Ecological Informatics 15 (2013) 66–73

Contents lists available at SciVerse ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Detection and attribution of large spatiotemporal extreme events in Earth observation data

Jakob Zscheischler [a,b,c,*], Miguel D. Mahecha [a], Stefan Harmeling [b], Markus Reichstein [a]

[a] Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany
[b] Max Planck Institute for Intelligent Systems, Speemanstr. 38, 72076 Tübingen, Germany
[c] Institute for Atmospheric and Climate Science, ETH-Zürich, 8092 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 27 January 2013
Received in revised form 24 March 2013
Accepted 25 March 2013
Available online 6 April 2013

Keywords:
Spatiotemporal extreme events
Climate change
Biosphere–atmosphere interaction
Scaling behavior
fAPAR

ABSTRACT

Latest climate projections suggest that both frequency and intensity of climate extremes will be substantially modified over the course of the coming decades. As a consequence, we need to understand to what extent and via which pathways climate extremes affect the state and functionality of terrestrial ecosystems and the associated biogeochemical cycles on a global scale. So far the impacts of climate extremes on the terrestrial biosphere were mainly investigated on the basis of case studies, while global assessments are widely lacking. In order to facilitate global analysis of this kind, we present a methodological framework that firstly detects spatiotemporally contiguous extremes in Earth observations, and secondly infers the likely pathway of the preceding climate anomaly. The approach does not require long time series, is computationally fast, and easily applicable to a variety of data sets with different spatial and temporal resolutions. The key element of our analysis strategy is to directly search in the relevant observations for spatiotemporally connected components exceeding a certain percentile threshold. We also put an emphasis on characterization of extreme event distribution, and scrutinize the attribution issue. We exemplify the analysis strategy by exploring the fraction of absorbed photosynthetically active radiation (fAPAR) from 1982 to 2011. Our results suggest that the hot spots of extremes in fAPAR lie in Northeastern Brazil, Southeastern Australia, Kenya and Tanzania. Moreover, we demonstrate that the size distribution of extremes follow a distinct power law. The attribution framework reveals that extremes in fAPAR are primarily driven by phases of water scarcity.

1. Introduction

Understanding the role of climate extremes is of major interest for global change assessments, because their effects on the terrestrial biosphere have the potential to substantially modify regional carbon budgets (Ciais et al., 2005; Schwalm et al., 2012). However, it has yet to be investigated whether these effects can influence the global carbon cycle climate feedback system. Hence, a global analysis of extreme events with an emphasis on data streams describing the state of the vegetation and estimating the potential impact on land–atmosphere fluxes of CO_2 is of uttermost importance (IPCC, 2012).

Many recent studies on extreme events focus on either air temperature extremes (Anderson and Kostinski, 2011; Barriopedro et al., 2011; Rahmstorf and Coumou, 2011, amongst others) and/or water scarcity (Mueller and Seneviratne, 2012; Sheffield et al., 2012). Analyses linking the effects of climate extremes to impacts on the biosphere and the associated biogeochemical cycles, instead, have primarily revolved around effects of specific climate extremes of limited geographical extent (Ciais et al., 2005; Kurz et al., 2008; Page et al., 2002; Zeng et al., 2009). In the context of disturbance analysis the direct investigation of extreme events in the terrestrial biosphere based on remotely sensed data is well established (e.g. Forkel et al., 2012; Kennedy et al., 2007; Potter et al., 2003; Sun et al., 2012). However, most of these studies focus on individual disturbance types, for example storms (Sun et al., 2012) and fires (Forkel et al., 2012), or investigate specific regions and their ecosystems (e.g. Kennedy et al., 2007). To the best of our knowledge, the question how to attribute these disturbances to either synchronous or antecedent climate extremes has been only treated marginally in the literature (Potter et al. (2003) provides some ideas in that direction).

As a consequence, it is difficult to draw a generalized and global picture on the impact of climate extremes on terrestrial ecosystems and the corresponding land–atmosphere fluxes (but see Zhao and Running (2010)). Especially the question how to detect and quantify extreme events in the terrestrial biosphere spanning over the three dimensions latitude, longitude and time on a global scale has still not been addressed systematically. Here we can learn from methodological advances in hydrology where a series of ideas have been proposed that consider both the spatial and temporal extent of large-scale droughts. For instance, Andreadis et al. (2005) use a true space–time representation of droughts, but eventually mainly

* Corresponding author. Tel.: +49 3641 576270; fax: +49 3641 577200.
E-mail address: jzsch@bgc-jena.mpg.de (J. Zscheischler).

1574-9541/$ – see front matter
http://dx.doi.org/10.1016/j.ecoinf.2013.03.004

focus on their areal extent. Corzo Perez et al. (2011) perform a "non-contiguous and contiguous drought area analysis", where they specifically aim at capturing both spatial and temporal information on droughts, but without investigating the corresponding climate trajectories corresponding to the detected events. Recently, Lloyd-Hughes (2012) presented the first analysis strategy for droughts that is truly giving full consideration to both space and time. In fact, in his approach space and time are exchangeable and a drought event is characterized by its three-dimensional extent. Thus, a further analysis of the drought's size, shape, temporal evolution and other interesting quantities can be performed.

The common notion of all the aforementioned studies is that extreme events have to be understood as spatiotemporal phenomena and detection algorithms have to be tailored to this peculiarity. This property gains additional relevance when we consider that Earth observations monitoring the terrestrial biosphere are often relatively short compared to the periods of relevance. For instance, satellite remote sensing observations are currently spanning up to maximal three decades, which prevents the application of classical extreme value theory (e.g. Coles, 2001).

In this paper we present an approach, firstly, to detect large spatiotemporal contiguous extreme events in Earth observation data, and secondly, to attribute these events to ancillary spatiotemporal climate observations. While the core ingredient is similar to the recent large-scale drought detection independently developed by Lloyd-Hughes (2012) we complement this work with a comprehensible preprocessing of the data and a new algorithm for attribution and by exploring a range of percentiles in data anomalies. Our goal is to obtain a generic method that is widely usable for a large class of variables, in particular gridded Earth observations. We illustrate the effects of the various steps of data preprocessing based on a synthetic data set. Then, we use the methodology for exploring extreme events in a composite of the fraction of absorbed photosynthetically active radiation (fAPAR, Jung et al., 2011), as an example for a remotely sensed data set. Because fAPAR can be related to the primary productivity of photosynthesis (Gobron et al., 2010), it is often used in diagnostic models that estimate the assimilation of carbon dioxide by the terrestrial vegetation. Finally, we attribute the detected extreme events in fAPAR to extreme anomalies in likely driver variables, i.e. air temperature, water availability, and fire.

2. Materials and methods

In the following, we first explain our two example data sets (Section 2.1). Secondly, we describe the different steps of data preprocessing whose effects will be analyzed later on (Section 2.2). Thirdly, we illustrate how to compute extreme events in typical three-dimensional Earth observations (Section 2.3). We then develop a method to identify relations of extreme events in one data set to anomalous conditions in ancillary observations (Section 2.4). In a very final step, we describe useful summary statistics that allow us to characterize the size distribution of extreme events over large data cubes (Section 2.5). A Matlab script for computing extreme events for a given data set and percentile is provided in the Supplementary materials.

2.1. Data

We demonstrate our extreme event detection approach by exploring two data sets. Firstly, we investigate a synthetic data set X designed to illustrate the effects of the different preprocessing steps. Its dimensionality is $90 \times 180 \times 120$, representing latitude (ϕ), longitude (λ), and time (t). In its basic form X consists of white noise (ξ). By adding a pixel-dependent positive trend (β), a global seasonal cycle (realized by a sine, one year $=$ 12 time steps) and increasing variability with increasing latitude (realized by a cosine dependent on latitude) we aim to sketch some important aspects of real world Earth observation data. These aspects will have different impacts on the results of our extreme event detection algorithm (cf. Section 3.1). In essence, we explore

$$X_{\phi,\lambda}(t) = \left(\beta_{\phi,\lambda}(t) + \sin\left[\frac{2t\pi}{12}\right] + \xi(t)\right)\left(\cos\left[\frac{\phi\pi}{180} + \pi\right] + 2\right), \quad (1)$$

where ϕ and λ run over all latitudes and longitudes, respectively. The $+2$ in the cosine secures that the variability ranges between 1 (at the equator) and 2 (at the poles). A visualization after imposing a land-sea mask is given in Fig. 1 left.

Our second data set is a global data set of fAPAR. It is a composite of three remote sensing products as described by Jung et al. (2011). In our application, we use fAPAR to detect spatiotemporal extreme events in the vegetation activity over the past 30 years.

Later on, we aim to translate extreme events detected in fAPAR into anomalies in the uptake of CO_2. For this inference step we rely on an upscaled data set of gross primary production (GPP), which has been produced by a machine learning approach that ingests the same fAPAR data stream along with FLUXNET observations and hydrometeorological records (Jung et al., 2011). For the attribution of extremes in fAPAR to meteorological conditions we use climate variables from ERA-interim (Dee et al., 2011), in particular air temperature (T) and monthly precipitation sums (P). In addition we use the water availability index (WAI), which is a more complete hydrometeorological indicator for the plant accessible water storage in the ecosystem. WAI works comparable to a single bucket model type water balance index that reflects the plant usable water column in the soil according to the principles described in Kleidon and Heimann (1998) and Teuling et al. (2006). To be able to relate extremes in fAPAR to fire we use burned area (BA) and CO_2 emissions from fires (FE, Giglio et al., 2010).

For all data sets, the spatial resolution is 0.5; the T, P, WAI, fAPAR and GPP data are available from 1982 to 2011, BA and FE only range from 1997 to 2010. In some sections we divide the globe into the six continents where we aggregated the 26 regions used in the Special Report of the IPCC (2012) as follows: North America (NA) 1–6, South America (SA) 7–10, Europe (EU) 11–13, Africa (AF) 14–17, Asia (AS) 18–23, and Oceania (OC) 24–26 (Fig. 2).

2.2. Preprocessing of the data

Given the global nature of our extreme event detection, the preprocessing is crucial for the further analysis and can shift the results significantly (see Section 3.1). In the case of Earth observations, most data sets are expected to contain a pronounced seasonality and (possibly nonlinear) trend components (Mahecha et al., 2010). Additionally, the variability of some data sets might scale with latitude. On the other hand, variables describing episodic events like precipitation have to be processed differently to guarantee the detection of extreme deviations from the normal state. In the following, we discuss the various preprocessing steps and their effects for a variety of data sets (a summary is provided in Table 1).

Subtracting the linear (or nonlinear) trend along with the mean annual cycle at each pixel allows us to compare values and extremes across time without being confounded by e.g. summer–winter differences. In order to achieve comparability across space, we normalize each pixel by its standard deviation. However, the question when a normalization of this kind is suitable depends on the application and specific research questions at hand. If one has the aim to obtain absolute extremes on a global scale whose impact can be interpreted in meaningful units, e.g. changes in carbon uptake, pixels should not be normalized. For variables that describe episodic events (for instance P, BA, FE) subtracting trend and seasonal cycle would lead to non-zero values in time steps that have been zero before. Normalizations by the standard deviation are also inappropriate because those variables are positive