Biosphere anomalies detection by regression models

Yanira Guanche[1,3], Maha Shadaydeh[1], Miguel Mahecha[2,3] and Joachim Denzler[1,3]

1 Computer Vision Group
Friedrich Schiller University Jena, Germany
http://www.inf-cv.uni-jena.de

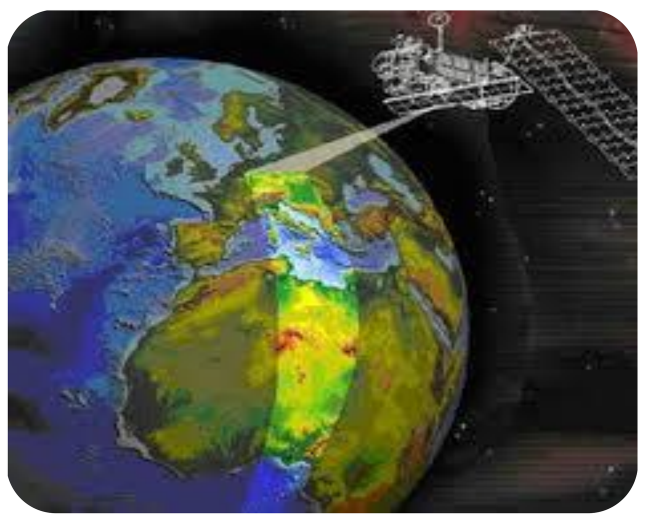
2 Max Planck Insititute for Biogeochemistry
Jena, Germany
http://www.bgc-jena.mpg.de

3 Michael Stifel Center
Jena, Germany
http://www.mscj.uni-jena.de

since 1558

Friedrich Schiller University Jena
Computer Vision Group

Motivation

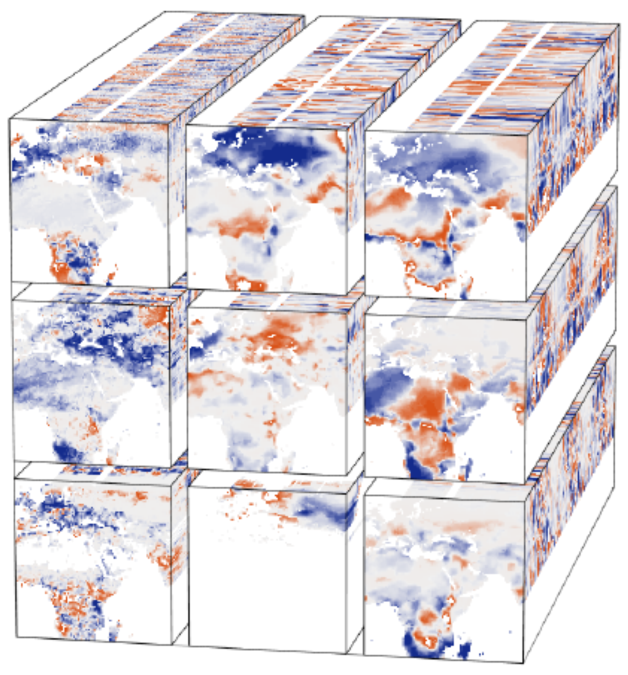


Satellites are constantly measuring environmental variables

Objective

Detect abnormal events in big amount of multivariate data and try to define an index for quantifying the Earth's health[1]

Data



CAB-LAB Earth System Data Cube [2]

0.25°global grid
8-daily data
2001-2012
Variables used:

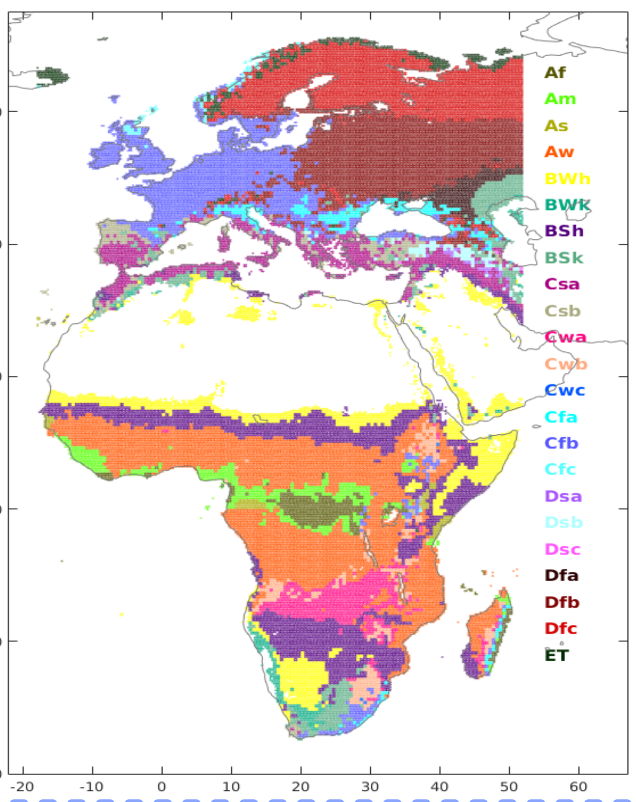
- Gross Primary Productivity (GPP)
- Latent Energy (LE)
- Net Ecosystem Exchange (NEE)
- Sensible Heat (SH)
- Terrestrial Ecosystem Respiration (TER)

Methodology and application

1. Preprocessing

1.1 Deseasonalization + Normalization

Seasonality was removed and the data were normalized (μ=0, σ=1) at each point of the grid.



1.2 Regionalization

Köppen climate classification[3] was used to cluster the grid into regions of similar conditions.

2. Feature Extraction

2.1 ARMA models

Anomalies whithin a a time series can be defined as those points that are not well represented by a previously fitted model[4].

We have used Univariate ARMA models; because of their simplicity and easier interpretation.

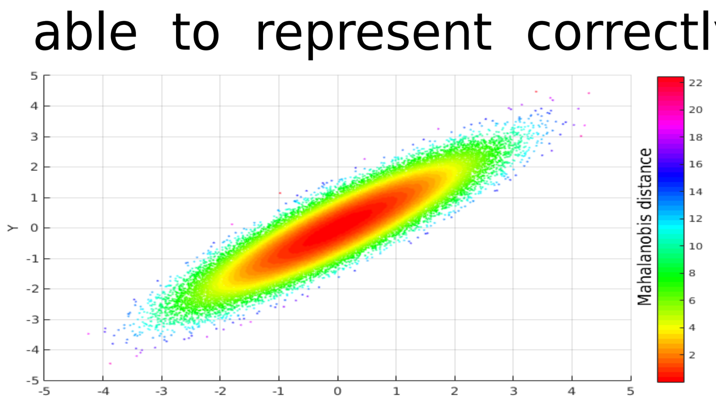
An ARMA(p,q) was fitted for each location and variable.

The values of p and q were selected for each variable and climate region with a Bayesian criteria[3]. All the combinations from (0,0) to (5,5) were tested.

2.2 Mahalanobis distance

We extract the residuals, which are the differences between the fitted models and the original time series.

The Mahalanobis distance[5] was used to measure the joint distribution of the residuals and find the extreme events. Those extreme events are observations; where the ARMA models were not able to represent correctly the time series.

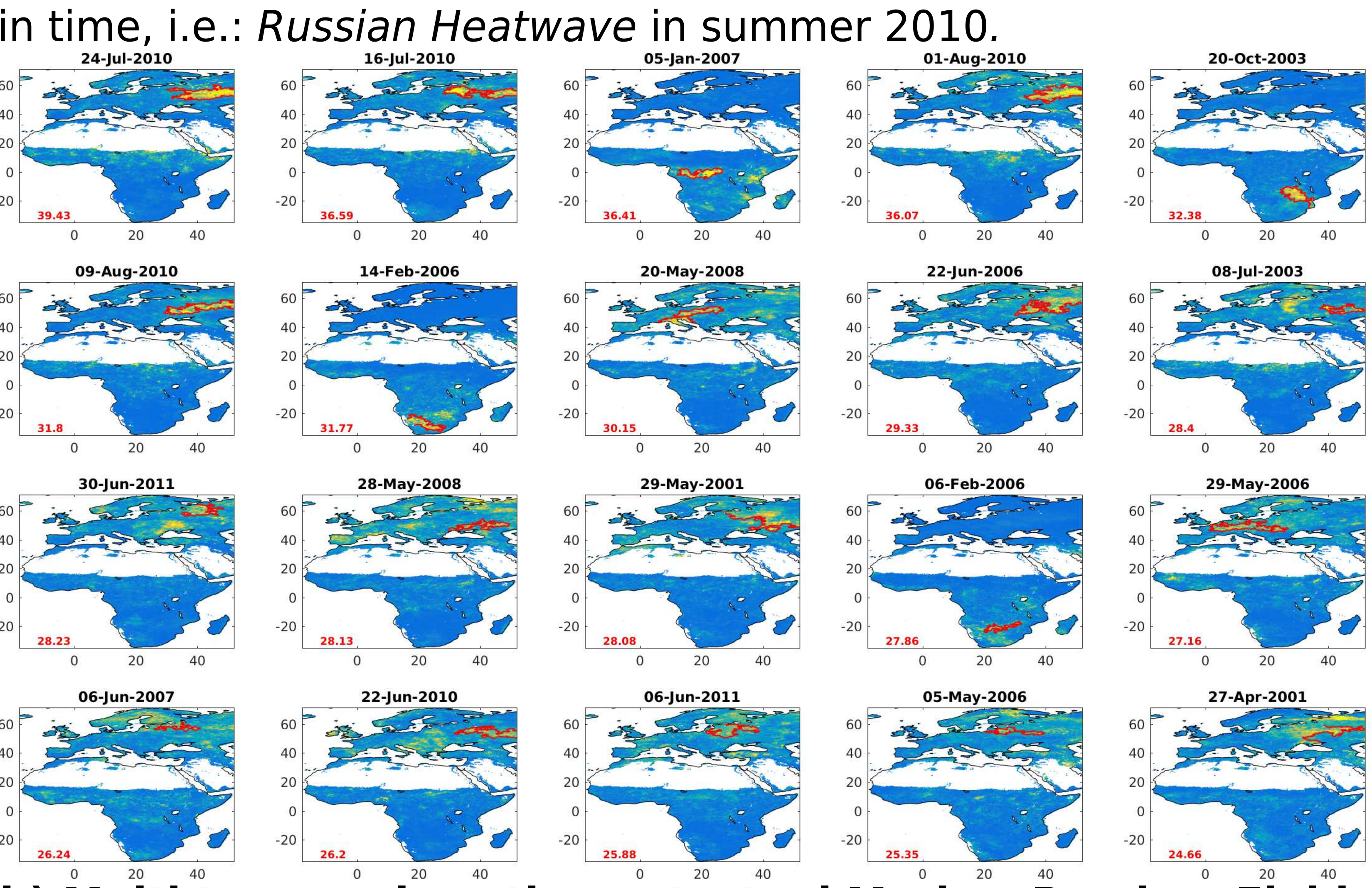


$$d_m(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^T \sum\nolimits^{-1} (\vec{x}-\vec{y})}$$

3. Event Detection

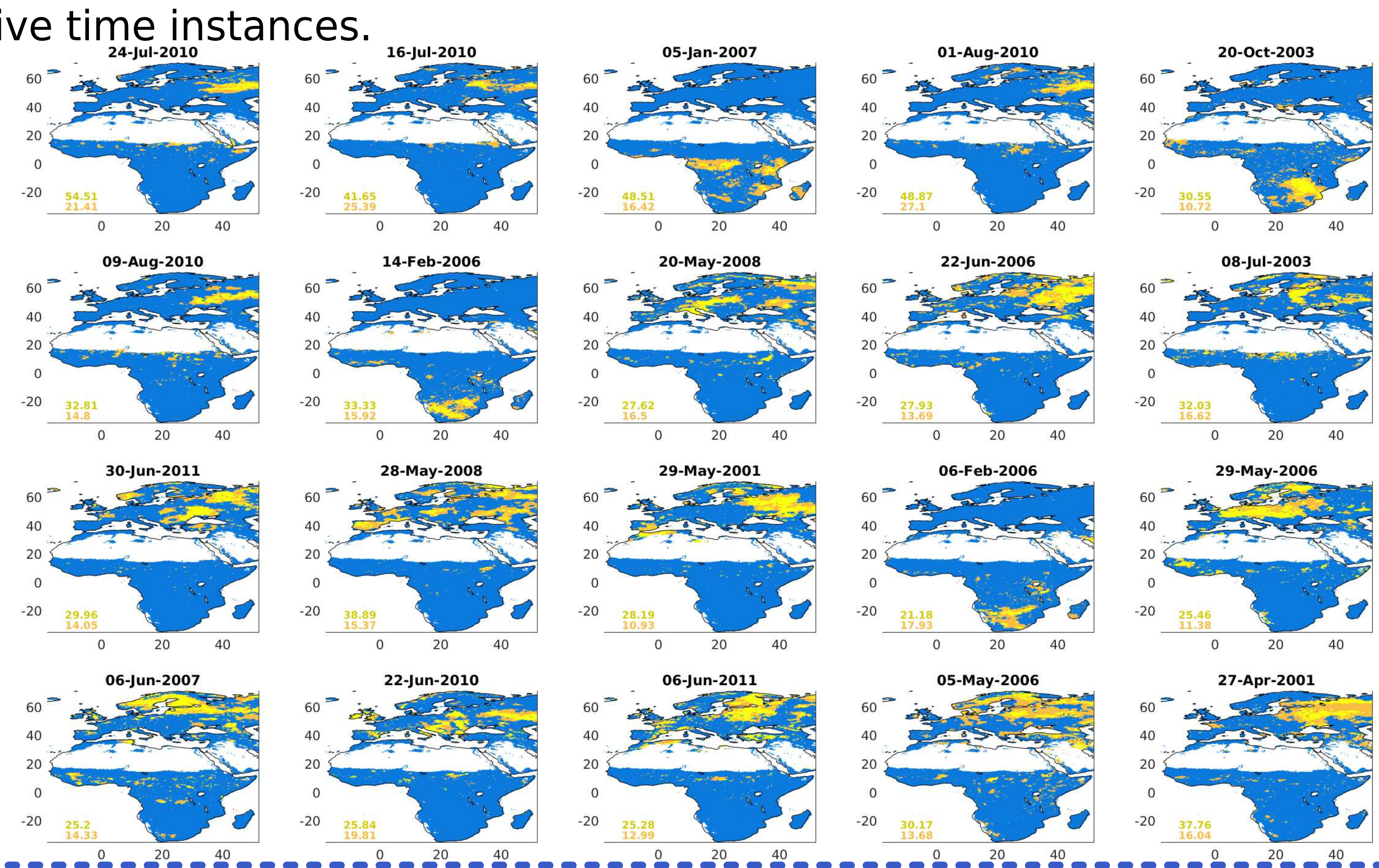
Option a) Fixed Percentile Threshold

A global threshold was fixed at the 97.5% percentile. Figure below shows the top 20 largest observations above that threshold (red line). The legend in red shows the mean intensity within the red contour. Note that each observation was assumed independent, therefore some observations might belong to the same event that extends in time, i.e.: *Russian Heatwave* in summer 2010.



Option b) Multi-temporal spatio-contextual Markov Random Field model

To tackle the spatio-temporal dependencies, the obtained Mahalanobis distance over the study area is treated as a time series of images. These images are segmented into three classes namely, ***intense anomaly***, ***possible anomaly*** and ***normal***. The figure shows the same observations as before with the mean intensity of the anomalies. The segmentation is performed using unsupervised K-means clustering followed by a multi-temporal spatio-contextual Markov Random Field model[6] applied recursively on each 3 consecutive time instances.



Acknowledgements and references

MichaelStifelCenterJena
for Data-Driven and Simulation Science

BACI

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 640176

esa

Cofunded by ESA Support to Science Element CAB-LAB

[1] baci-2020.eu
[2] earthsystemdatacube.net
[3] Chen, D. and Chen, H.W. 2013. Using the Köppen classification to quantify climate variation and change: An example for 1901-2010. Environmental Development, 6, 69-79
[4] Chandola, V. et al. 2009. Anomaly detection: A survey. ACM Computing surveys (CSUR),41, n3.
[5] Schwarz, G.E., 1978. Estimating the dimension of a model. Annals of Statistics, 6 (2): 461-464.
[6] Sziranyi, T and Shadaydeh, M. 2014. Segmentation of remote sensing images using similarity-measure-based fusion-MRF model. IEEE Geosci. Remote Sens. Lett. 11 (9).