



Extreme anomaly event detection in biosphere using linear regression and a spatiotemporal MRF model

Yanira Guancho García^{1,2} · Maha Shadaydeh¹ · Miguel Mahecha^{2,3} · Joachim Denzler^{1,2}

Received: 2 January 2018 / Accepted: 17 July 2018 / Published online: 26 July 2018
© The Author(s) 2018

Abstract

Detecting abnormal events within time series is crucial for analyzing and understanding the dynamics of the system in many research areas. In this paper, we propose a methodology to detect these anomalies in multivariate environmental data. Five biosphere variables from a preliminary version of the Earth System Data Cube have been used in this study: Gross Primary Productivity, Latent Energy, Net Ecosystem Exchange, Sensible Heat and Terrestrial Ecosystem Respiration. To tackle the spatiotemporal dependencies of the biosphere variables, the proposed methodology after preprocessing the data is divided into two steps: a feature extraction step applied to each time series in the grid independently, followed by a spatiotemporal event detection step applied to the obtained novelty scores over the entire study area. The first step is based on the assumption that the time series of each variable can be represented by an autoregressive moving average (ARMA) process, and the anomalies are those time instances that are not well represented by the estimated ARMA model. The Mahalanobis distance of the ARMA models' multivariate residuals is used as a novelty score. In the second step, the obtained novelty scores of the entire study are treated as time series of images. Markov random fields (MRFs) provide an effective and theoretically well-established methodology for integrating spatiotemporal dependency into the classification of image time series. In this study, the classification of the novelty score images into three classes, *intense anomaly*, *possible anomaly*, and *normal*, is performed using unsupervised *K*-means clustering followed by multi-temporal MRF segmentation applied recursively on the images of each consecutive $L \geq 1$ time steps. The proposed methodology was applied to an area covering Europe and Africa. Experimental results and validation based on known historic events show that the method is able to detect historic events and also provides a useful tool to define sensitive regions.

Keywords Autoregressive models · Mahalanobis distance · Markov random field model · Spatiotemporal event detection

✉ Maha Shadaydeh
maha.shadaydeh@uni-jena.de

1 Introduction

Technological developments from the last decades offer unprecedented opportunities to monitor the Earth system. In particular, the derived downstream data products are very valuable to understand processes at the land surface. International research projects like ESDL¹ and BACI² are joint efforts to provide free-of-charge, unified, and high quality Earth Observations (EOs) from satellite-based remote sensing measurements. Within this framework, the concept of the ‘Earth System Data Cube’ arose as a practical and intuitive way of storing and representing multivariate spatiotemporal databases.

The ability to detect and monitor anomalous behavior in multivariate environmental time series is crucial. These events are signals of changes in the underlying dynamical system and their detection can be used as an early warning system for land ecosystems. Classical extreme value theory (Coles 2001; Dey and Yan 2016) cannot be an option since the length of existent EOs data so far is relatively short (up to maximal three decades). Recently, Zscheischler et al. (2014) and Zscheischler et al. (2014) proposed an univariate approach based on threshold exceedances to analyze the global interannual variability of gross primary production. The presented methodology in contrast aims to tackle the problem from a multivariate point of view. Then, the definition of an extreme event should also include those constellations where not a single variable is an extreme but its combination is an extreme (multivariate extreme or compound event) (Reichstein et al. 2013; Flach et al. 2017; Zscheischler et al. 2018). Therefore, the extrapolation from the univariate to the multivariate case is not trivial.

A common approach for multivariate analysis in geoscience is to look for those events where multiple variables present abnormal behavior simultaneously, often called co-exceedances (Donges et al. 2011; Donges et al. 2016). This approach is based on fixing a threshold at each variable and analyzing the probability of occurrence of events above those thresholds either simultaneously or with a certain lag between variables. However, this might be a very conservative approach. A good alternative which has become very popular lately is the application of copulas. Copula models are based on Sklar’s theorem which states that any multivariate distribution can be written in terms of univariate marginal distribution functions of the variables involved and a copula function that describes the dependence between these variables (Sklar 1959). Whereas copulas are well studied in the bivariate case, higher-dimensional cases still present some limitations. Elliptical (i.e., multivariate Gaussian and Student’s *t* distributions) and Archimedean (i.e., Clayton, Frank and Gumbel) families are the most suitable ones for practical multivariate applications (Ma et al. 2013; Corbella and Stretch 2013). Nonetheless, there are authors arguing against the use of copulas and that they do not present any particular advantage when dealing with multivariate distributions (Mikosch 2005).

The main objective of this study is to propose a methodology to detect abnormal events in multivariate environmental time series. By combining different statistical methods, we are able to tackle the spatiotemporal dependencies. The methodology we propose can be divided into two main steps: feature extraction and event detection. The first step is based on the assumption that the time series of each variable can be represented by an autoregressive moving average (ARMA) process, and anomalies are those time instances that are not well represented by the estimated ARMA model (Chandola et al. 2009). We use the

¹ earthsystemdatalab.net/.

² <http://baci-h2020.eu/>.

Mahalanobis distance as a measure of the deviation of the multivariate residuals (difference between the observations and ARMA model output) at certain time step from their joint distribution.

For the second step, two event detection methods are presented in this paper. The first is to use a fixed threshold at a certain percentile of the Mahalanobis distance distribution applied on each time step independently. However, adjacent points in time and space are most likely to belong to the same event, whether it is normal or anomalous. The exploitation of the spatiotemporal regularity of the obtained novelty score can, on one hand, help to reduce the uncertainty in the estimation of the Mahalanobis distance from the noisy observations of a single point in time and space, and on the other hand, can help to directly define the spatial and temporal extent of the detected events. Hence, as an alternative solution, we propose to approach the problem of detecting abnormal events as detection of spatiotemporal clusters of high novelty score (Mahalanobis distance). Based on the proposed approach, the statistics (mean and variance) of the detected clusters, rather than a fixed percentile threshold, can be used to define the intensity of the anomalies. This is advantageous since the optimal selection of a fixed percentile threshold might vary according to the season as well as the climate area.

Markov random fields (MRFs) (Geman and Geman 1984) provide an effective and theoretically well-established mathematical tool for integrating spatiotemporal dependency into the classification of image time series (Melgani and Serpico 2003; Benedek et al. 2015). To this end, the obtained Mahalanobis distance over the entire study area is treated as a time series of images. We use an adaptation of the multi-layer fusion MRF classification model presented in Sziranyi and Shadaydeh (2014) and Shadaydeh et al. (2017) for the classification of this Mahalanobis distance images into three classes, *intense anomaly*, *possible anomaly* and *normal*.

The remainder of this article proceeds as follows. Section 2 gives a short description of the used data and study area. In Sect. 3, the steps of the methodology are explained in detail. Experimental results and validation based on known historic events are presented in Sect. 4. Finally, a conclusion is drawn in Sect. 5.

2 Data and study area

Data from the Earth System Data Cube (ESDC) developed within the ESDL project have been used as the primary source of biosphere data for this study. The ESDC comprises spatiotemporal data consisting of: time, latitude, longitude and multivariate Earth Observations. The version used in this study covers the period from January 2001 to December 2012 with 8 daily observations and a spatial grid with a resolution of 0.25°. More than 30 biosphere and atmosphere parameters are included in this database. Out of these variables, we have used those 5 that mainly measure the terrestrial biosphere activities: Gross Primary Productivity (GPP), Latent Energy (LE), Net Ecosystem Exchange (NEE), Sensible Heat (SH) and Terrestrial Ecosystem Respiration (TER), which were kindly provided by the FLUXCOM initiative (Tramontana et al. 2016).

The study area comprises Africa and Europe (see Fig. 1). This area was defined as the main study area within the EU project BACI. The BACI project aims to develop a 'Biosphere Atmosphere Change Index' to detect climate-induced ecosystem changes and to assess their impacts in socioeconomical and ecological processes.

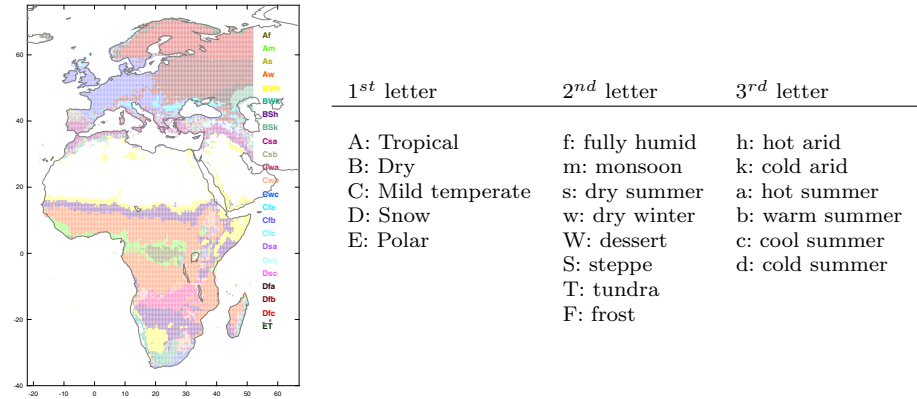


Fig. 1 Area of study clustered according to the Köppen climate classification. Gaps represent areas where there is no data available

3 Methodology

The methodology we propose can be divided into the following three steps: preprocessing, feature extraction and event detection. Methodology workflow is illustrated in Fig. 2. Each step is described in more detail in the following subsections.

3.1 Preprocessing

Desasonalization and normalization To avoid inconsistencies later, data needs to be pre-processed. We have applied techniques commonly used in environmental sciences; initially, the seasonal pattern usually present in environmental variables has been removed. In order to do so, we have subtracted the mean seasonal cycle. Then the remaining variables were normalized by subtracting its mean, μ , and dividing by its variance, σ . This is done for all the 5 variables locally at each pixel of the grid.

Regionalization Once the seasonality has been removed and the variables have been normalized, the grid was clustered into regions of similar climate conditions. This regionalization was done according to the climate types defined by the Köppen Climate Classification (Chen and Chen 2013). The Köppen Climate Classification is a widely used vegetation-based empirical clustering that divides the world in up to 31 climate regions.

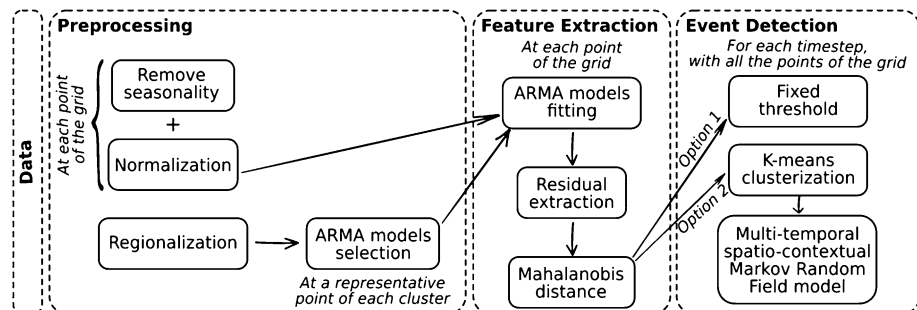


Fig. 2 Flowchart of the proposed methodology

From these 31 climate regions, 23 are present in our study area. Figure 1 shows the climate regions with the legend explaining the codes that define them.

3.2 Feature extraction

3.2.1 ARMA models

An abnormal event can be defined as those points within the time series that are not well represented by a previously fitted statistical model (Chandola et al. 2009). Following this intuitive concept, we have applied an autoregressive moving average (ARMA) model and afterward computed the residuals between the model and the data. Those points where the differences (residuals) between model and data are significantly high can be considered as abnormal events that the model is not able to represent correctly. 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 ε_t is an error term assumed to be i.i.d. Gaussian noise.

For each climate region, a representative point that is geographically centered in the region and hence reflects its average behavior has been selected. A univariate ARMA model for each of the 5 variables has been fitted, for every representative point. In order to select the best model order (p, q), a Bayesian Criterion (Schwarz 1978) was applied to all the possible combinations between (0,0) and (5,5). Table 1 shows the selected ARMA order for each variable and climate region.

Although there are multivariate approaches available (e.g., Cai 2011; Soares and Cunha 2000), we have decided to work with univariate ARMA models independently fitted to each variable at each point due to higher flexibility and easier interpretation. In Preez and Witt (2003), the authors 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 cross-correlations between variables are not particularly strong. Multivariate models involve a greater number of parameters, which becomes a disadvantage for rather short time series, while their performance is comparable to univariate approaches.

Accordingly for each climate region and variable, we have estimated the parameters (p, q) of the ARMA models to be fitted. Then we proceed with the entire grid, fitting for each point an ARMA(p_{ij}, q_{ij}), where i refers to the climate region and j stands for the variable (see Table 1). Note that there are some variables where the selected ARMA model is of order (0, 0), in those cases, the Bayesian Criterion indicates that is better to work directly with the variables themselves instead of working with ARMA model residuals.

We have additionally tested the use of ARIMA models (autoregressive integrated moving average models) that are worthy to be used when the variables present non-stationarity. Comparing the use of ARMA and ARIMA models by means of the Bayesian Criterion, we found that including the extra parameter of the ARIMA models does not lead to better results for the relatively short-term (12 years) ESDL data. Table 2 shows the comparison between the ARMA and ARIMA models for each climate region and variable. The ARIMA(p, D, q) models introduce a non-seasonal integration term defined by the

Table 1 (p, q) ARMA parameters selected for each climate region and variable

Region	Variables				
	GPP	LE	NEE	SH	TER
Af	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]
Am	[1,0]	[0,0]	[1,0]	[0,0]	[1,0]
As	[1,0]	[0,0]	[1,0]	[0,0]	[0,1]
Aw	[1,0]	[2,2]	[1,0]	[0,0]	[1,0]
BWh	[1,1]	[1,1]	[1,1]	[1,0]	[2,0]
BWk	[1,1]	[1,1]	[1,1]	[1,0]	[0,0]
BSh	[1,0]	[0,0]	[1,0]	[0,0]	[0,0]
BSk	[1,0]	[1,0]	[1,0]	[0,0]	[1,1]
Csa	[1,1]	[1,0]	[1,1]	[1,0]	[1,0]
Csb	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]
Cwa	[1,1]	[1,1]	[1,1]	[1,0]	[1,1]
Cwb	[1,0]	[1,1]	[0,0]	[1,0]	[1,1]
Cwc	[1,1]	[1,1]	[1,1]	[1,1]	[1,1]
Cfa	[0,0]	[3,2]	[0,0]	[0,0]	[0,0]
Cfb	[4,2]	[1,0]	[1,1]	[1,1]	[1,0]
Cfc	[0,3]	[1,0]	[4,0]	[1,0]	[1,0]
Dsa	[4,2]	[1,1]	[2,0]	[1,1]	[1,1]
Dsb	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]
Dsc	[0,0]	[0,0]	[0,0]	[0,1]	[0,0]
Dfa	[1,1]	[1,1]	[1,1]	[1,1]	[1,1]
Dfb	[1,0]	[1,0]	[1,0]	[0,0]	[1,0]
Dfc	[1,0]	[0,1]	[1,0]	[1,0]	[0,1]
ET	[1,0]	[0,0]	[0,0]	[0,0]	[0,1]

parameter D . As it can be seen, D is equal to 0 in the majority of the climate regions and variables. Therefore, following the principle of parsimony, we have decided for the ARMA models.

3.2.2 Residuals

Next we proceed with the model fittings for all the pixels of the grid; by comparing the predictions of the ARMA models with the variables themselves, we obtain the time series of residuals at each point. These residuals' time series will be used to detect abnormal events. To ensure the fitness of the estimated ARMA models, we have checked the autocorrelation pattern of the residuals to ensure the absence of seasonal pattern and low correlation.

3.2.3 Mahalanobis distance

At each pixel of the grid, we have combined the residuals of the 5 variables in a vector \mathbf{x} and estimated its Mahalanobis distance (Mahalanobis 1936; Hotelling 1947). This distance measure compared to other metrics has the advantage of taking into account the shape of the joint distribution. Compared to the Euclidean distance, it does not only take the mean

Table 2 Comparison between (p, q) ARMA parameters and (p, D, q) ARIMA parameters for each climate region and variable

	ARMA					ARIMA				
	GPP	LE	NEE	SH	TER	GPP	LE	NEE	SH	TER
Af	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]
Am	[1,0]	[0,0]	[1,0]	[0,0]	[1,0]	[1,0,0]	[0,0,0]	[1,0,0]	[0,0,2]	[1,0,0]
As	[1,0]	[0,0]	[1,0]	[0,0]	[0,1]	[1,0,0]	[1,0,0]	[1,0,0]	[0,0,0]	[1,0,0]
Aw	[1,0]	[2,2]	[1,0]	[0,0]	[1,0]	[2,0,0]	[1,0,1]	[2,0,0]	[1,0,0]	[1,0,0]
BWh	[1,1]	[1,1]	[1,1]	[1,0]	[2,0]	[1,0,1]	[1,0,1]	[2,0,0]	[1,0,1]	[1,0,0]
BWk	[1,1]	[1,1]	[1,1]	[1,0]	[0,0]	[2,0,0]	[1,0,0]	[1,0,1]	[1,0,0]	[1,0,0]
BSh	[1,0]	[0,0]	[1,0]	[0,0]	[0,0]	[1,0,0]	[1,0,1]	[1,0,0]	[1,0,0]	[1,0,0]
BSk	[1,0]	[1,0]	[1,0]	[0,0]	[1,1]	[1,0,2]	[1,0,0]	[2,0,0]	[1,0,1]	[1,0,0]
Csa	[1,1]	[1,0]	[1,1]	[1,0]	[1,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]
Csb	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[1,0,0]	[1,0,0]	[3,1,3]	[0,0,0]	[1,0,0]
Cwa	[1,1]	[1,1]	[1,1]	[1,0]	[1,1]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]
Cwb	[1,0]	[1,1]	[0,0]	[1,0]	[1,1]	[1,0,1]	[1,0,0]	[3,0,0]	[1,0,0]	[1,0,1]
Cwc	[1,1]	[1,1]	[1,1]	[1,1]	[1,1]	[1,0,1]	[1,0,0]	[1,0,1]	[1,0,0]	[1,0,0]
Cfa	[0,0]	[3,2]	[0,0]	[0,0]	[0,0]	[1,0,0]	[1,0,0]	[1,0,1]	[1,0,1]	[1,0,0]
Cfb	[4,2]	[1,0]	[1,1]	[1,1]	[1,0]	[0,0,0]	[0,0,0]	[0,0,0]	[0,0,0]	[0,0,0]
Cfc	[0,3]	[1,0]	[4,0]	[1,0]	[1,0]	[1,0,0]	[1,0,0]	[2,0,0]	[0,0,2]	[1,0,0]
Dsa	[4,2]	[1,1]	[2,0]	[1,1]	[1,1]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]
Dsb	[1,0]	[1,0]	[1,0]	[1,0]	[1,0]	[1,0,0]	[0,0,1]	[1,0,0]	[1,0,0]	[1,0,0]
Dsc	[0,0]	[0,0]	[0,0]	[0,1]	[0,0]	[1,0,0]	[1,0,0]	[2,0,0]	[0,0,0]	[1,0,0]
Dfa	[1,1]	[1,1]	[1,1]	[1,1]	[1,1]	[1,0,0]	[2,0,0]	[2,0,0]	[1,0,1]	[1,0,0]
Dfb	[1,0]	[1,0]	[1,0]	[0,0]	[1,0]	[1,0,1]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]
Dfc	[1,0]	[0,1]	[1,0]	[1,0]	[0,1]	[1,0,0]	[2,0,0]	[2,0,0]	[0,0,0]	[0,0,1]
ET	[1,0]	[0,0]	[0,0]	[0,0]	[0,1]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[0,0,1]

but also take the covariance matrix into account. The Mahalanobis distance (in squared units) is defined as:

$$d_m(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \tag{2}$$

where $\bar{\mathbf{x}}$ and Σ are the mean and covariance matrix of the multivariate residuals vector \mathbf{X} respectively. The mean and the covariance were estimated considering the entire time series. This was the best way to do so in our case due to the short length of the time series used together with its coarse temporal resolution.

Figure 3 shows the scatter-plot matrix of the residuals at a certain location (50,875°N, 11,625°E). The diagonal of the matrix shows the autocorrelation plots of the 5 variables, while the rest of the subplots represent all the pair-wise scatter plots. The colors assigned to the dots in the scatter plots are associated to the Mahalanobis distance estimated with the 5 variables. Although the residuals' joint distribution does not follow a multivariate Gaussian distribution, it does not present a clear multimodality. Therefore, Mahalanobis distance is still a robust approach as argued by Warren et al. (2011). The negligible autocorrelation

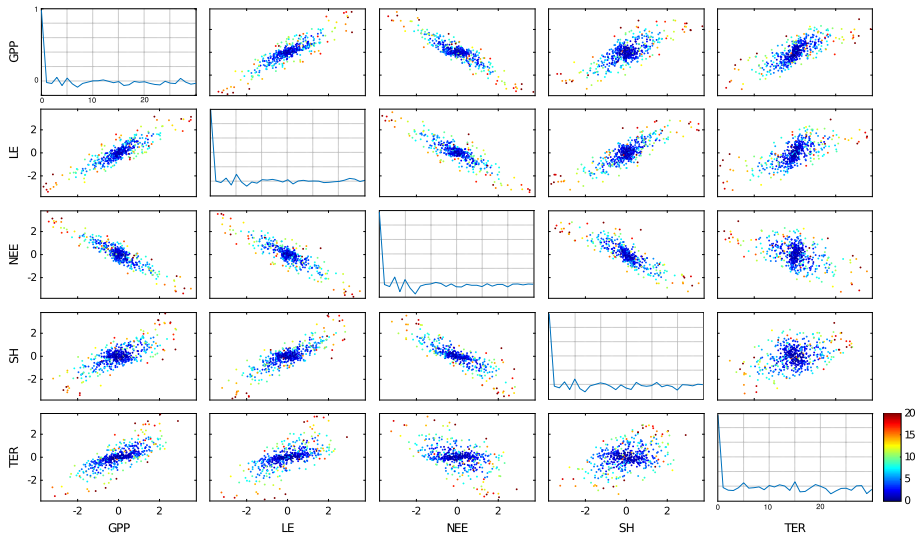


Fig. 3 Scatter-plot matrix and autocorrelation plots of the ARMA residuals at the location 50,875°N, 11,625°E. The color of the dots represents the Mahalanobis distance associated to the residuals

values for lags greater than zero indicate that the models are well fitted. This corroborates the use of univariate ARMA models and their correct fit for this study case considering the low temporal resolution of the data. However, multivariate autoregressive models could be implemented at the feature extraction step for higher temporal resolution data without changing the next event detection step.

3.3 Event detection

Once the Mahalanobis distance for all the points of the grid has been estimated, the following question arises: how could we discern between normal and abnormal values of this metric? At this point, we have considered two options. As a first approach, we have used a fixed threshold at a certain percentile of the Mahalanobis distance distribution. And as a second and more complex approach, we define abnormal events as those spatiotemporal clusters of high novelty score (Mahalanobis distance). To this end, we use unsupervised K-means clustering followed by MRF spatiotemporal smoothing. Details of these two approaches are described in the following two sections.

3.3.1 Event detection using fixed threshold

The easiest way to distinguish between normal and abnormal events is to set a threshold and look for the events surpassing this threshold. We are interested only in the largest values of the Mahalanobis distance; therefore, we have set the threshold at the 97.5th percentile of its distribution (all the Mahalanobis distance values along the entire region). We have then looked for the events above this threshold, which are the 2.5% of observations with the largest Mahalanobis distance.

3.3.2 Event detection using spatiotemporal MRF model

Markov Random Field models are widely adopted to quantify the spatial/temporal dependency among adjacent pixels in time series of images. It represents an undirected graph where graph nodes denote image pixels, and graph edges denote conditional dependencies. The dependency between adjacent pixels can be modeled by conditional probabilities within a neighborhood system.

Sziranyi and Shadaydeh (2014) proposed a multi-layer fusion MRF model for change detection in multi-temporal optical images. This method has been further improved in Shadaydeh et al. (2017) to deal recursively with time series of images.

In this study, the obtained Mahalanobis distance over the entire study area is treated as time series of images. We use an adaptation of the model proposed in Shadaydeh et al. (2017) for the classification of time series of the Mahalanobis distance images into K classes using unsupervised K -means clustering followed by spatiotemporal MRF-based segmentation applied recursively on each $L \geq 1$ consecutive images. The K classes represent K different intensities of the novelty score clusters.

An image $\mathcal{S} = \{s_1, s_2, \dots, s_H\}$ is considered as two-dimensional grid of pixels. Let \mathcal{G}_s be the set of pixels which are neighbors of a pixel $s \in \mathcal{S}$; $\mathcal{G} = \{\mathcal{G}_s \mid s \in \mathcal{S}\}$ is a neighborhood system (Kato and Zerubia 2012) if: $\forall s, r \in \mathcal{S} : s \notin \mathcal{G}_s$; and $s \in \mathcal{G}_r \iff r \in \mathcal{G}_s$.

Each of the image pixels may take a label λ from a finite set of labels Λ . Let $\Omega = \{\omega = (\omega_{s_1}, \dots, \omega_{s_H}) : \omega_i \in \Lambda, 1 \leq i \leq H\}$ be the set of all possible labels assigned to the image pixels. The image segmentation is equivalent to a global labeling $\Omega = \{\omega_s \mid s \in \mathcal{S}\}$. The label field Ω is modeled as a Gaussian Markov Random Field. There are several methods to estimate the global optimum labeling through iteration process or graph-cutting (Kato and Zerubia 2012).

Let $d_s(t)$ denote the novelty score (Mahalanobis distance) at location s and time point t . The proposed method consists of the following four steps applied recursively for each time step t (cf. Fig. 4):

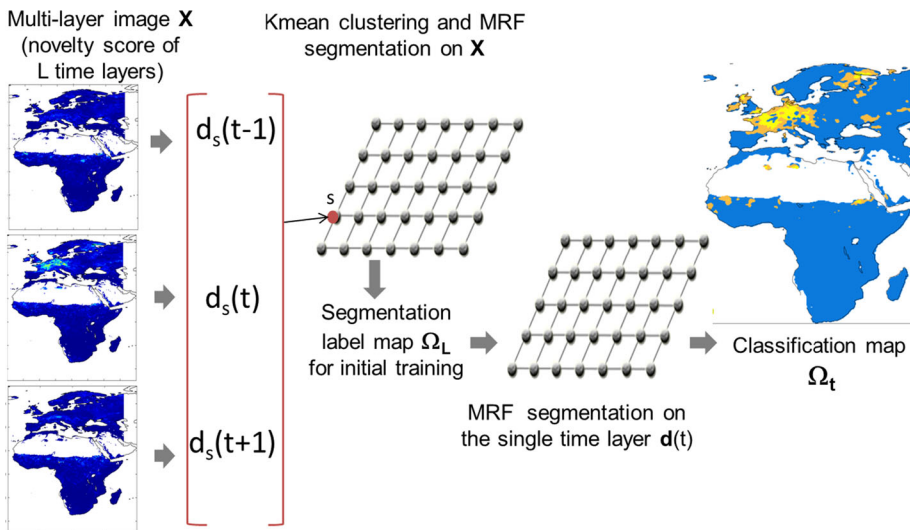


Fig. 4 The workflow of the proposed spatiotemporal MRF classification model applied on each $L = 3$ consecutive novelty score images

1. The novelty scores $d_s(t)$ for $L = 2N + 1$ successive time instants are presented by the feature vector:

$$\mathbf{d}_s = [\mathbf{d}_s(\mathbf{t} - \mathbf{N}), \dots, \mathbf{d}_s(\mathbf{t}), \dots, \mathbf{d}_s(\mathbf{t} + \mathbf{N})]^T. \quad (3)$$

2. Finding K clusters in the multi-layer image $\mathbf{X} = \{\mathbf{d}_s, s \in \mathcal{S}\}$ using the unsupervised K -means clustering algorithm.
3. Running MRF segmentation on the multi-layer image \mathbf{X} using the K -means clustering parameters to obtain the multi-layer labeling Ω_L .
4. The multi-layer labeling Ω_L is used as training map for another MRF segmentation applied on the novelty score image $\mathbf{d}(\mathbf{t}) = \{\mathbf{d}_s(\mathbf{t}), s \in \mathcal{S}\}$ resulting in a labeling Ω_t .

We use a maximum a posteriori (MAP) estimator for the label field of the MRF. The MAP estimator combines the conditional random field of the observed data $P(\mathbf{f}_s | \omega_s)$ ($\mathbf{f}_s = \mathbf{d}_s$ in Step 1, and $\mathbf{f}_s = d_s(t)$ in Step 4) and the unconditional Potts model (Potts 1952). The global labeling $\hat{\Omega}$ is defined by the energy minimum:

$$\hat{\Omega} = \operatorname{argmin}_{\Omega} \left[\sum_{s \in \mathcal{S}} -\log P(\mathbf{f}_s | \omega_s) + \sum_{r, s \in \mathcal{S}} \Theta(\omega_r, \omega_s) \right], \quad (4)$$

where $\Theta(\omega_r, \omega_s)$ is the neighborhood-energy term. It is set to zero if s and r are not neighboring pixels, otherwise Θ can be modified by applying the β homogeneity weight:

$$\Theta(\omega_r, \omega_s) = \begin{cases} 0 & \text{if } \omega_r = \omega_s \\ \beta & \text{if } \omega_r \neq \omega_s \end{cases}. \quad (5)$$

4 Experimental results and discussion

4.1 Validation process

Validating models that try to reproduce environmental processes is not trivial task. There are no well-defined ground-truth events which can be used to compare the models' performance and level of accuracy. Here, with help of experts on the topic, we have selected a list of well-known historical events that caused perturbations in the biosphere within the time span of our data.

This validation is not meant to be as detailed and precise as the topic would require. For this purpose, within the BACI project, there are specific work packages with experts in the field currently working and analyzing the models from a geophysical point of view and studying the socioeconomic and biodiversity impacts of these anomalies as well as the causes behind them. With this validation, we only want to check quantitatively the performance of the proposed methodology without detailed knowledge on the physical drivers behind the events.

Table 3 encompasses the list of historical events selected for this validation. We have selected three main events that happened at different locations within the area of study and with different causes and impacts. These three events are well documented in the literature: 1. the drought-floods at the Horn of Africa along 2006; 2. the Russian Heatwave from 2010 and 3. the severe floods in South Africa between the end of 2010 and the beginning of 2011. In 2006, the Horn of Africa, specially Somalia, confronted a series of disasters both natural and human-made. From the beginning of that year, a severe drought has caused a

Table 3 Known historical extreme events used in the validation

Event	Period	Description
1	10-11/2006	Horn of Africa floods and food crisis (Kijazi and Reason 2009)
2	06-08/2010	Russian Heatwave (Trenberth and Fasullo 2012; Coumou and Rahmstorf 2012)
3	12/2010-01/2011	Floods in Southern Africa (Nilsson 2012)

dramatic decrease in crops threatening almost 2 million people with starvation. This was aggravated by the ongoing political conflict as well as the worst flooding in a decade by the end of the year (Isar 2010). In summer 2010, western Russia, Ukraine, Belarus, Georgia and Kazakhstan experienced a heat wave that led to historical warm records in several cities and a considerable high number of casualties related to the extremely hot temperatures (Barriopedro et al. 2011; Trenberth and Fasullo 2012; Coumou and Rahmstorf 2012). By the end of 2010 and the beginning of 2011, a series of floods took place in the southern part of Africa, mainly in South Africa, Mozambique, Zimbabwe and southern Botswana. The heavy rainfalls caused river-floods on the main rivers in that area. Severe damages on property and loss of human lives were reported in the countries affected. This event has been linked to the La Niña event that occurred that winter and led to similar events in several other countries around the world (Nilsson 2012).

In the following two sections, the results given by the two event detection methods for those 3 events are presented.

4.2 Results for fixed threshold event detection

For each of the events presented before, the areas with higher values of Mahalanobis distance are shown in Figs. 5, 6 and 7 with the largest contour line at the 97.5th percentile threshold marked in red.

Figure 5 shows the months of October and November 2006. At that time, an important food crisis caused by large floods devastated the Horn of Africa. Figure 6 clearly depicts the Russian Heatwave of 2010. It can be seen how the anomaly moves from East to West and finally hits the Nordic countries. The time steps between December 2010 and January 2011 are plotted in Fig. 7. During those months, serious floods took place in southern Africa.

4.3 Results for spatiotemporal MRF model event detection

Considering the same events, we applied the four steps of the spatiotemporal MRF classification model presented in Sect. 3.3.2 to each time step. The classification maps of the Mahalanobis distance into the classes *intense anomaly*, *possible anomaly* and *normal* are shown in Figs. 8, 9 and 10. The intensity of the *possible* and *intense anomaly* classes for each time step are annotated in the subplots' legends, respectively, as PA and IA.

In all the experiments, we set the homogeneity weight $\beta = 1$. We used a graph cut-based α -expansion algorithm for the energy minimization of the MRF with the implementation accompanying Szeliński et al. (2006). The selection of the value of L is based on data's temporal resolution. In our experiments, we set $L = 3$ ensuring that the data belong

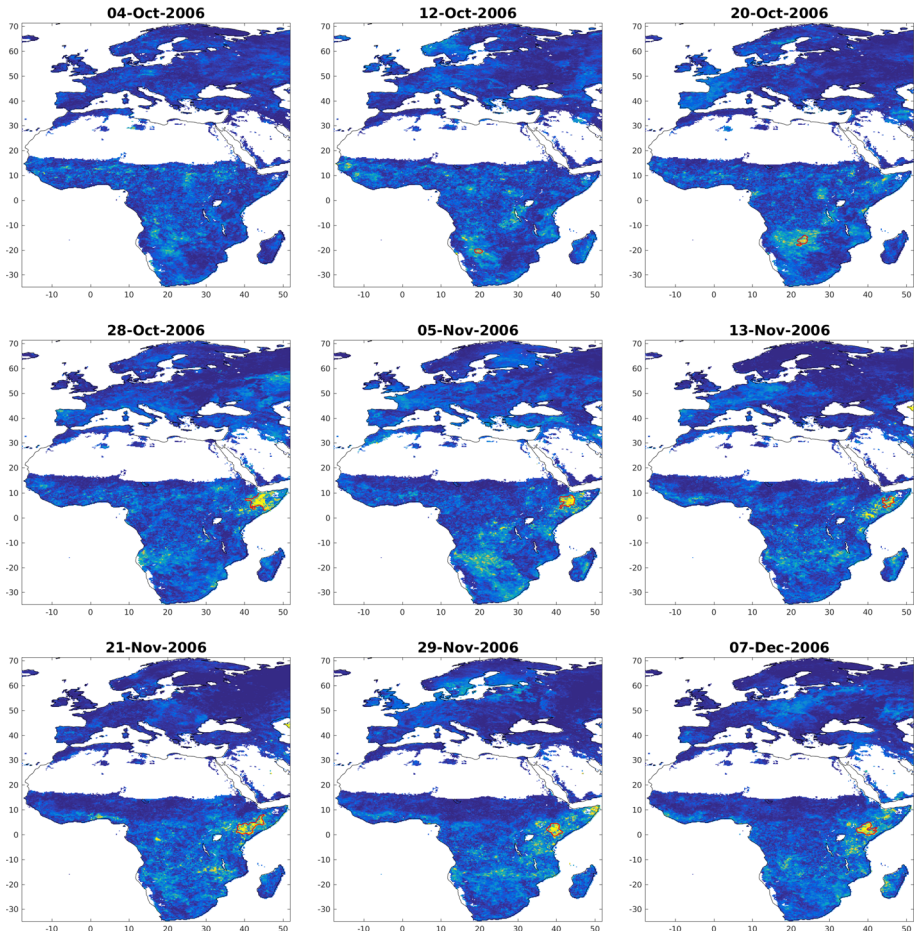


Fig. 5 Floods at the Horn of Africa from October to November 2006 (Mahalanobis distance images). The largest contour line at the 97.5th percentile of the Mahalanobis distance distribution is marked in red

to the same season. However, it is possible to use higher values of $L = 3$ when using data with a higher temporal resolution. We initially assume that we have $K = 9$ clusters. We use MRF segmentation with 9 classes representing 9 different intensities of the Mahalanobis distance. For anomaly detection, we then merge all classes with mean values exceeding the value of the Mahalanobis distance distribution (chi-square distribution) at the 97.5th percentile as the *intense anomaly* class; classes with mean values between the 97.5th and the 95th percentiles are classified as *possible anomaly*, and classes with mean value below the 95th percentile are considered *normal*. The mean and variance of the new three clusters are calculated again from the points of the merged sub-clusters.

4.4 Discussion

The plots shown in the previous sections demonstrate the methodology's capability of effectively detecting abnormal events in the biosphere. Both of the proposed event

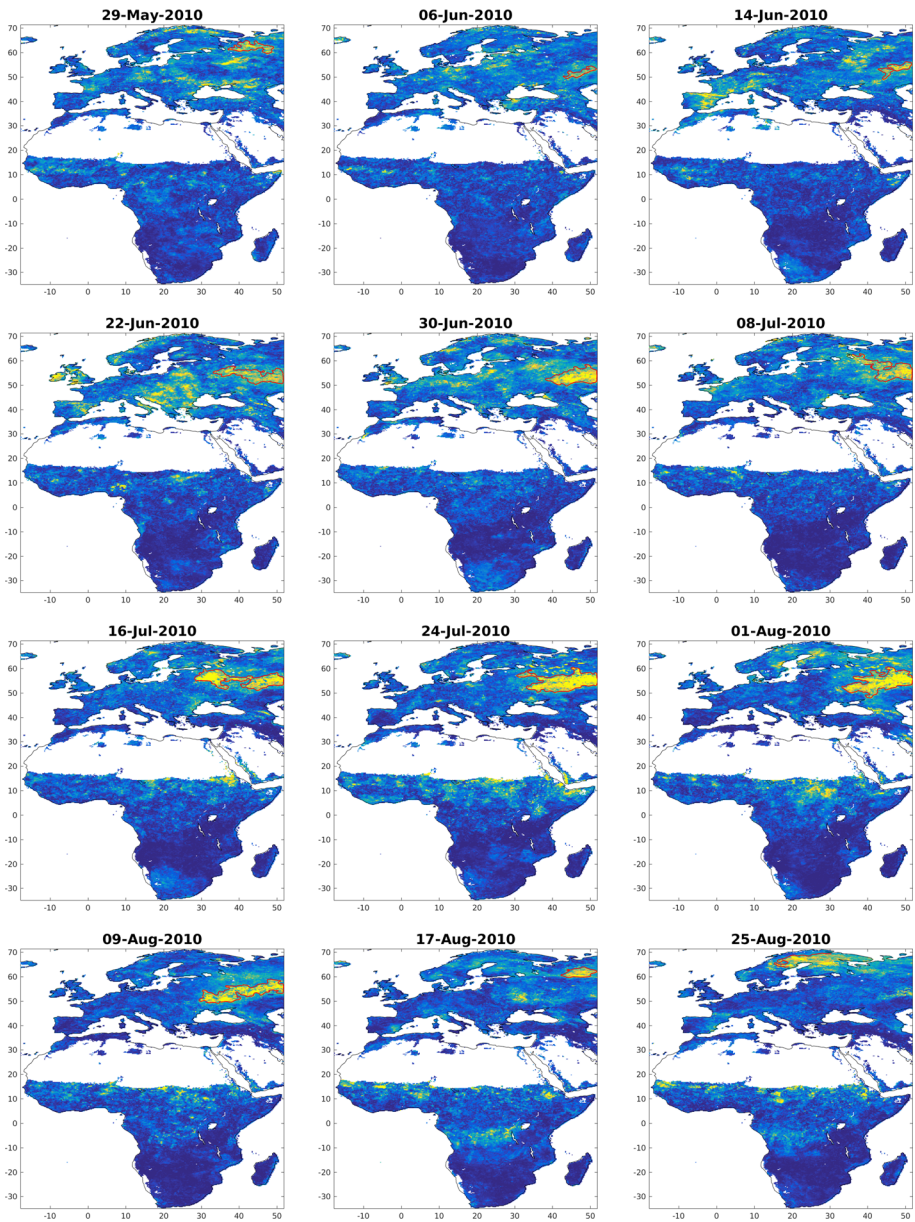


Fig. 6 Russian Heatwave from May to October 2010 (Mahalanobis distance images). The largest contour line at the 97.5th percentile of the Mahalanobis distance distribution is marked in red

detection methods have advantages and disadvantages: fixing a threshold is a simple approach that allows for a quick analysis of the results. This approach ensures that the detected events at least have the same intensity (threshold) but the selection of the threshold requires a meditated decision. The second proposed method, on the contrary, automatically detects the spatial extents as well as the thresholds between classes at each

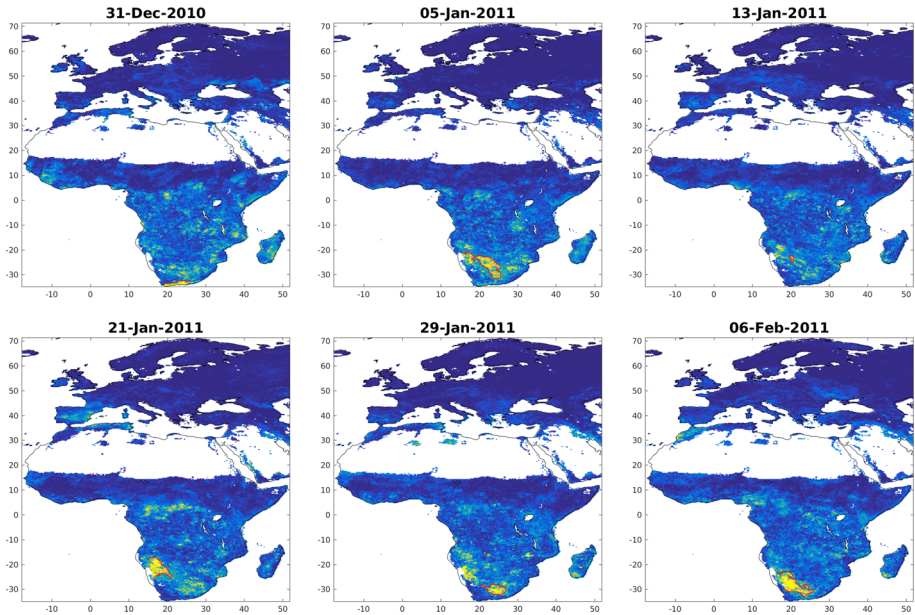


Fig. 7 Floods in southern Africa from December 2010 to January 2011 (Mahalanobis distance images). The largest contour line at the 97.5th percentile of the Mahalanobis distance distribution is marked in red

step. This allows for a more flexible definition of the detected anomalies but makes the comparison of events more complicated since they might have different intensity levels.

Taking a closer look on the plots of the three selected events, we can observe the following: Larger events like the Russian Heatwave are well represented by both methods (Figs. 6 and 9). The sequence of time steps clearly shows how the anomaly moved from East to West with a later displacement toward northern latitudes. While initially the anomalies seem to be sparse across entire Europe, they become more concentrated around western Russia, Ukraine and Belarus. In terms of intensity from the MRF models, the event increases in intensity starting in May until it reaches a peak in the last 2 weeks of July before decreasing again.

The event at the Horn of Africa in 2006 (Figs. 5 and 8) is detected more clearly using the MRF-based method than with the fixed threshold. This can be related to the fact that the intensities for this event are a bit lower than the ones obtained for the Russian Heatwave.

Finally, the event in southern Africa between the end of 2010 and the beginning of 2011 is shown in Figs. 7 and 10. Although being shorter in time (only 6 time steps), this anomaly is clearly detected by both methods. Here, another difference between both methods is present: the definition of three classes established from the MRF model-based method allows for a better spatial definition of the event compared to the fixed threshold method.

It should be noted that the low temporal and spatial resolution of the data used in this study represents a limitation for obtention of anomalies of short duration or small spatial extension.

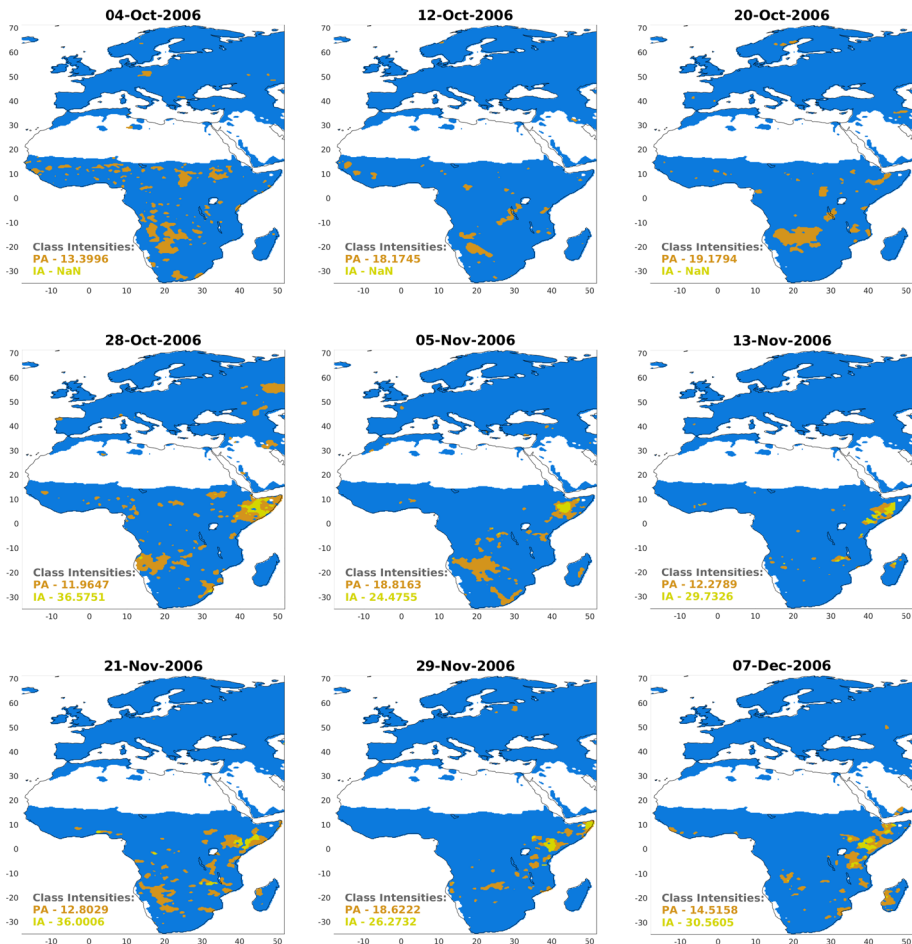


Fig. 8 Floods at the Horn of Africa from October to November 2006 (spatiotemporal MRF classification maps). The intensity of the *possible* and *intense anomaly* classes is annotated in the subplots’ legends as PA and IA, respectively

5 Conclusions

A new methodology to detect anomalies in biosphere time series has been described. The procedure is computationally efficient and hence practical to apply. Our approach comprises two main steps after preprocessing the data: feature extraction and event detection. The feature extraction is achieved by means of autoregressive models followed by the estimation of Mahalanobis distance of the multivariate residuals. Event detection is based on the concept of detecting clusters of high novelty score using a spatiotemporal MRF classification model.

The proposed methodology has been applied to a large area that covers Europe and Africa. Results show that the method is able to detect the spatial extent of three known historic events and also provides a useful tool to define sensitive regions.

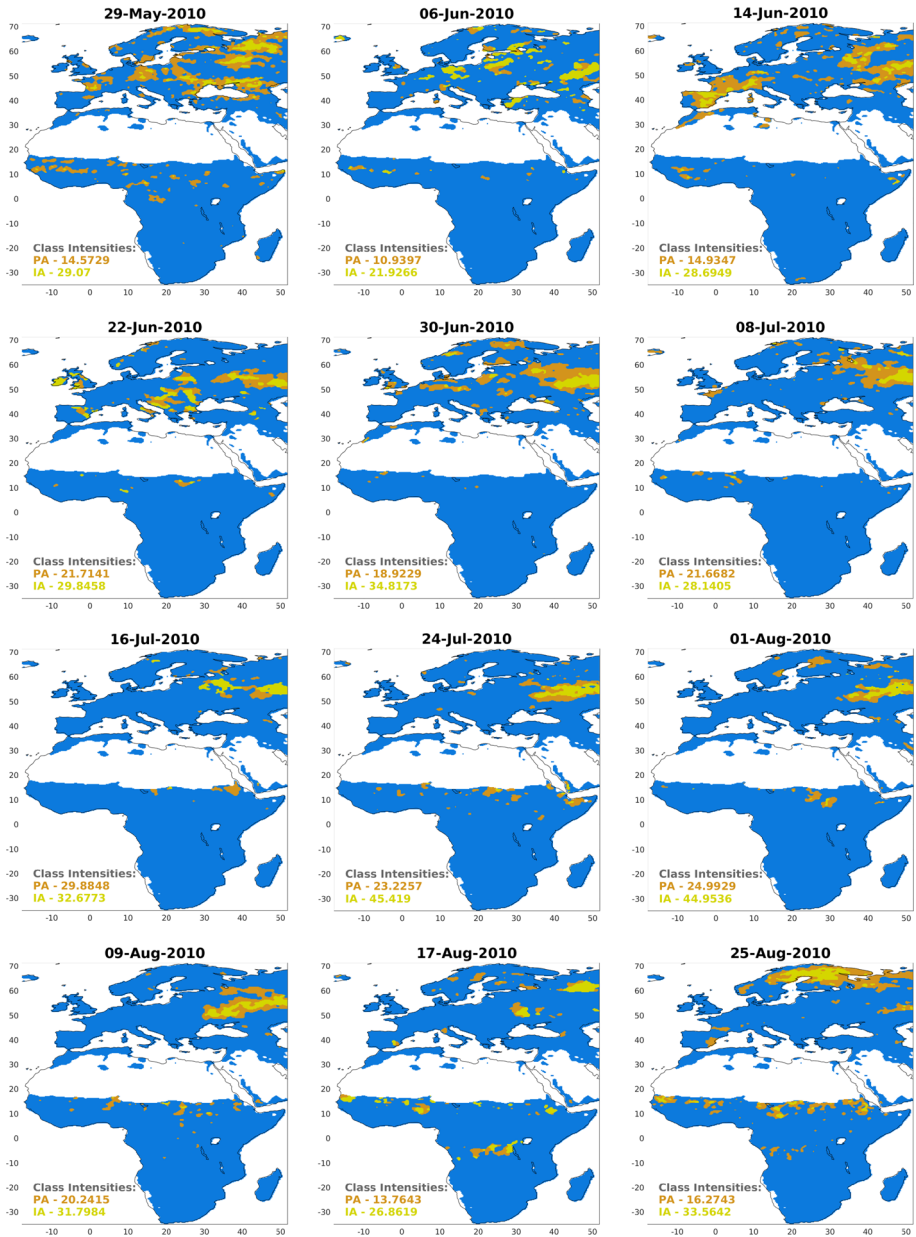


Fig. 9 Russian Heatwave from May to October 2010 (spatiotemporal MRF classification maps). The intensity of the *possible* and *intense anomaly* classes are annotated in the subplots' legends as PA and IA, respectively

The modular nature of the methodology allows for further improvements in different sub-steps. It is also envisaged that the methodology can be applied to other kind of data (e.g., atmospheric parameters). This however could require different preprocessing steps.

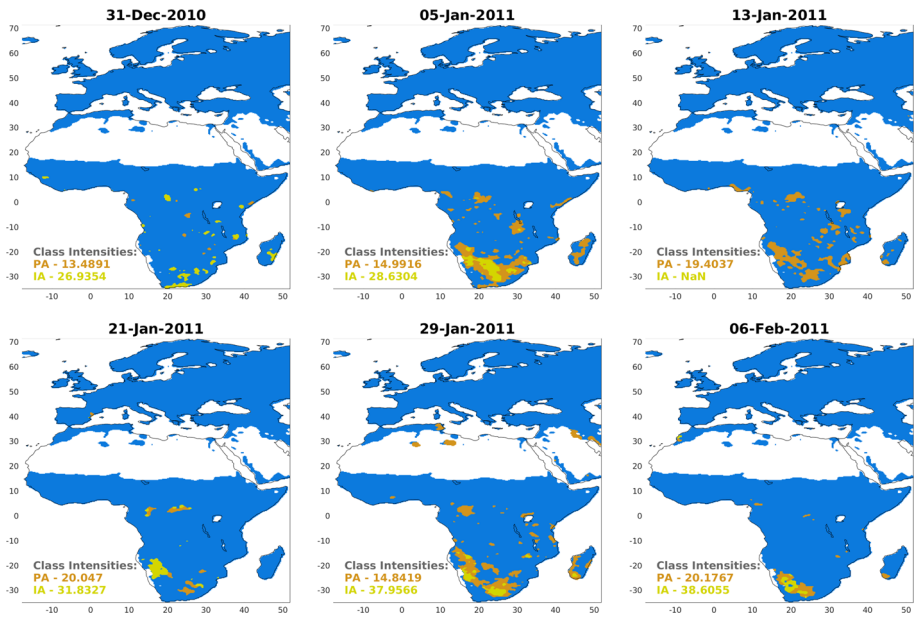


Fig. 10 Floods in southern Africa from December 2010 to January 2011 (spatiotemporal MRF classification maps). The intensity of the *possible* and *intense anomaly* classes are annotated in the subplots' legends as PA and IA, respectively

The proposed methodology could be enhanced by the use of other models than the univariate ARMA models suggested here; for example, with longer time series, multivariate regressive models could be implemented at the feature extraction step without changing significantly the event detection step. Further research is also needed for a better suited climate classification. Although the use of the Köppen Climate classification provides a useful approach, a better definition of climate regions with transitional borders between regions, instead of hard ones could avoid seasonal effects observed in the models' behavior. Finally, regarding the spatiotemporal MRF classification model, one potential improvement would be the use of a hyper-parameter-free clustering method instead of the *K*-means. In case of availability of training data for normal and/or anomalous classes, it is possible to use supervised training for the spatiotemporal MRF model instead of the unsupervised *K*-means clustering.

Acknowledgements The data used in this study were kindly provided by the ESA project ESDL: Earth System Data Lab. The authors are grateful to the FLUXCOM initiative (<http://www.fluxcom.org>) who provided the data accessed through the Earth System Data Cube. This study has been conducted within the framework of the project BACI: Toward a Biosphere Atmosphere Change Index, a project funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 640176.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R (2011) The hot summer of 2010: redrawing the temperature record map of Europe. *Science* 332(6026):220–224
- Benedek C, Shadaydeh M, Kato Z, Szirányi T, Zerubia J (2015) Multilayer Markov random field models for change detection in optical remote sensing images. *ISPRS J Photogramm Remote Sens* 107:22–37
- Cai Y (2011) Multi-variate time-series simulation. *J Time Ser Anal* 32(5):566–579
- Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. *ACM Comput Surv (CSUR)* 41(3):15
- Chen D, Chen HW (2013) Using the Köppen classification to quantify climate variation and change: an example for 1901–2010. *Environ Dev* 6:69–79
- Coles S (2001) An introduction to statistical modeling of extreme values. Springer, Berlin
- Corbella S, Stretch DD (2013) Simulating a multivariate sea storm using archimedean copulas. *Coast Eng* 76:68–78
- Coumou D, Rahmstorf S (2012) A decade of weather extremes. *Nat Clim Change* 2(7):491–496
- Dey Dipak K, Yan Jun (2016) Extreme value modeling and risk analysis: methods and applications. CRC Press, Boca Raton
- Donges JF, Donner RV, Trauth MH, Marwan N, Schellnhuber HJ, Kurths J (2011) Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. *Proc Natl Acad Sci* 108(51):20422–20427
- Donges JF, Schluessner C-F, Siegmund JF, Donner RV (2016) Event coincidence analysis for quantifying statistical interrelationships between event time series. *Eu Phys J Spec Top* 225(3):471–487
- Du Preez J, Witt SF (2003) Univariate versus multivariate time series forecasting: an application to international tourism demand. *Int J Forecast* 19(3):435–451
- Flach M, Gans F, Brenning A, Denzler J, Reichstein M, Rodner E, Bathiany S, Bodesheim P, Guaniche Y, Sippel S, Mahecha MD (2017) Multivariate anomaly detection for earth observations: a comparison of algorithms and feature extraction techniques. *Earth Syst Dyn* 8(3):677–696
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. In: IEEE transactions of pattern analysis and machine intelligence, pp 721–741
- Hotelling H (1947) Multivariate quality control. *Techniques of statistical analysis*
- Isar A S (2010) Progressive development: to mitigate the negative impact of global warming on the semi-arid regions. Springer, Berlin
- Kato Z, Zerubia J (2012) Markov Random Fields in image segmentation. Collection foundation and trends in signal processing. Now Editor. World Scientific, Singapore
- Kijazi AL, Reason CJC (2009) Analysis of the 2006 floods over northern Tanzania. *Int J Climatol* 29(7):955–970
- Ma M, Song S, Ren L, Jiang S, Song J (2013) Multivariate drought characteristics using trivariate gaussian and student t copulas. *Hydrol Process* 27(8):1175–1190
- Mahalanobis P (1936) On the generalised distance in statistics. *Proceedings National Institute of Science, India*, vol 2, pp 49–55. Retrieved from <http://ir.isical.ac.in/dspace/handle/1/1268>
- Melgani F, Serpico SB (2003) A markov random field approach to spatio-temporal contextual image classification. *IEEE Trans Geosci Remote Sens* 41(11):2478–2487
- Mikosch T (2005) Copulas: tales and facts. *Laboratory of Actuarial Mathematics*. University of Copenhagen, Copenhagen
- Nilsson E (2012) Flood impact assessment and proposals for improved flood management in Keimoes, South Africa
- Potts, R (1952) Some generalized order-disorder transformation. In: *Proceedings of the Cambridge Philosophical Society*, vol 48, p 106
- Reichstein M, Bahn M, Ciais P, Frank D, Mahecha MD, Seneviratne SI, Zscheischler J, Beer C, Buchmann N, Frank DC et al (2013) Climate extremes and the carbon cycle. *Nature* 500(7462):287
- Schwarz G et al (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
- Shadaydeh M, Zlinszky A, Manno-Kovacs A, Sziranyi T (2017) Wetland mapping by fusion of airborne laser scanning and multi-temporal multispectral satellite imagery. *Int J Remote Sens* 15:7422–7440
- Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. *publications de l'institut de statistique de l'université de paris*
- Soares C Guedes, Cunha C (2000) Bivariate autoregressive models for the time series of significant wave height and mean period. *Coast Eng* 40(4):297–311
- Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C (2006) A comparative study of energy minimization methods for Markov random fields. In: *European conference on computer vision*. volume 2. Graz, Austria, pp 16–29

- Sziranyi T, Shadaydeh M (2014) Segmentation of remote sensing images using similarity-measure-based fusion-mrf model. *IEEE Geosci Remote Sens Lett* 21:1544–1548
- Tramontana G, Jung M, Schwalm CR, Ichii K, Camps-Valls G, Ráduly B, Reichstein M, Arain MA, Cescatti A, Kiely G et al (2016) Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms. *Biogeosciences* 13:4291–4313
- Trenberth KE, Fasullo JT (2012) Climate extremes and climate change: the Russian heat wave and other climate extremes of 2010. *J Geophys Res Atmos* 117:D17
- Warren R, Smith RF, Cybenko AK (2011) Use of Mahalanobis distance for detecting outliers and outlier clusters in markedly non-normal data: a vehicular traffic example. Technical report, DTIC Document
- Zscheischler J, Mahecha MD, Von Buttler J, Harmeling S, Jung M, Rammig A, Randerson JT, Schölkopf B, Seneviratne SI, Tomelleri E et al (2014) A few extreme events dominate global interannual variability in gross primary production. *Environ Res Lett* 9(3):035001
- Zscheischler J, Reichstein M, Harmeling S, Rammig A, Tomelleri E, Mahecha MD (2014) Extreme events in gross primary production: a characterization across continents. *Biogeosciences* 11(11):2909–2924
- Zscheischler J, Westra S, Hurk Bart JJM, Seneviratne SI, Ward PJ, Pitman A, AghaKouchak A, Bresch DN, Leonard M, Wahl T et al (2018) Future climate risk from compound events. *Nat Clim Change* 19:1

Affiliations

Yanira Guancho García^{1,2} · Maha Shadaydeh¹ · Miguel Mahecha^{2,3} · Joachim Denzler^{1,2}

¹ Department of Mathematics and Computer Science, Friedrich Schiller University Jena, Jena, Germany

² Michael Stifel Center for Data-Driven and Simulation Science Jena, Jena, Germany

³ Max Planck Institute for Biogeochemistry, Jena, Germany