the interval in order to reward longer detections. This approach also is theoretically justified, since it can be shown that the resulting distribution of KL divergence scores on white noise is independent from the length of the intervals and follows a χ^2 -distribution which depends on the number of attributes only [14]. This insight can also be used to normalize the obtained scores regarding the number of attributes in order to make them comparable to scores obtained on different data.

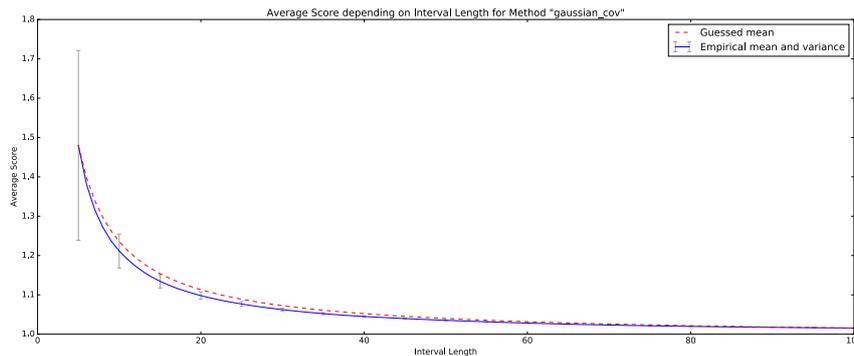


Figure 12: Empirical and theoretical mean KL divergence on white noise under the assumption of Gaussian distributions depending on the interval length. The theoretical mean is $\frac{1}{m} - \frac{1}{n-m}$, where m is the length of the interval and n is the length of the entire time series.

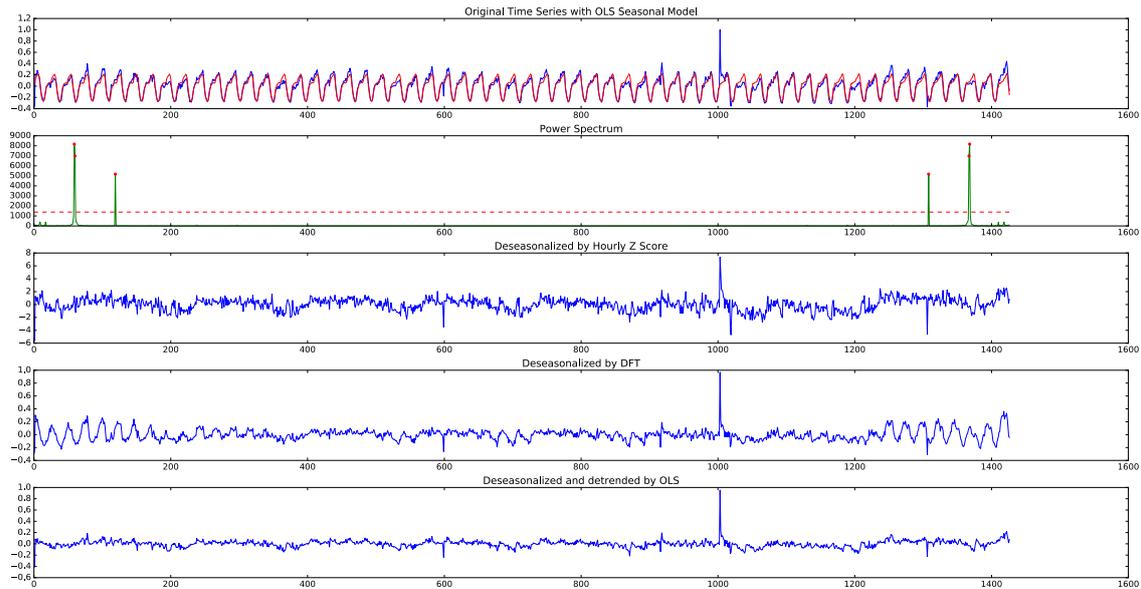


Figure 13: First plot: An exemplary time series with a diurnal seasonal pattern. Second plot: The power spectrum of that time series with an automatically determined threshold and all frequencies above that threshold marked by dots. Plots 3-5: Comparison of different deseasonalization methods applied to that time series.

2.1.6 Deseasonalization

Periodical patterns are very common in real data, particularly in climate data, which exhibits seasonal influences. We have implemented and tested several methods to remove this seasonality from such a time series:

Z-Score Partition the time series into seasonal groups $(x_{t \cdot S + p})_{t=1}^{n/S}$, $p = 0, \dots, S - 1$, where S is the number of seasonal groups, and normalize each group separately. This method is very common in literature, but is also known to over-emphasize small deviations in groups of rather low overall variance. In addition, prior knowledge about the number seasons is required.

OLS Deseasonalization Model the time series according to $x_t = a_0 + b_0 \cdot t + a_j + b_j \cdot t + \varepsilon_t$, where a_0 is an overall offset, $b_0 \cdot t$ is a global linear trend, a_j is the offset specific to the season j which time step t belongs to, $b_j \cdot t$ is a linear trend of that seasonal offset and the residual ε_t is the deseasonalized time series. The coefficients can be estimated by ordinary least squares (OLS) regression. This method does also require prior knowledge about the number and length of each season, but is more flexible regarding the association of the time steps with the seasons. This does also allow modelling multiple overlapping seasonal effects.

Fourier Transform Remove strong frequencies from the Fourier spectrum of the time series. This method can handle multiple overlapping seasonal effects and does not require any prior knowledge about the seasonality. In turn, it may result in artefacts, especially at the borders, due to the hard removal of frequencies and the periodicity assumption of the Fourier transform.

Figure 13 shows an example of the different deseasonalization methods applied to a time series with a diurnal seasonal pattern.

During our experiments on real data we have found that none of the methods generally performs better than the other ones. Which one is suitable to assist anomaly detection again depends on the data and the application. Moreover, for some applications removing the seasonality may even be disadvantageous (see section 2.2.3 for an example).

2.1.7 Implementation details

The entire algorithm, including KDE and Gaussian distribution models, interval proposals, time-delay and deseasonalization, has been implemented in a very efficient fashion in a C++ library called `libmaxdiv`. Bindings for Python are available to make the functionality of the library easily accessible.

The algorithm has been parallelised so that multiple intervals can be analysed concurrently on machines with multiple processors. But even without parallelisation the C++ implementation is about 50 times faster than the pure Python implementation of the algorithm which was available as of Deliverable 5.1.

When working with high-dimensional spatio-temporal data, time is not the only constraint, but memory limitations will become a problem too. Two particularly critical issues regarding memory consumption have been identified and solved using techniques described in the following:

Partial Cumulative Sums Time-Delay usually lead to very high-dimensional data. For the efficient computation of the covariance matrices needed by the Gaussian model, the number of values which have to be stored for each sample in the tensor of cumulative sums is quadratic in the number of dimensions. But since our algorithm scans through the possible intervals linearly, we do not need access to that entire tensor at once. Instead, we limit its size to a few gigabytes and compute a partial cumulative sum over a subset of time steps until that size is reached. Whenever a sum over samples not covered by that partial cumulative sum is requested, it will be dismissed and a new partial cumulative sum starting at the beginning of the current range will be computed. This way, memory consumption is always under control, while speed is not significantly impaired. Moreover, this approach mitigates numerical issues like round-off errors and cancellation which would arise from the use of large cumulative sums.

Concurrent Non-Maximum Suppression Normally, during the scan over all possible intervals our algorithm stores the scores of the intervals in a list. Afterwards, this list is sorted and non-maximum suppression is applied to get non-overlapping intervals only. But again, the number of possible intervals grows polynomially when working with spatio-temporal data. For example, during our experiment on the CoastDat dataset described in section 2.2.3, which was not even spatial, 90% of the memory (around 0.5 GB) have been occupied by the list of interval scores. Thus, non-maximum suppression in that form is not feasible for large datasets.

To solve this issue we have implemented an approximation which we call *concurrent non-maximum suppression*. Instead of collecting all scores before starting to eliminate local non-maxima, this algorithm maintains a sorted list of detections and applies non-maximum suppression for each detection at the moment it is to be inserted into the list: if there is already an overlapping detection with a higher score, the new detection will be dismissed; if there is not, it is inserted at the appropriate position in the list and all overlapping detections with a lower score will be removed from the list. As opposed to original non-maximum suppression, the output of this algorithm depends on the order of insertion. In turn, it consumes much less memory and does even run an order of magnitude faster than original non-maximum suppression, which took several minutes for the CoastDat experiment.

2.2 Experiments

2.2.1 Synthetic Data

For a quantitative evaluation of our algorithm, we have created a data set of synthetic time series sampled from a Gaussian Process. Different types of anomalies have been injected into each of these time series: a simple shift of the mean value, a change of amplitude, a change of frequency and a replacement of a random part of the time series with a part of another one, interpolating them at the borders to make a smooth transition. For each anomaly type there is a univariate as well as a multivariate test case. In case of the mean-shift anomaly we have also set up a test case where 5 anomalies are injected into a single time series. There are 11 test cases in total, each with 100 different time series of length 1000. The size of the anomalies ranges between 2% and 10% of the length of the time series.

Note that this test set would have been intractable for our prototype implementation of Deliverable 5.1, but can be processed easily by the new efficient implementation.

We compare the performance of our algorithm to three baselines:

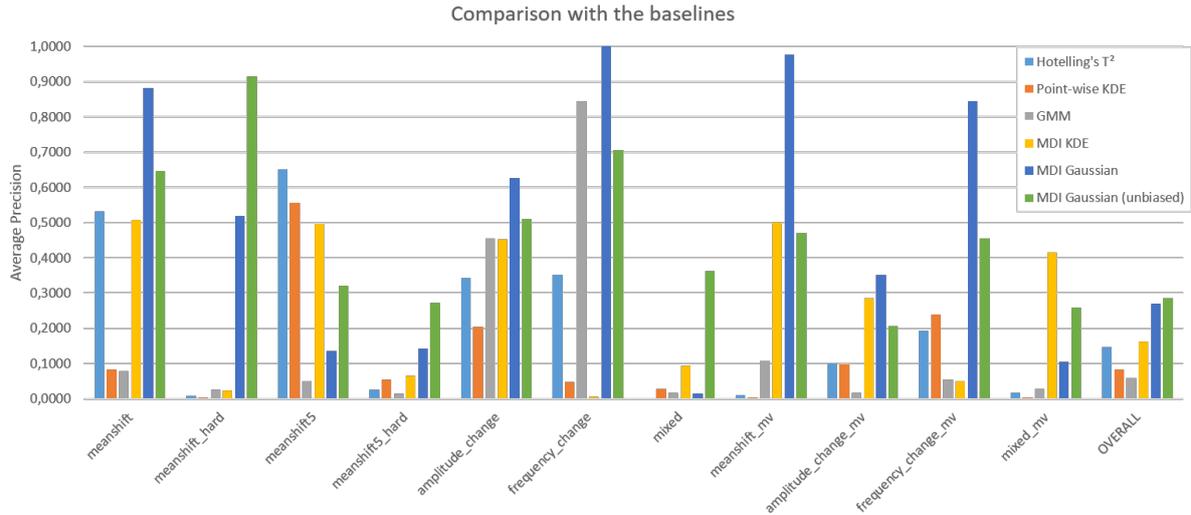
Hotelling's T^2 The Mahalanobis distance of each sample to the mean value.

Point-Wise Kernel Density Estimation Use KDE to estimate the probability density function $p(x_t)$ and score each sample x_t by its unlikelihood $1 - p(x_t)$.

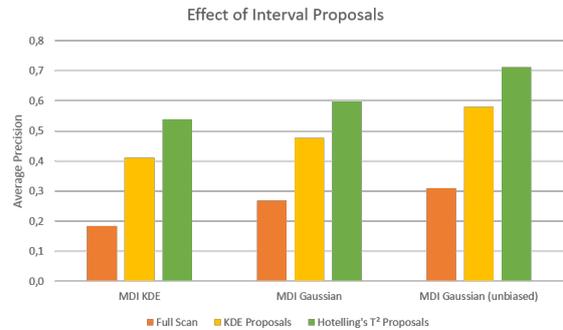
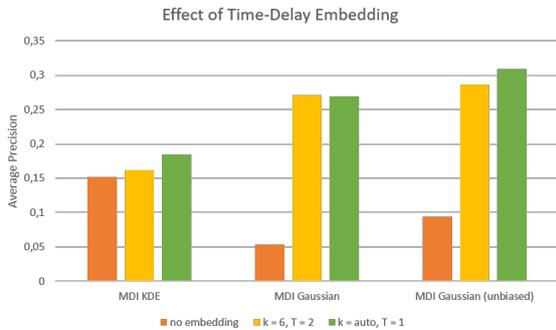
Gaussian Mixture Models Fit a GMM with 2 components to the data, denote the component with the higher weight as the *nominal component* and the other one as the *anomalous component* and score each sample by its posterior probability for belonging to the anomalous component.

All these baselines are point-wise anomaly detection methods, but we obtain interval detections by grouping contiguous detections based on multiple thresholds.

We use *Average Precision* as evaluation criterion, since it is more suitable for detection tasks than AUC.



(a) Comparison of the performance of the MDI method with the baselines on the synthetic data set.



(b) Comparison of the performance of the MDI algorithm with different Time-Delay Embedding parameters (without any embedding, with fixed parameters and with automatically optimized parameters).

(c) Comparison of the effect of different proposal methods on the performance of the MDI algorithm.

Figure 14: Evaluation results on the synthetic data set

As can be seen from the results of the first experiment shown in Figure 14a, our method, particularly with the Gaussian model, clearly outperforms the baselines in almost all test cases as well as regarding overall performance. Interestingly, the Gaussian method with the original KL divergence performed better than the unbiased variant in the majority of the test cases, but fails if there are multiple anomalies. In such a case, the unbiased version does a better job, which is also the case for most of the experiments with real data we describe in the following.

For the first experiment, we have fixed the Time-Delay Embedding parameters to $k = 6, T = 2$, which we have empirically found to be a good value for this dataset. Figure 14b shows that the automatic optimization of the context window size based on mutual information as described in section 2.1.4 works equally well without any manual tuning.

Interestingly, the use of the interval proposal method described in section 2.1.3 does not only lead to a significant gain in speed, but also in performance, as can be seen from Figure 14c.

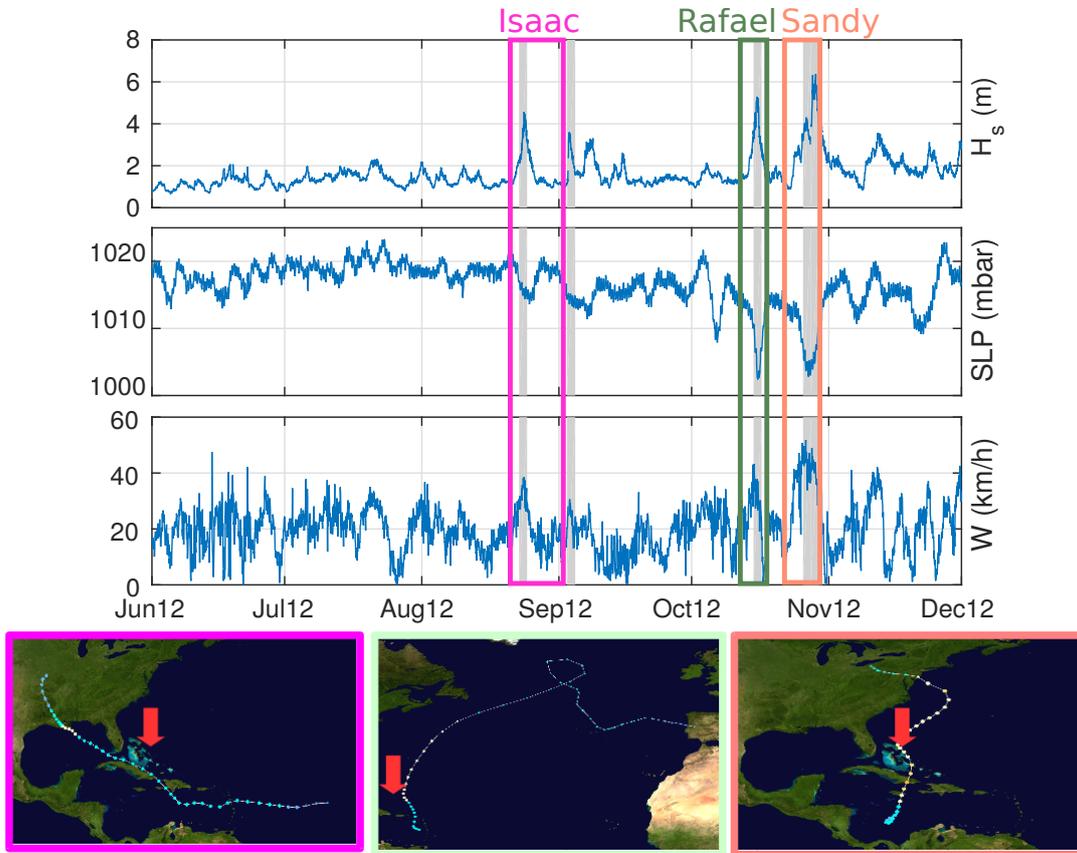


Figure 15: Upper plot: Boxes in colors represent historical hurricanes *Isaac*, *Rafael* and *Sandy* and grey shaded areas MDI Gaussian detections. The false-positive right after Isaac might be related either to a local storm or to the reminiscences from hurricane *Leslie* passing these days by Bermudas. Lower plot: Trajectories of the historical hurricanes detected (from left to right: *Isaac*, *Rafael* and *Sandy*) and location of the buoy database used (red arrow).

2.2.2 Hurricane Detection

Meteocean data (significant wave height, H_s , wind speed, W and sea level pressure SLP) in a location near the Bahamas in the Atlantic Sea (23.838°N , 68.333°W) were used in these tests. Six months of hourly data, from June 2012 until November 2012 were extracted from the National Data Buoy Center from the NOAA². This period corresponds to the Atlantic hurricane season, which in that year was specially active with 19 tropical cyclones (winds above 52 km/h) where 10 of them became hurricanes (winds above 64 km/h). In contrast to our synthetic dataset, the anomalies have an effect on multiple variables at once.

²<http://www.ndbc.noaa.gov/>

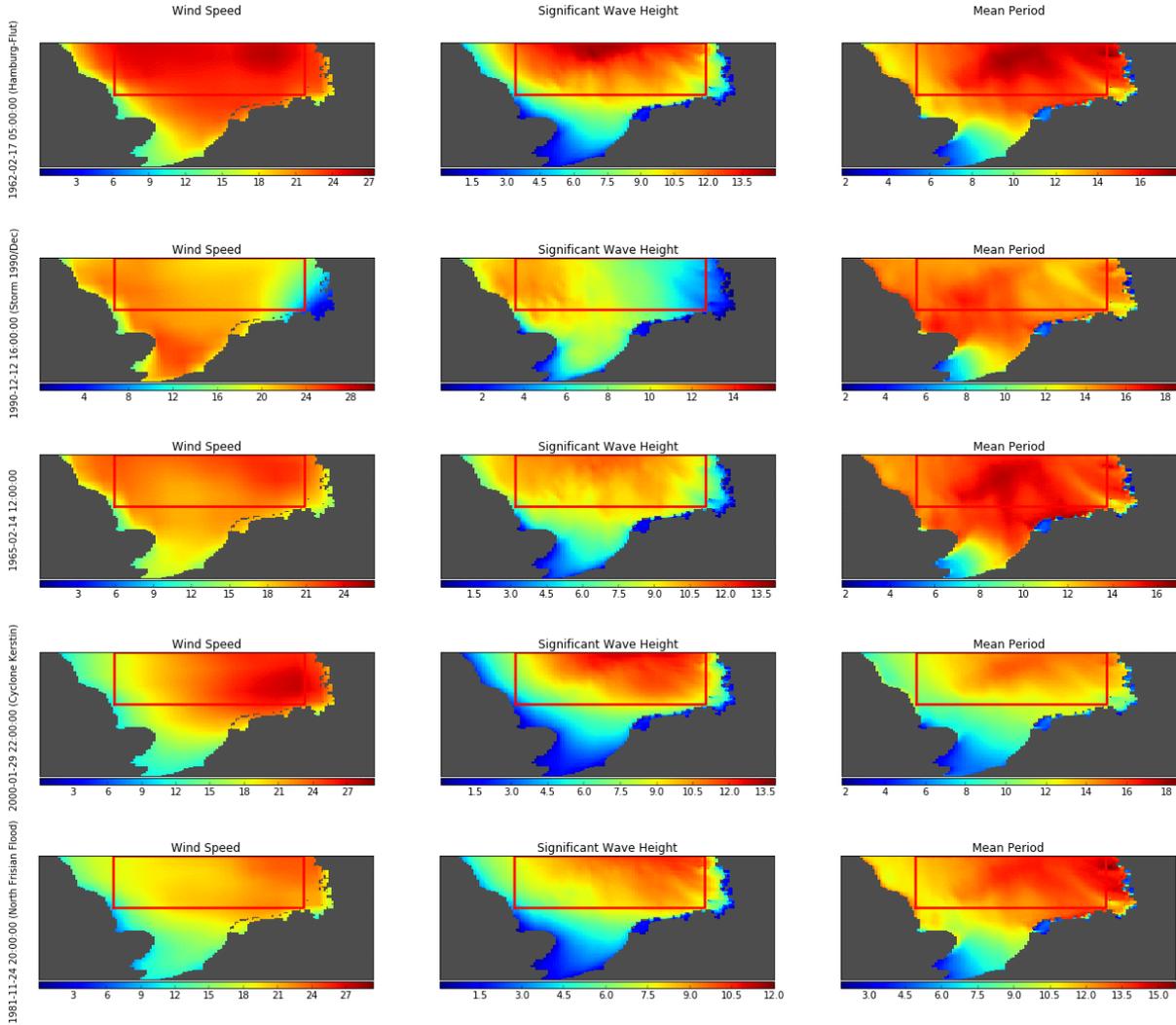


Figure 16: Heatmaps during the center of the top 5 detections on the CoastDat dataset. The red frame spans the region considered in this experiment.

We have applied the MDI Gaussian method to these three variables and compared the results with the historical hurricanes at Bahamas (Figure 15). The boxes in color represent the official duration of the three main events of that season that passed near our location, hurricanes *Isaac*, *Rafael* and *Sandy* respectively. Grey shaded areas represent the MDI intervals detected by the model. Note that in general the ground-truth areas are larger than the detections, because they span the entire lifetime of the hurricane and not just its presence at the Bahamas, see Figure 15.

2.2.3 Detection of North Sea Storms

Next, we tried to apply our method to a large-scale data set in order to detect winter storms over the North Sea: the coastDat-1 hindcast of various marine climate variables in a region over the southern North Sea over 50 years from 1958 to 2007, sampled hourly, which leads to approximately 450,000 time steps [28]. We use the variables Wind Speed, Significant Wave Height and Mean Wave Period and consider the area between 53.9°N , 0°E and 56°N , 7.7°E . Since cyclones and other storms are usually

large enough to span the entire region, we aggregate all values in that area by taking their mean to obtain a non-spatial time series. We then run a full scan over all intervals of size between 12 and 72 hours using the unbiased Gaussian method with a Time-Delay Embedding of $k = 3, T = 1$ to obtain the top 50 detections.

To assess the results we have compiled a database of 89 historic North Sea storms from several sources. Matching the detections of the algorithm against that database exposed promising results: 28 out of the top 50 detections (7 out of the top 10) can be associated with historic storms. We have manually inspected the remaining 22 detections and found that almost all of them are north sea storms as well, which we just do not have in our rather small database. Only 4 of them 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.

Figure 16 shows a heatmap of the three variables under consideration during the middle of the detected time-frame for the top 5 detections.

Using the interval proposal method described in section 2.1.3 instead of a full scan enables us to process this entire large time series in less than a second without changing the top 10 detections significantly.

We have also repeated this experiment after applying Z-Score deseasonalization, 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.

2.3 Preliminary tests with BACI data

After we have extensively tested our algorithm with synthetic and real data from different institutions outside the project, we are ready to start testing it with data provided by other partners of the BACI consortium.

2.3.1 Albedo and reflectance

We have used albedo and reflectance data from MODIS Terra and Aqua covering from 2001 until 2015. From these variables we have two locations: a 9x9 pixels area around the Hainich FLUXNET tower in Germany (51.0792° N, 10.4530° E) and another 9x9 pixels area in Somalia (6.0° S, 47.05° E). These data were developed within WP 2.

In both locations we have analyzed separately the 3 albedo measurements: visible (albedo_vis), near infrared (albedo_nir) and short wave infrared (albedo_swir) and 6 from the 7 bands of surface bidirectional reflectance (refl_b1 to refl_b7). Band 6 in MODIS collection presented some issues and we have decided to remove it during these tests.

For the tests performed in each location we have used the following settings to extract the 10 most divergent intervals.

- Divergence model used: Kullback-Leibler
- Distribution model for the distribution: Gaussian with full covariance.
- Minimum length of the intervals: 10 timesteps.
- Maximum length of the intervals: 120 timesteps.
- Number of events extracted: 10.
- Preprocessing to the time series: Standarization.
- Time-delay embedding dimension: 3.
- Search strategy: full scan.
- Overlap threshold: 0%.

We have reduced the spatial dimension by estimating the mean behavior of each variable along the 81 points that conform the 9x9 spatial grid. This way he have obtained a unique time series for each variable to which apply our method.

Somalia Figure 17 upper plot represents the 10 most divergent intervals in the albedo measurements (albedo_vis, albedo_nir and albedo_swir) for the analyzed area in Somalia. The lower plot represents the uncertainties associated to these albedo estimations and as it can be seen they have very low values.

In Figure 18 are depicted the 10 most divergent intervals in the reflectance measurements (refl_b1 to refl_b7 without refl_b6) for the same area in the upper plot and their uncertainties in the lower one.

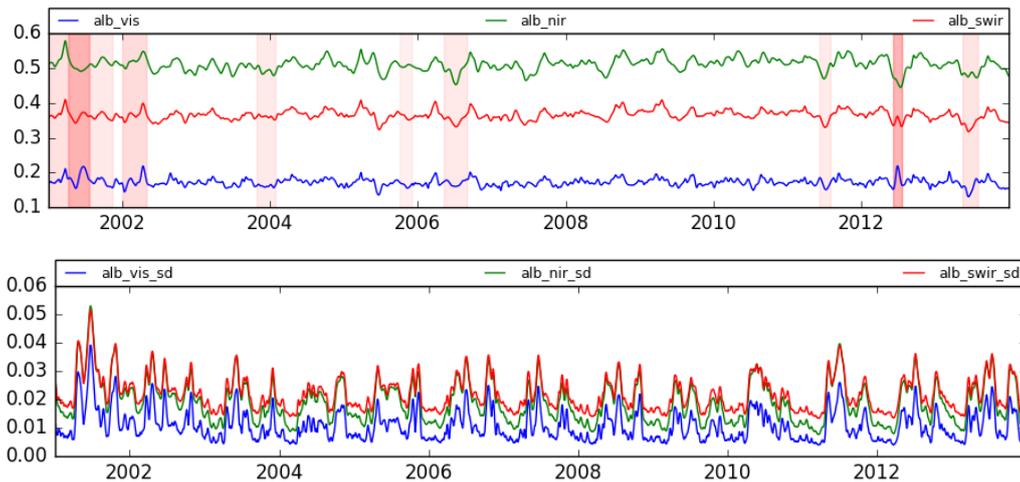


Figure 17: Somalia - Top 10 albedo intervals (upper plot) and albedo uncertainties (lower plot).

Table 3 summarizes all the events detected for both data sets in Somalia.

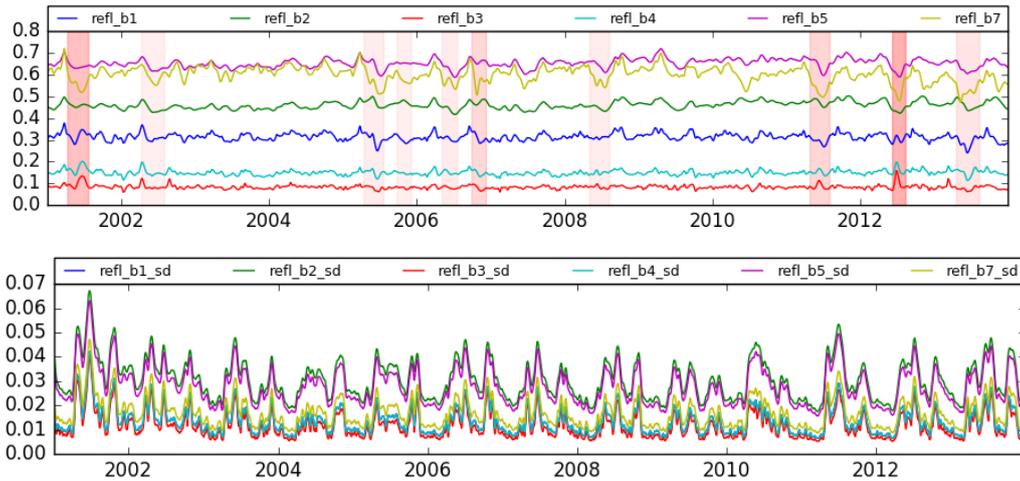


Figure 18: Somalia - Top 10 reflectance intervals (upper plot) and reflectance uncertainties (lower plot).

Table 3: Somalia - Top 10 intervals details.

Albedo			Reflectance		
Start	End	Score	Start	End	Score
2001-04-10	2001-07-20	246.41	2012-06-06	2012-08-12	334.74
2012-06-07	2012-07-22	240.05	2001-04-11	2001-07-21	299.25
2001-12-31	2002-04-30	155.48	2011-04-28	2011-08-02	250.94
2001-01-07	2001-04-10	145.39	2006-09-30	2006-12-06	246.84
2013-05-17	2013-08-02	138.62	2013-04-18	2013-08-10	216.88
2001-07-21	2001-11-13	131.1	2006-05-05	2006-07-22	200.57
2006-05-10	2006-09-04	129.32	2008-05-02	2008-08-07	198.84
2003-10-25	2004-01-28	126.97	2005-04-11	2005-07-21	198.73
2011-06-10	2011-08-04	123.91	2005-09-24	2005-12-04	192.83
2005-10-01	2005-12-04	115.75	2002-04-09	2002-08-03	192.27

Hainich Analogous to the previous figures, Figure 19 shows the events detected in the albedo measurements for the Hainich area together with its uncertainties. Figure 20 shows the 10 most divergent intervals in the reflectance measurements (refl_b1 to refl_b7 without refl_b6) for the same german area in the upper plot and their uncertainties in the lower one.

In this case, most of the events detected correspond to winter. During winter, albedo and reflectance estimations are less stable; which can be also seen in the values of its uncertainties. One reason for that is that during autumn-winter-spring there are more cloudy days and so much less satellite observations. This can lead into negative values. Although this is not physically possible it is mathematically plausible with the estimation methods used. Our Maximally Divergent Intervals method in its current version is not able to take into account the uncertainty associated to the data. But we are currently working on this, with an implementation that allow us to deal with gaps in the time series. This will let us discard data that present very high uncertainty.

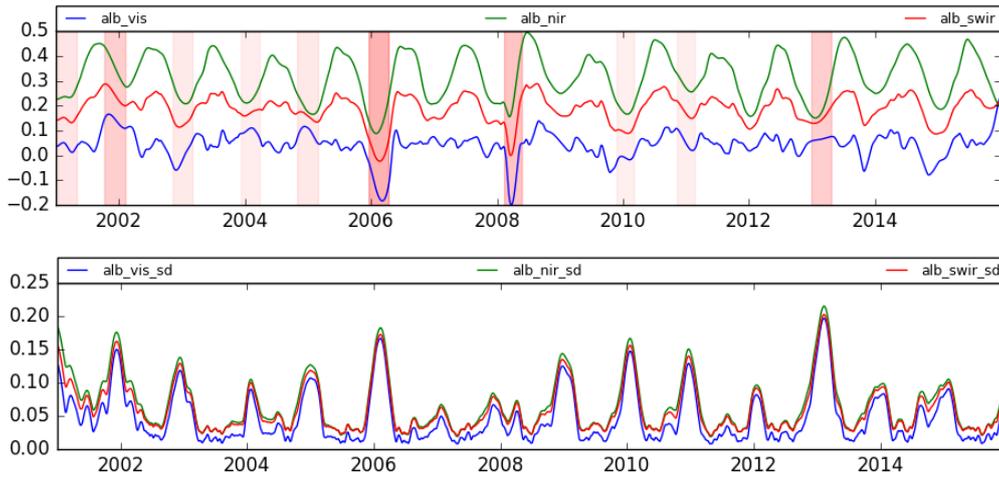


Figure 19: Hainich - Top 10 albedo intervals (upper plot) and albedo uncertainties (lower plot).

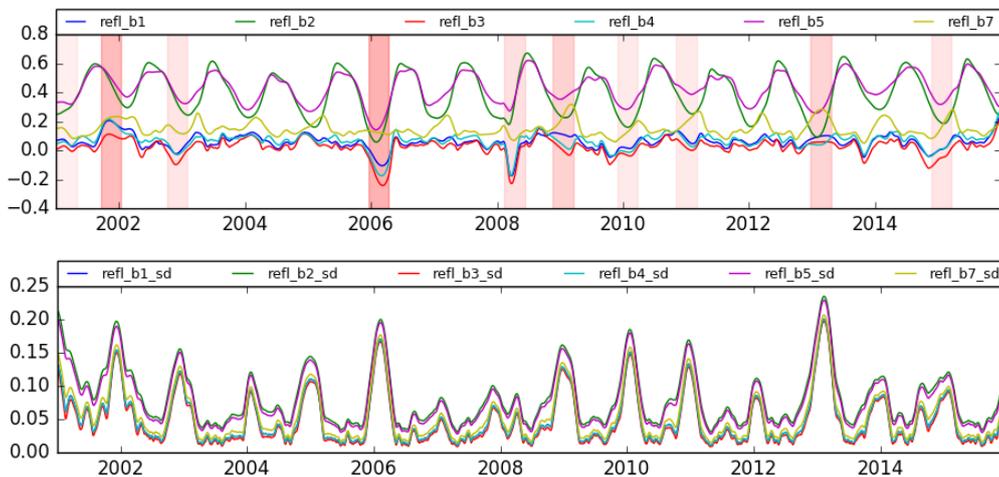


Figure 20: Hainich - Top 10 reflectance intervals (upper plot) and reflectance uncertainties (lower plot).

To summarize, Table 4 compiles all the details for the events detected in both albedo and reflectance measurements in Hainich.

2.4 Conclusions

In this section, the advances developed for the Maximally Divergent Intervals algorithm were presented. These advances have been specially focused on a more efficient computation of the Kullback-Leibler divergence and a better identification of the intervals in order to be able to deal with larger datasets.

We have carried out several experiments with artificial and real data. Additionally we have started to test data developed within the BACI project by other partners.

Table 4: Hainich - Top 10 intervals details.

Albedo			Reflectance		
Start	End	Score	Start	End	Score
2005-12-17	2006-04-16	520.81	2005-12-15	2006-04-14	740.93
2008-02-10	2008-05-26	447.96	2001-09-16	2002-01-14	692.96
2001-10-07	2002-02-04	407.62	2008-11-18	2009-03-18	580.77
2012-12-25	2013-04-24	407.11	2012-12-21	2013-04-20	571.46
2001-01-01	2001-04-29	319.36	2008-02-13	2008-06-12	513.86
2004-11-01	2005-03-01	311.48	2002-10-05	2003-02-02	490.89
2002-11-05	2003-03-05	300.01	2014-11-22	2015-03-22	486.71
2003-12-11	2004-03-27	289.25	2010-11-05	2011-03-05	464.61
2010-11-07	2011-02-22	286.83	2009-11-27	2010-03-27	463.28
2009-11-26	2010-03-10	281.65	2001-01-01	2001-05-01	439.91

Our near future work in this direction will be focused on the following points:

- *i) Spatio-temporal intervals detection (Task 5.3 - Incremental novelty detection, automatic dataset cleanup and going from novelty scores to direct detections);*
- *ii) Development of an Graphical User Interface (Task 5.3 - Incremental novelty detection, automatic dataset cleanup and going from novelty scores to direct detections);*

3 Conclusions

This report refers to the works done related to the *Task 5.2 - Near-real-time processing of non-normalized time series and large-scale adaptations* within the *Work Package 5 - Synthetic Index and Attribution Scheme: the BACIndex*.

The works done during the last period can be divided into two main parts:

We have been working on the development of a methodology to detect extreme events based on autorregressive models. We have tested two approaches, one based on coexceedances above a certain threshold and another one based on a distance metric of the joint distribution of the variables. These two methods have been applied to biosphere variables at two scales: locally and globally.

The second main focus of our work during this last period was the improvement of our Maximally Divergent Intervals method already presented in Deliverable 5.1. We have achieved better speeds in the computation time and a more efficient evaluation of the Kullback-Leiler divergence. This allow us to deal with larger databases. Artificial and real data have been tested. Within those experiments we include some initial tests with data provided by other partners of the BACI consortium.

This deliverable states the successful accomplishment of Task 5.2.

References

- [1] B Algieri, M Kalkuhl, and N Koch. A tale for two tails: Explaining extreme events in financialized agricultural markets. In *Conference Paper for the Australian Agricultural and Resource Economics Society*, 2015.
- [2] G EP Box, G M Jenkins, G C Reinsel, and G M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [3] Y Cai. Multi-variate time-series simulation. *Journal of Time Series Analysis*, 32(5):566–579, 2011.
- [4] V Chandola, A Banerjee, and V Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.
- [5] Ph Ciais, M Reichstein, N Viovy, A Granier, J Ogée, V Allard, M Aubinet, N Buchmann, C Bernhofer, A Carrara, et al. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437(7058):529–533, 2005.
- [6] R D Cook and D M Hawkins. Unmasking multivariate outliers and leverage points: comment. *Journal of the American Statistical Association*, 85(411):640–644, 1990.
- [7] Deutscher Wetterdienst. *Wetterrekorde in Deutschland und weltweit*, (5)000/08.15, edition, 2013.
- [8] Randall Dole, Martin Hoerling, Arun Kumar, Jon Eischeid, Judith Perlwitz, Xiao-Wei Quan, George Kiladis, Robert Webb, Donald Murray, Mingyue Chen, et al. The making of an extreme event: putting the pieces together. *Bulletin of the American Meteorological Society*, 95(3):427–440, 2014.
- [9] JF Donges, C-F Schleussner, JF Siegmund, and RV Donner. Event coincidence analysis for quantifying statistical interrelationships between event time series. *The European Physical Journal Special Topics*, 225(3):471–487, 2016.
- [10] Johann Du Preez and Stephen F Witt. Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, 19(3):435–451, 2003.
- [11] J Duchi. Derivations for linear algebra and optimization. *Berkeley, California*, 2007.
- [12] Y Guancho García, E Rodner, M Flach, S Sippel, M Mahecha, and J Denzler. Detecting multivariate biosphere extremes. In *Proceedings of the 6th International Workshop on Climate Informatics:CI 2016*, 2016.
- [13] H Hotelling. Multivariate quality control. *Techniques of statistical analysis*, 1947.
- [14] T Kanungo and R M Haralick. Multivariate hypothesis testing for gaussian data: Theory and software. Technical report, Citeseer, 1995.

- [15] M B Kennel, R Brown, and H Di Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, 45(6):3403, 1992.
- [16] G M Ljung and G EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.
- [17] P Mahalanobis. On the generalised distance in statistics (vol.2, pp.49–55). *Proceedings National Institute of Science, India*. Retrieved from <http://ir.isical.ac.in/dspace/handle/1/1268>, 1936.
- [18] N H Packard, J P Crutchfield, J D Farmer, and R S Shaw. Geometry from a time series. *Physical review letters*, 45(9):712, 1980.
- [19] A Rammig, M Wiedermann, J F Donges, F Babst, W von Bloh, D Frank, K Thonicke, and M D Mahecha. Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity. *Biogeosciences*, 12(2):373–385, 2015.
- [20] M Reichstein, P Ciais, D Papale, R Valentini, S Running, N Viovy, W Cramer, A Granier, J Ogee, V Allard, et al. Reduction of ecosystem productivity and respiration during the european summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Global Change Biology*, 13(3):634–651, 2007.
- [21] C Schär, P L Vidale, D Lüthi, C Frei, C Häberli, M A Liniger, and C Appenzeller. The role of increasing temperature variability in european summer heatwaves. *Nature*, 427(6972):332–336, 2004.
- [22] J F Siegmund, T GM Sanders, I Heinrich, E van der Maaten, S Simard, G Helle, and R V Donner. Meteorological drivers of extremes in daily stem radius variations of beech, oak, and pine in northeastern germany: An event coincidence analysis. *Frontiers in Plant Science*, 7:733, 2016.
- [23] C Guedes Soares and C Cunha. Bivariate autoregressive models for the time series of significant wave height and mean period. *Coastal Engineering*, 40(4):297–311, 2000.
- [24] P A Stott, D A Stone, and M R Allen. Human contribution to the european heat-wave of 2003. *Nature*, 432(7017):610–614, 2004.
- [25] D MJ Tax and R PW Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004.
- [26] P Viola and M J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.
- [27] 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. Technical report, DTIC Document, 2011.

- [28] R Weisse, P Bisling, L Gaslikova, B Geyer, N Groll, M Hortamani, V Matthias, M Maneke, I Meinke, E Meyer, et al. Climate services for marine applications in europe. *Earth Perspectives*, 2(1):1, 2015.
- [29] J Zscheischler, R Orth, and S I Seneviratne. A submonthly database for detecting changes in vegetation-atmosphere coupling. *Geophysical Research Letters*, 42(22):9816–9824, 2015.