Associate Editor Decision: Publish subject to minor revisions (Editor review) (14 Aug 2017) by Christopher A. Williams

Comments to the Author:

The response to reviewer comments and the revised version appear to adequately address all of the concerns raised and the paper should be acceptable for publication with final submission of the revised format.

Please note that a LaTeX-diff-file does not make much sense. The rewording and restructuring really changed too much.*Thank you very much for accepting our responses and revised paper. We have now uploaded all files for the final production as a zip file.*

Please note that a LaTeX-diff-file does not make much sense. The rewording and restructuring really changed too much.

Detecting impacts of extreme events with ecological in-situ monitoring networks

Miguel D. Mahecha[1,2,3], **Fabian Gans**[1], **Sebastian Sippel**[1,4], **Jonathan F. Donges**[5,6], **Thomas Kaminski**[7], **Stefan Metzger**[8,9], **Mirco Migliavacca**[1], **Dario Papale**[10,11], **Anja Rammig**[12], **and Jakob Zscheischler**[4]

[1]Max Planck Institute for Biogeochemistry, 07745 Jena, Germany
[2]German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany
[3]Michael Stifel Center Jena for Data-Driven and Simulation Science, 07743 Jena, Germany
[4]Institute for Atmospheric and Climate Science, ETH Zürich, Switzerland
[5]Earth System Analysis, Potsdam Institute for Climate Impact Research, Telegrafphenberg A62, 14473 Potsdam, Germany
[6]Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 114 19 Stockholm, Sweden
[7]The Inversion Lab, Tewessteg 4, 20249 Hamburg, Germany
[8]National Ecological Observatory Network, Fundamental Instrument Unit, Boulder, Colorado, USA
[9]University of Colorado, Institute for Arctic and Alpine Research, Boulder, Colorado, USA
[10]Department for Innovation in Biological, Agro-Food and Forest Systems, University of Tuscia, Viterbo, Italy
[11]Euro-Mediterranean Center on Climate Change (CMCC), 01100 Viterbo, Italy
[12]Technische Universität München, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany

Correspondence to: M. D. Mahecha (mmahecha@bgc-jena.mpg.de)

Abstract. Extreme hydrometeorological conditions typically impact ecophysiological processes ~~of terrestrial vegetation~~on land. Satellite based observations of the terrestrial biosphere provide an important reference for detecting and describing the spatiotemporal development of such events. However, in-depth investigations of ecological processes during extreme events require additional in-situ observations. The question is ~~if~~ whether the density of existing ecological in-situ networks is sufficient for analyzing the impact of extreme events, or what are expected event detection rates of ecological in-situ networks of a given size. To assess these issues, we build a baseline of extreme reductions in the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), identified by a new event detection method tailored to identify extremes of regional relevance. We then investigate the event detection success rates of hypothetical networks of varying sizes. Our results show that large extremes can be reliably detected with relatively small ~~network~~networks, but also reveal a linear decay of detection probabilities towards smaller extreme events in log-log space. For instance, networks with $\approx$100 randomly placed sites in Europe yield a ~~90~~$\geq 90\%$ chance of detecting the 8 largest (typically very large) extreme events; but only a $\geq 50\%$ chance of capturing the 39 largest events. These findings are consistent with probability-theoretic considerations, but the slopes of the decay rates deviate due to temporal autocorrelation and the exact implementation of the extreme event detection algorithm. Using the examples of AmeriFlux and NEON, we then investigate to what degree ecological in-situ networks can capture extreme events of a given size. Consistent with our theoretical considerations, we find that today's systematically designed networks (i.e. NEON) reliably detect the largest extremes, but that the extreme event detection rates are not higher than would be achieved by randomly designed networks. Spatio-temporal expansions of ecological in-situ monitoring networks should carefully consider the size distribution characteristics of extreme events if the aim is also to monitor the impacts of such events in the terrestrial biosphere.

1 Introduction

Many lines of evidence point towards an intensification of certain hydrometeorological extreme events, such as hot temperature extremes or droughts in many regions of the world over the next few decades (?). Consequently, much research focuses on understanding how extreme hydrometeorological events affect ecosystems and their functioning (overviews of the state of research and concepts are given e.g. in ????). For instance, ecosystem responses could be manifested in extreme anomalies of phenology (?), biogeochemical fluxes (?), or even in altered ecosystem structure due to induced mortality (?). Global analyses of the geographical extent and integrated anomalies of extremes in the terrestrial biosphere reveal that only a very few extremes affect large areas, whereas most events are only of very local relevance (?). Nevertheless, the integrated effects of extreme events may have global relevance. For instance, ? showed that extreme anomalies in gross primary production (GPP) to a large extent explain global inter-annual variability in gross carbon uptake.

Earth observations (EOs), especially satellite remote sensing data, encode relevant information on anomalous ecosystem functioning (??). Examples include the exploration of soil moisture anomalies in tandem with climate patterns to understand anomalous vegetation responses (?), snow cover induced albedo anomalies with consequences for local climate (?), and the impact of weather extremes on vegetation indices to track anomalies in productivity and explain vector-borne disease outbreaks (?), among many others. The consistent and contiguous spatiotemporal data coverage, and, more importantly, the fact that observations of the land surface typically integrate a plethora of processes, make EO very attractive for detecting extremes affecting the land surface.

Although EOs enable the detection of extremes in the terrestrial biosphere, a deeper understanding of impacts on ecosystem functioning can be gained from combining EOs with in-situ observations (??). In fact, ecological in-situ networks play an increasingly important role in analyzing ecological phenomena and often provide a complementary perspective on natural phenomena to EO (???) and complement model analyses (??). One prominent example is FLUXNET, with its proven record of advancing our understanding of the functioning of terrestrial ecosystems (?). FLUXNET assembles data on the turbulent land-atmosphere exchanges of CO_2, H_2O, and energy via the eddy-covariance (EC) technique (??) as they are collected in regional networks at the country or continent scale (e.g. the pan-European Network Integrated Carbon Observation System ICOS, AmeriFlux, AsiaFlux etc.). Today, many additional networks are operational or are concatenating data from past campaigns. For instance, the International Soil Moisture Network (ISMN) includes a wide range of soil-moisture observations at different depths (??); phenological observations are collected in EUROPhen (?) or Phenocam (?), and one could easily extend this list.

The site distribution in space of ecological in-situ monitoring networks is typically sparse. One obvious and common critique is that networks emerging either as voluntary associations of sites or being constructed on the basis of existing sites (naturally) cannot provide an equitable representation of the world's ecosystems (?). And in fact, geographic clustering of sites (?) as well as incoherence in their temporal continuity is problematic. However, it has also been shown that the problems of network spatiotemporal representation and the limitations of spatiotemporal extrapolations ("upscaling" *sensu* ???) are relatively minor compared to the advantages of the sheer size of the network (?).

In this paper we aim to understand the potential of ecological in-situ networks of varying size for monitoring the impact of extreme events. This paper addresses this issue in three steps: 1) We propose an approach for detecting extremes that are of regional relevance. This step is important to avoid a bias toward considering extremes that take place only in high-variance regions, and may be a relevant contribution beyond our application. 2) We explore a series of random networks of varying sizes to explore the expected detection rates. We aim to understand the observed patterns using probabilistic approaches and formulate a theoretical expectation of detection probabilities of extremes. 3) We then analyze the detection probabilities in two real networks (NEON and Ameriflux) and compare these to random networks of identical size. The paper concludes with an outlook on how our remarks could lead to improvements in network design that could be implemented to improve the detection of extreme events.

2 Data

2.1 Earth observations, EO

We required a catalogue of extreme events experienced by terrestrial ecosystems in the past several years to analyze the suitability of in-situ networks for detecting them. To create such a catalogue of extreme impacts, we used extreme negative anomalies of the Fraction of Absorbed Photosynthetically Active Radiation, FAPAR. These values are a dimensionless spatiotemporal indicator of how much solar radiation energy (in the PAR domain) is effectively absorbed by vegetation i.e. converted by photosynthesis (??).

FAPAR is considered an "Essential Climate Variable (ECV)" (?) because it supports a large variety of studies on the states and variability of the biosphere (*e.g.* ??) and plays an increasingly important role in the investigation of global biogeochemical cycles (in particular carbon and water fluxes). For instance, FAPAR can be conceptually related to GPP (typically estimated from eddy covariance (EC) tower measurements). This relationship is of the general form $\mathrm{GPP} = \varepsilon \times \mathrm{FAPAR} \times \mathrm{PAR}$, where ε is some "light use efficiency", and PAR is the "photosynthetically active radiation" (e.g. ?); one may also include other limiting factors. Consequently, FAPAR is an important basis for empirical estimates of GPP (???) and other relevant ecosystem-atmosphere fluxes *e.g.* evapotranspiration (ET; ?) or is directly used as input to diagnostic biosphere models (??). Given the tight link between FAPAR and land-surface fluxes, this variable has been used in various studies as a reference for monitoring extremes affecting terrestrial ecosystems (??).

The temporal variability of FAPAR is influenced by vegetation development, but likewise encodes e.g. fire events and other extreme reductions of FAPAR that are assumed to have a pronounced effect on GPP. Here we use FAPAR data derived by the JRC-TIP approach (TIP-FAPAR, ?). These estimates are based on the MODIS broadband visible and near-infrared surface albedo products from NASA Collection 5 at 1 km spatial resolution (MCD43B.005, ?, available on demand from co-author T. Kamins These satellite data cover the entire surface every 16 days and the data range from 2000 to 2014; in this study we use data covering Europe and the continental US (excluding Alaska). In the following we denote this data set as a 3D data cube $\mathbf{X} = \{x_{uvt} : \forall\, u \in 1,\dots,U;\ v \in 1,\dots,V;\ t \in 1,\dots,T\}$ where u is the index across the U grid longitudes, v the corresponding index on V latitudes, and t is the index on the T time steps. Each element x_{uvt} is called a voxel and is characterized by a well-defined space-time volume.

2.2 In-situ networks

First, we create artificial random in-situ networks in order to systematically study the effects of varying network sizes and as a reference for the analysis of existing networks. Then we analyze existing or recently established in-situ networks for their capability to detect the impacts of extreme events.

We use the geographical locations of eddy-covariance flux tower networks but to the actual measurements. Our main target is **FLUXNET**, a global collection of eddy covariance data collected (www.fluxdata.org; for in-depth descriptions see ??). FLUXNET is a bottom-up initiative of regional networks which decided to bring their data to a central repository. Hence, there is no systematic sampling design, resulting in unbalanced spatial coverage biased towards central Europe and the contiguous US (?). In the US, FLUXNET is mainly composed of the regional network **Ameriflux** https://ameriflux.lbl.gov/ and we use the geographical coordinates of their towers. In Europe, an overview of the most widely used EC can be found in the **European Fluxes database** http://www.europe-fluxdata.eu, which will be partly maintained in the future by **ICOS** https://www.icos-cp.eu. Here, we rely on the site distribution described in the LaThuile data set (?).

The National Ecological Observatory Network, **NEON** (http://www.neoninc.org/; ?) is an initiative to monitor ecosystems of the United States and was constructed using a systematic sampling design chosen to equitably represent the dominant ecoregions across the US. Comparable to Ameriflux, NEON sites are equipped with eddy covariance towers, but also a large suite of additional instrumentation (?), and human-based observations are recorded frequently (?). We also use the site coordinates of NEON to compare these with Ameriflux in the US.

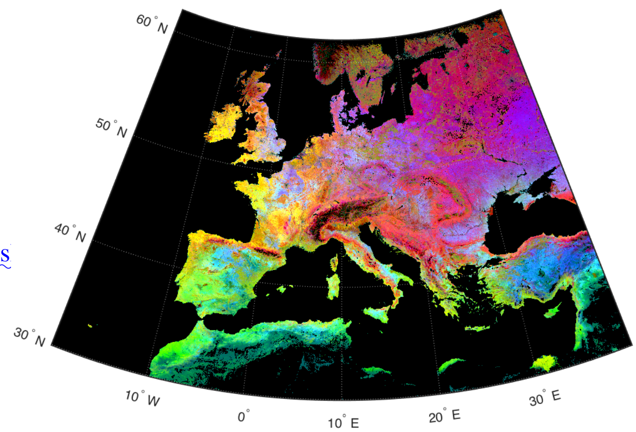

Figure 1. The top three principal components of the mean seasonal cycles of FAPAR over Europe visualized as red (R), green (G), blue (B) channels. The first component accounts for 84% of the variance. The cumulative explained variances in the first two component explain 95% of the variance, and the first three components sum up to 97%. Similar RGB colour combinations indicate comparable mean phenological patterns. These similarities are used to define overlapping regions of comparable phenology. Within each phenological region we estimate suitable and spatially varying thresholds as references for flagging potential extreme reductions in FAPAR.

3 Methods

3.1 Regional extreme event flagging

The question of how to define extreme events in spatiotemporal data cubes (see eq. 2.1) is key to the evaluation of the suitability of ecological in-situ networks. One approach would be to define some global threshold and identify values exceeding this threshold as potential extremes ("peak over threshold"). Choosing a global threshold setting is suitable when the question is about how extremes add up to global anomalies (?), i.e. when one is working with extensive data properties where the target is the integral over space and time. However, the consequence of setting a global threshold is that values that are flagged as potential extremes will occur exclusively in high variance regions, whereas low variance regions will apparently never experience extreme events. An alternative would be using only highly local thresholds (defined over time at each spatial point x_{uv}). However, the latter approach would necessarily lead to an equal spatial distribution of extreme event occurrences, which is also not desirable. We want to define extremes relative to regions that are characterized by a similar ecophysiology i.e. we want to compare each grid cell with other grid cells that have a comparable phenology and search for extremes across these geographical locations. However, as our approach should be entirely data driven, we refrain from using precomputed definitions of ecoregions.

In the following we develop a strategy to define thresholds of regional relevance. This is an attempt to find a compromise between fully local and global thresholding. Our idea builds on the concept of optical types (**?**), as they have been concretely elaborated for EOs by **?**. The key idea offered by them is that similar autocorrelation functions allows to classify ecosystems according to their temporal dynamics (see also **?**). **?** use the leading principal components of the autocorrelation estimated at each pixel across time-lags. We have developed a similar scheme to identify regions in the EOs that are of similar dynamics, but we use mean seasonal cycles instead of the autocorrelation patterns. The rationale of our choice is that want to also maintain differences in amplitude and phasing. The main steps applied for obtaining a regional threshold are the following (for a full description of the regional event detection method see Appendix A):

1. Estimate mean seasonal cycles of the datasets under scrutiny at each grid cell u,v. The mean seasonal cycles are centered around a mean of zero.

2. Reduce the temporal dimensionality of the mean seasonal cycles (MSCs) by a principal component analysis such that each principal component (PC) represents a main feature underlying the seasonal cycles. The orthogonal basis for the PCs can be approximated using a random subset of MSCs, rendering the approach very efficient in dealing with this very large data set. Figure 1 shows the first three PCs as an RGB image map for Europe. Although the nonlinearity of color perception by the human eye limits the quantitative informative value of the map, similar colors still represent regions of similar phenological dynamics in FAPAR, so one can gain an impression of environmental heterogeneity in the investigated area.

3. Identify pixels of comparable phenology by binning the scores of the MSCs on the three leading PCs as illustrated in Fig. A1 into bins of equal size. Note that the bins are very small compared to the length of the PC, guaranteeing a very fine binning.

4. Estimate a characteristic FAPAR anomaly threshold in each bin, considering all grid cell u,v belonging to this bin and grid cell u,v in the adjacent bins. Note that in the case of binning the leading 3 PCs, we have all grid cell u,v in 27 bins to estimate an FAPAR anomaly threshold as a quantile of the anomalies. Figure 2 illustrates the resulting regional threshold of FAPAR anomalies. In southern European ecosystems, smaller negative anomalies of FAPAR (i.e. higher values in Fig. 2) would be used to flag values as potential extremes. The overall geographical pattern suggests that low-variance regions (i.e. arid ecosystems) typically require smaller deviations from the expected variability to be considered abnormal situations.

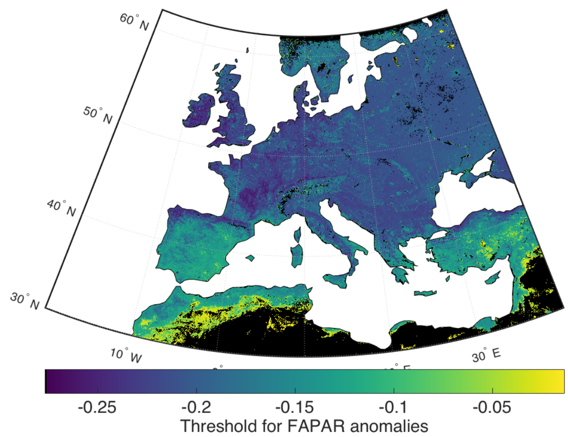

Figure 2. Map of the regionally varying FAPAR threshold used for detecting extreme events. These thresholds are derived within each subregion as defined by the leading PCs of the mean seasonal cycles. The gradient between central and southern Europe indicates that we may classify an event as extreme in one ecosystem that would be considered part of the normal variability elsewhere, i.e. arid ecosystems have lower thresholds of extremeness in FAPAR compared to humid areas.

The rationale behind this approach is primarily that similar mean seasonal cycles indicate which pixels form a "phenological cluster", requiring the application of similar quantiles. Additionally, the identification of these clusters based on the leading PCs avoids complications of an analogous analysis in geographical space where regions of similar phenology might be spatially separated by some barrier like a different land cover type, orography, or a body of water.

3.1.1 Contiguous spatiotemporal extremes

Based on the regional extreme threshold (Fig. 2) one may flag individual events as potential ("candidate") extremes. However, these initially flagged values may likewise reflect observational noise. **?** therefore proposed only considering events as extremes if larger geographical areas are synchronously affected or if the extreme persists over some temporal threshold (a very similar idea was proposed in the context of monitoring droughts by ?). This idea is realized by identifying clusters in the data cube where the spatial or temporal voxel neighbors are likewise flagged as potential ("candidate") extremes. Each of these clusters is subsequently considered a singular event; for a conceptual illustration see Fig. 3.

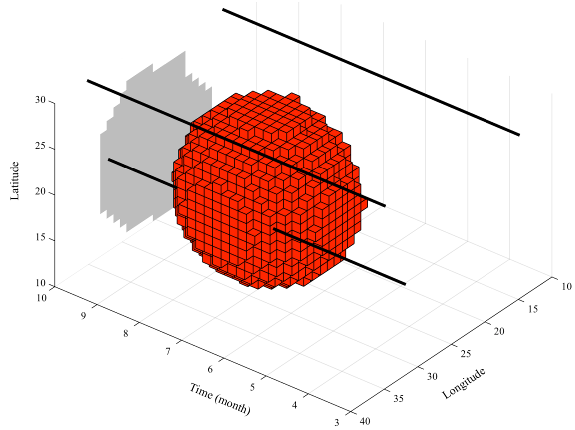

Figure 3. Conceptual visualization of the presented approach. An extreme occurs over a well-defined spatiotemporal domain (which could be asymmetric as shown here e.g. on the latitude/longitude projection). The rank of an extreme can be determined e.g. by the anomaly integrated by the red voxels, or the maximum spatial extent (gray area), or the duration along the time axis, amongst other properties. Black lines indicate the spatial position and active time of three in-situ measurement stations. In this example, only one site would have coincided with the extreme and would be considered as a potential basis for exploring the in-situ effects of the event.

A critical step of this process is defining the search space around each voxel for detecting potential neighbor extremes that should be concatenated. Throughout this paper we consider the direct neighborhood around a central voxel as follows:

- We define a spatial search space z. Two voxels x_{uvt} and $x_{u'v't}$ $(u \neq u'; v \neq v')$ are connected if $|u-u'| \leq z$ and $|v-v'| \leq z$ to obtain a spatial connectivity structure for a given t.
- We also define a temporal search horizon τ from the central voxel to compare x_{uvt} and $x_{uvt'}$ $(t \neq t')$ connecting them if $|t-t'| \leq \tau$.

Visually speaking, we search a square in space and a short line structure in time centered on a locally detected extreme event. Note that a wide range of alternative spatiotemporal connectivity structures could be used, for instance emphasizing the temporal dimension by extending the search space along the t-axis. Our choices of $z=5$ (corresponding to 25km) and $\tau=1$ (16days) are adjusted *ad-hoc* to the specific properties of the TIP-FAPAR data with its relatively high spatial resolution. By setting $z=5$ we guarantee that e.g. similar vegetation types (from which we would assume a similar responsiveness to some extreme event) could be concatenated to one extreme, even if these vegetation types are spatially fragmented due to a mosaic of land cover types. In time we search only starting from the central voxel, but given that we do this at each v,u combination, relatively complex spatiotemporal structures are allowed. Each event may consist of a set of voxels with characteristic geometric properties such as the event average or maximum duration across all affected grid cells, or the maximum areal extent. Another interesting property is the average duration of an extreme per affected grid cell. Another way of looking at these events is to integrate the variable anomaly over the voxels affected by an event, and one could also define additional metrics.

3.1.2 Specific setting for this study

In summary, in this study we used the following settings:

- Mean seasonal cycles computed over a time-span from 2001 to 2014.
- The first three PCs binned using a grain size of 4% of the range of the first PC.
- For each bin in the PC space and its surrounding 26 cells we estimate the quantile $= 0.025$. The FAPAR-anomaly values corresponding to this quantile are assigned as the threshold for the grid cells corresponding to this central bin.
- The search space for detecting extreme events is parameterized with $z=5$ and $\tau=1$ corresponding here to a search space of ± 5 km and ± 16 days.

3.2 Coinciding in-situ observations and 3D extremes

In-situ observations typically capture subgrid-level processes or footprints. For the sake of simplicity, here we assume that each point measurement is representative of one pixel x_{uv} [1 km^2] and we intersect geographical positions u and v of the in-situ data with the occurrences of 3D extremes. This approach allows us to answer the hypothetical question of whether a certain observation site would have detected an extreme in the past. An intersection considering the time domain as well would allow us to understand if an extreme had a chance of being effectively observed. Along these lines, we can also investigate whether random placement of towers would have improved or deteriorated the capability to detect extreme events.

4 Results

4.1 Random networks

To better understand expected extreme event detection rates, we initially explore random networks and their hypothetical capability to detect extreme FAPAR reductions. We focus on Europe and vary the network sizes from $n = 5, \ldots, 10000$ sites on a logarithmic scale, asking how many of the detected extremes can be identified for each size class. More precisely, we investigate the probability that an extreme event of a given size m (measured in terms of affected area) will be detected by n hypothetical towers $P(m,n)$. All following analyses are based on repeating the tower placement 100 times per size class. We mimic real site placement by assuming that a tower is not mobile, i.e. it remains active at a given location over the entire period covered by the FAPAR observations.

Figure 4 shows the average detection success rates for the random networks. The ranks r shown in Fig. 4a are derived here from the integrated spatiotemporal FAPAR anomalies (i.e. the total impact); the latter are displayed in Fig. 4b. Across network sizes we find that empirical event detection probabilities increase with event impact. These increases typically follow a straight line in the log-log plot (power-law-like behavior) for small extremes and then level off for very large event sizes. To better illustrate this pattern, we selected the network of size $n = 103$ and display it as black lines in Fig. 4. This specific network size has a $P \geq 90\%$ chance of detecting the 8 largest extreme events (according to the ranks of integrated FAPAR anomaly, see Fig. 4a). This success rate declines rapidly for smaller events, e.g. we have only a $\geq 50\%$ chance of capturing the $r = 39th$ largest event. An analogous pattern is found for the detection probabilities assessed in terms of spatial extents (Fig. 4c). In contrast, investigating the event durations (Fig. 4d) did not reveal such a clear pattern, which could be explained by the fact that we are dealing with a relatively short time series, in which only a few discrete duration classes can be recognized. The fact that global impacts of extreme events in the terrestrial biosphere behave similarly to those at smaller spatial extents is expected because these properties are known to be strongly correlated as shown e.g. in **?**. This study also reported that the duration of extreme events is less strongly correlated with their impact, as we would also suspect from Fig. 4.

A different view on this phenomenon is offered by Fig. C1 showing the detection likelihood for extremes of a given rank r across varying network sizes. Extremes of low rank (i.e. large in impact) need very small networks to be detected with rates near to 100%, whereas high rank events (of small impact) need much larger networks to reach similar detection rates. The detection probability scales linearly in log-log space with network size, indicating that one would need to inflate in-situ networks by orders of magnitude in order to detect small scale events at comparable rates to large-scale extremes.

4.1.1 Statistical considerations

The results shown in Fig. 4c are an empirical approach to describe the detection probability of extremes characterized by a given spatial extent m (measured e.g. in terms of the number of pixels or area affected during an event) using a network constructed with n randomly placed towers. In other terms, this figure reports the probability $P(m,n)$ that *at least one* tower detects the extreme and a single extreme event of spatial extent m is detected by a single randomly placed tower with probability

$$p = \frac{m}{m_{\max}}, \tag{1}$$

where $m_{\max}$ is the maximum possible extent m (in our case the maximally affected area across all time steps). However, an equivalent question is the probability that one extreme *is not detected* by any of the n towers. According to the binomial distribution, the latter probability is $(1-p)^n$, and our estimated probabilities should be described by

$$\begin{aligned} P(m,n) &= 1-(1-p)^n \\ &= 1-\left(1-\frac{m}{m_{\max}}\right)^n. \end{aligned} \tag{2}$$

This formulation helps explain the parallel decline (linear in log-log) in the detection probabilities for small extremes: We can rewrite Eq. 2 as

$$P(m,n) = 1-\exp\left(n\ln\left(1-\frac{m}{m_{\max}}\right)\right) \tag{3}$$

A Taylor expansion of Eq. 3 for a small number of towers n and small event sizes $m/m_{\max}$ (here realized by assuming that $|n\ln(1-\frac{m}{m_{\max}})| \ll 1$) yields

$$P(m,n) \approx -\ln\left(1-\frac{m}{m_{\max}}\right)n. \tag{4}$$

Further adjusting this formula for small extremes with $|\frac{m}{m_{\max}}| \ll 1$ gives

$$P(m,n) \approx \frac{m}{m_{\max}}n, \tag{5}$$

which, in a logarithmic form reads

$$\ln P(m,n) \approx \ln m + \ln n - \ln m_{max}. \tag{6}$$

We expect that this equation explains the empirically identified parallel lines of positive slope in Fig. 4 and compare our empirical findings to this theoretical expectation. Fig. 5 compares the expected and observed detection probabilities. The leveling off of event detection probabilities for large events is indeed theoretically expected,

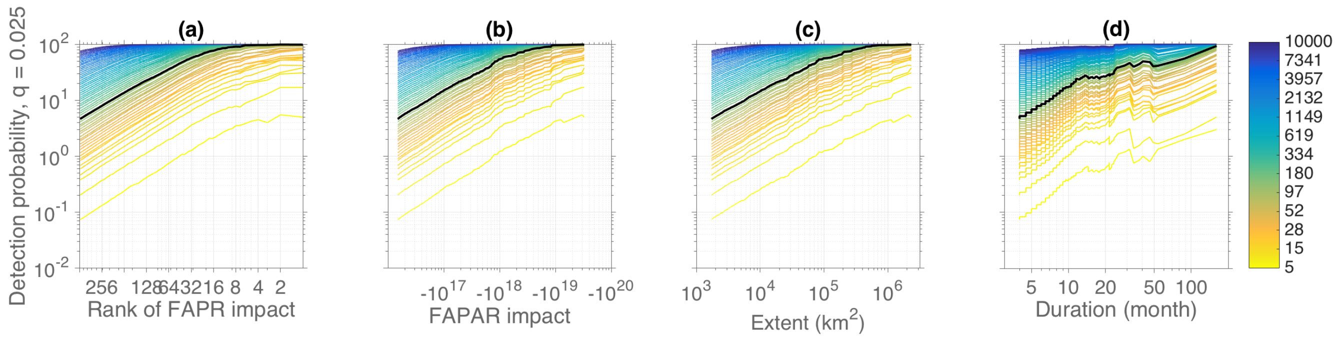

Figure 4. Comparison of average detection rates for randomly placed networks of different sizes in Europe for the period from 2000 to 2014. The color code shows the moderately exponentially increasing size of networks under consideration. Lines show the average percentage of detected events by (a) rank, (b) integrated FAPAR anomaly, (c) affected spatial area, and (d) event duration. The black line shows the case of a hypothetical network of 103 towers.

but the log-linear scaling for small events is expected to be steeper sensu Eq. 2.

In other words: the observed detection probabilities for small extremes are higher than expected, whereas detection probabilities of large extremes are lower in random networks compared to theoretical expectations. Our hypothesis is that these discrepancies are related to the spatiotemporal correlation structure of the extreme events, which is not taken into account in the above theoretical analysis.

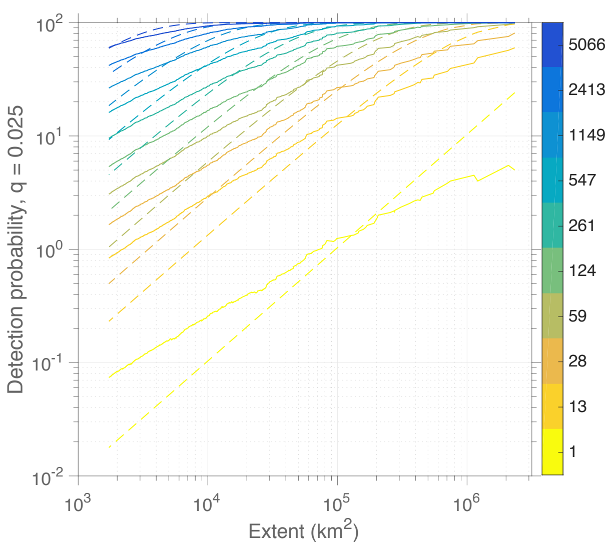

Figure 5. Comparison of the affected area of extremes (continuous lines are a subset from Fig 4c) and the theoretical expectation according to a binomial distribution and uncorrelated data (dashed lines) for varying network sizes (shown as different colours). Our empirical detection probability is lower than the the theoretical expected ones for large extremes and higher for small extremes. However, the overall pattern of the expected detection probabilities is well captured by the theoretical expectation.

In order to investigate the discrepancy revealed in Fig. 5, we performed a series of simulations using artificial data that are characterized by varying spatiotemporal correlation structures, and compared these to the expected detection rates. The results of these experiments are reported in Appendix B and let us conclude that there are very few effectively independent observations because the extremes are highly autocorrelated in time. Hence, these strong correlations lead to the fact that the largest spatio-temporal extremes tend to occur at some distance from the boundary of the domain (i.e. from the coasts). Because the networks are randomly placed, i.e. without regard to the differentiated occurrence probabilities of large vs. small extremes, this leads to the observed underestimation of detection probabilities for large extremes. A simple thought experiment can intuitively explain this effect: Imagine a landscape that consists of a contiguous, relatively large mainland (e.g. Europe) *and* a number of islands or otherwise disconnected regions (e.g. Great Britain, Ireland, Sicily) that are all far enough from the mainland that spatio-temporal extremes can by definition not be connected, i.e. exceeding the search space z. In addition, imagine that the few largest extremes that affect the mainland exceed the size of any of the islands. In this case, any tower randomly placed on an island cannot contribute to detecting large extremes, which intuitively illustrates why not taking into account the effects of autocorrelation and edge effects in our analysis results in overly optimistic theoretical predictions of detection rates based on the binomial distribution for real world landscapes. Contrarily, for medium-sized and small events, the chosen spatial search space of $z = 5$ leads to an overestimation of detection probabilities in the real data as compared to the theoretical predictions. Nonetheless, the theoretical predictions provide an exact expectation under simplified settings (i.e. no boundary effects, and an event search only in directly adjacent grid cells ($z = 1$), see Appendix B); and are thus useful for illustrating and understanding the almost linear scaling of detection rates and the size of extremes in log-log space.

4.2 Scaling issues

One doubt in applying a regional event detection approach was whether key aspects of extreme event distributions would be affected. Occurrence probabilities of extreme events in the terrestrial biosphere have often been reported to follow a power-law of the form $p(m) \propto m^{-\alpha}$ in the tails, i.e. for some values $\geq m_{\min}$ (see ??, for scaling examples in FAPAR and gross primary production respectively). Using a maximum likelihood estimator as suggested by ?? we analyze the scaling characteristics of contiguous areas affected by extreme events. We find that the event properties follow a power law (see Fig. C3). The probabilities of areas affected by extremes in both areas decline with $\alpha = 1.85 \pm 0.007$ (uncertainties given as standard errors from 1000 bootstrap samples).

Without over-interpreting these patterns (i.e. many processes could lead to the emergence of these power-laws, some of which are discussed in ?) we consider that this property could be exploited to inform network design issues: According to ?, and others there are a few considerations pointing in this direction: the expectation value $E[m(r)]$ of an extreme event of rank r (in this formulation, the largest event has rank 1 as in Fig. 4a) has the form

$$E[m(r)] = cr^{-\frac{1}{\alpha-1}}. \tag{7}$$

where α is the scaling exponent, and c is some normalization constant - both can be obtained from a fit to the empirically obtained rank function $m(r)$. Applying Eq. 7 would allow us to study the network detection probability as a function of rank (see Figs. 4a and C1) and we can insert the expressions into Eq. 3:

$$P(m,n) = 1 - \left(1 - \frac{m(r)}{m_{max}}\right)^n \tag{8}$$

$$= 1 - \left(1 - \frac{cr^{-\frac{1}{\alpha-1}}}{m_{max}}\right)^n \tag{9}$$

Furthermore, using the approximated log-log form of the network detection probability (Eq. 7) yields

$$\ln P(m,n) \approx -\frac{1}{\alpha-1}\ln r + 1\ln n + \ln c - \ln m_{max}. \tag{10}$$

This equation may explain the parallel lines for ranks r corresponding to small extreme event extents $m(r)$ (see e.g. Fig.C1). More importantly, it relates the scaling exponent to the expected detection probabilities. In other words: gaining insights about the scaling behaviour of the extremes can be used to formulate clear expectations about event detection probabilities of a given rank and size.

4.3 Comparing AmeriFlux and NEON

Our results so far show that random networks may differ somewhat from our expected detection rates for various reasons. But the overarching hypothesis is that even relatively small networks may have a good chance of detecting large scale extreme events. We therefore consider the configuration of real eddy covariance networks. We now focus on the US (continental areas only) instead of Europe. We have two networks with very different histories and therefore configuration: Ameriflux and NEON, and we consider them both together. Again, we compare our results to random networks of equal size.

The starting point for our considerations was whether ecological in-situ networks have effectively been able to detect the most relevant extreme events experienced by land ecosystems due to their network construction, or if these were lucky circumstances. We therefore ranked the 100 largest events detectable in the continental US by their integrated FAPAR anomalies. We then counted the number of events that could have been detected by at least one of the Ameriflux or NEON towers, or, by taking both together (if all towers would have been active over the entire monitoring period). Fig. 6 shows the number of detected events for these three network configurations of NEON, AmeriFlux, and both together, as a function of their rank.

Due to its large network size, AmeriFlux detects many more extremes than NEON (128 vs. 39 sites in the contiguous US, excluding Alaska and islands). Concatenating both networks helps increase the detection rates for small events. Our next question was whether these detection rates are comparable to random networks of the same size. For the case of NEON we find that the median detection rate of randomly designed networks is slightly higher compared to the real network - which still remains above the 2.5%ile. At first glance this is an unexpected finding: we would expect that undesired vicinity may occur by chance in a random network, increasing redundancy among towers in space compared to the very systematic sampling design of NEON (?). We conclude here that while the design efforts used in establishing NEON may pay off for certain studies, they are not an effective means to maximize the detection of extremes. This observation again reflects the lack of spatial regularity in the occurrence of extremes.

The equivalent experiment conducted on the AmeriFlux network yields much higher detection rates for the random networks compared to the established network (Fig. 6). We attribute this difference to one particular characteristic of AmeriFlux: many of the sites in this network are co-located on purpose (e.g. to explore spatial heterogeneity or to monitor different disturbance regimes in adjacent and hence climatologically similar ecosystems). Fig. 6 shows that AmeriFlux sites have a relatively high degree of spatial clustering. If the target were to analyze continental extreme events and guarantee monitoring the largest events, the

AmeriFlux configuration would be suboptimal. In other words: the spatial autocorrelation in an ecological in-situ network that was not systematically designed can be outperformed by a random (and hence spatially independent) network.

Another aspect to investigate in this context is concatenating NEON and AmeriFlux (both data sets are intended to be freely available to the research community, Fig. 6 dashed line). Our results show that this approach would marginally increase the detection capacity. One reason for this marginal improvement is again that AmeriFlux and NEON sites are partly geographically co-located and that AmeriFlux—despite of being a bottom-up activity—already has a significant spread across the country that is competitive with a novel network designed for the purpose of capturing large scale extremes.

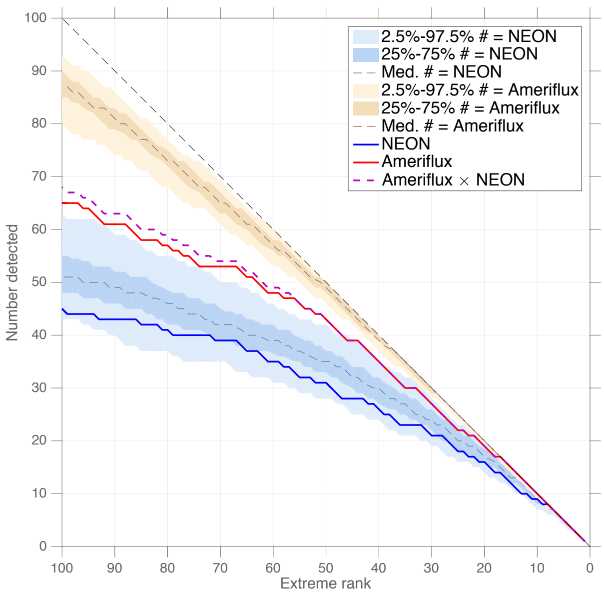

Figure 6. Comparison of the potential of NEON (39 terrestrial sites) and AmeriFlux (128 sites) for detecting extremes defined by varying thresholds in the contiguous continental US (excluding Alaska and islands). The purple dashed line shows a merged AmeriFlux–NEON network. Dashed lines enveloped by a 95 percentile range are detection rates of random networks. The sizes of the random networks correspond to NEON (blue) and AmeriFlux (brown) and summarize 100 repetitions. We also show the 1:1 line, which would correspond to perfect detection performance and is the theoretical limit.

5 Discussion

5.1 Regionalized event detection

Reliable event detection algorithms are a prerequisite to addressing the question of how effective in-situ networks are for detecting extreme events of a given geographical extent. Our aim here is to classify events as “extreme” if they exceed an anomaly value that is unusual across regions that follow the same main phenological pattern. This contribution could be relevant to other studies beyond the present application. This method has advantages over using a global threshold, which fundamentally changes the obtained picture and leads to a few hotspots of extremes in regions where the data have high variability (for the case of GPP see **?**). The effect of building on regional thresholds to delineate which anomalies should be considered “extreme” (recall Fig. 2) is that we find only very moderate geographical clustering of event occurrences (not shown). From our viewpoint, this is very logical, as there is no reason why relative extremes should preferentially happen in certain regions. Methods of this kind are particularly relevant in times of increasing availability of EOs to detect impacts rather than referring to anomalous observations in the meteorological records, which may or not affect terrestrial ecosystems. In fact, all of the largest extreme events that have had severe impacts on agriculture and human well-being and attracted the attention of the media are well detectable with our approach. Prominent examples are e.g. the 2003 European heat wave (e.g. **?**), the 2010 Russian heat wave (e.g. **?**), or the 2012 US drought (e.g. **?**), which are all easily detectable both from climate records and remote sensing data. However, the smaller the spatial extents become, the more relevant a remote sensing based regional assessment will be. We also expect that a regionalization of this kind could be useful when using more advanced multivariate event detection algorithms (see e.g. **?**) that can tap into the full potential of many EOs.

Regarding the details of the chosen methodological approach, one may question why we propose simply binning the leading PCs derived from the MSC of our EO. This approach was mainly developed to effectively deal with the very high resolution of the underlying data, seeking a very efficient subgridding approach. One alternative would have been to e.g. cluster the PCs directly. However, besides the computational costs, conventional clustering methods lead to a non-uniform partitioning of the space spanned by PCs. This non-uniform partitioning makes it slightly more complicated to identify neighbouring clusters, which is necessary to stabilize the quantile-based computation of anomaly thresholds. Having an equal meshgrid over the PCs that we can also compute on a subset of MSCs renders the approach very efficient for very large data sets and is completely data adaptive. It was very important for this exercise to have many small classes, in order to compute a very well regionalized anomaly threshold (shown in Fig. 2), which would not have been achievable using classical climate classifications of ecoregions. A more detailed follow-up study should explore the question of how the choice of the various parameters affects the event detection accuracies. A crucial question in this context will be whether one can tune these parameters such that a baseline of events is well detected.

A further argument in favor of our approach was that we rely on a limited number of events detected in a finite time horizon of available satellite data. Monitoring 15 years of extreme events probably does not allow us to conclude anything about the future occurrences of extreme events. In this sense, this study can only be read as a call for (re)considering the density of ecological networks in network design studies. An alternative would be to also consider climate projections and put more emphasis on more "vulnerable" ecoregions. Non-stationary climate and environmental conditions notwithstanding, we have to acknowledge that extremes are too rare to derive a spatial occurrence probability using data from the satellite era only.

5.2 Relevance for network design

To the best of our knowledge, there are only a few realized examples of systematically designed in-situ ecological networks. One of the best examples is NEON, which is therefore particularly interesting in the context of this study. The underlying design principle is to cluster environmental conditions and states, including e.g. precipitation, radiation, topography, and water table depth, among others (**?**). These delineated ecoregions are taken to be representative of approximately homogeneous areas in the mean land-climate system state, and yield an equitable representation of land surface processes in upscaling activities (e.g. the spatiotemporal inter- and extrapolation of land-atmosphere fluxes of CO_2, H_2O, and others **???**) or model-data integration studies (sensu **?**).

Our finding that concatenating NEON and AmeriFlux would have yielded only a minimal increase in detection capacities for extreme events can be understood as a call to avoid co-locating towers in relatively close vicinities - at least when the objective of detecting extreme events is highly relevant. In fact, when the objective is to monitor and understand the impacts of climate extremes on ecosystems, we show here that probability theoretical expectations should be be taken into account but would need to be extended to consider temporal autocorrelation as well as the event detection approaches chosen. In our case, the latter had a relatively large footprint ($z = 5$) in order to not miss events that may appear fragmented due to e.g. heterogeneous landscape characteristics. Clearly, one would need to determine such parametric choices depending on the type of extreme events and underlying question.

Nevertheless, we think that the remarks presented here could become useful elements for quantitative network design studies. In our area, earlier considerations in this direction have put their emphasis on reducing the uncertainties for upscaling fluxes from the site level to continental or global flux fields (**?**). Focussing on this first-order question is of course essential, before focussing on detecting rare anomalies. This is also reflected in the alternative methodological avenues that were used for addressing the network design problem. For instance, carbon cycle data assimilation systems (CCDAS; **?**) were very useful for quantitative network design (QND; see, e.g. **??**) i.e. to evaluate real or hypothetical candidate networks in terms of their ability to constrain target quantities of interest. The QND approach within a CCDAS allows the combination of terrestrial, atmospheric and ultimately also oceanic data streams. A key finding so far was that eddy covariance networks with one site per ecosystem type achieve excellent performance. QND studies have also been performed for EO data streams such as column integrated atmospheric CO_2 (**??**). But again, none of these studies so far have attempted to unravel the impacts of extreme events on the terrestrial biosphere, which might be a relevant pursuit for subsequent studies.

Overall, this study can be also seen as a prototype. In appendix B we show that analogous studies can be effectively implemented. There we use the International Soil Moisture Network ISMN and detect EO anomalies using a drought indicator. This very brief analysis stresses one additional aspect that we have effectively ignored through the main paper: the importance of keeping network measurements alive over time. Many of the sites have only been active for short monitoring periods, leading to substantial losses in event detection rates. It is the continuously sustained measurement networks that will substantially improve event detection rates in the long-term.

6 Conclusions

This study tries to understand to what degree ecological in-situ networks such as AmeriFlux or NEON can capture extreme events of a given size that ~~hit~~ affect land ecosystems. We find, for instance, that the ~~largest~~ 10 ~~extremes having~~ largest that have occurred in the US between 2000 and 2014 would ~~have all~~ all have been identified with the current networks, offering a good perspective for in-depth site level analyses of these phenomena. Concretely, this finding means that there is a high chance of capturing ~~the~~ major extreme events – beyond the very few ~~prominent~~ (2-3) prominent events that may receive major media coverage such as the 2003 heatwave in Europe or the 2012 US drought. In general, we find that "large" extreme events could have been detected in a very reliable way, ~~while~~ whereas there was a linear decay of detection probabilities for smaller extreme events in log-log space. We can explain this general behavior with ~~straighforward~~ straightforward considerations in probability theory, but the slopes of the decay rates deviate: While we find lower detection rates for the very large extremes, the opposite is the case for very small extremes. Experiments with artificial networks reveal that these deviations stem both from ~~the temporal autocorrelation~~ autocorrelation issues and the exact implementation of the detection algorithm.

Our original motivation for pursuing this study ~~is the question if~~ was the question of whether one could optimize the design of ecological in-situ networks for maximizing the detection rates of extreme events. And indeed, we find some general rules, i.e. when the goal is detecting very large events (i.e. low rank events), network sizes can differ by up to two orders of magnitude but still yield nearly comparable detection rates. Only if the goal was to reliably enhance the detection probabilities of small-scale events ~~,~~ would a disproportionate "investment" in large networks ~~would~~ be required, ~~but~~ which would then also become orders of magnitude more efficient compared to the small networks.

However, any inference on the future spatial occurrence probability of extremes is not tenable based on data from a decade of observation. ~~But it~~ It is not only data paucity that limits our insights here: quantitative network design is per se non-trivial in a changing world. We find, however, that certain general patterns could be taken into consideration, for instance the fact that event occurrence probabilities are clearly inversely related to detection probabilities on a very well defined and ~~a robust scale. Also~~ robust scale, and that the power-law distribution of extreme event size seems to have practical relevance for network design purposes.

Author contributions. The first three authors equally contributed to analyses presented in this study, J.F.D. helped in deriving the probability-theoretic explanations for the identified patterns, all authors provided substantial input to the design of the study and discussion of the results.

Acknowledgements. This study was supported by the European Space Agency with the Support to Science Element STSE "Coupled Atmosphere Biosphere virtual LABoratory project CAB-LAB" see http://earthsystemdatacube.org/ and we thank the EU for supporting the BACI project funded by the Horizon 2020 Research and Innovation Programme under grant agreement 640176; D.P. further thanks the EU for the ENVRIplus project funded in the same programme under grant agreement 654182. The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by Battelle Ecology, Inc. This material is based upon work supported by the National Science Foundation under the grant DBI-0752017. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. J.F.D. thanks the Stordalen Foundation (via the Planetary Boundary Research Network PB.net) and the Earth League's EarthDoc program for financial support. The reviews by Dr. Bohrer, Dr. Brunsell, and Anonymous as well as many suggestions by Dr. Durso have greatly improved the quality of the manuscript.

Appendix A: Regional event detection

In the following we develop a strategy for defining thresholds of regional relevance that are computationally suitable for dealing with high-resolution remote sensing data like the 1 km FAPAR data considered here. Our aim is to find regions of comparable phenology. Our assumption is that the expected seasonal cycle in FAPAR is a good representation of overall phenology, and hence ecosystem type.

The first step considers the data set of mean seasonal FAPAR patterns $\mathbf{F} = \{f_{n,s)} : \forall n \in 1,\ldots,N; s \in 1,\ldots S\}$, where each point n is pointing to a geographical location u,v and contains the local mean of seasonal observations s.

In the second step, we use principal component analysis (PCA) to reduce this S-dimensional data set. In other words, we seek orthogonal components that represent the main gradient along the covariances of the seasonal cycles. More formally, the covariances of these centered mean seasonal cycles are given as

$$\mathbf{C} = \mathbf{F}^t\mathbf{F}. \tag{A1}$$

Common patterns of seasonality are identified by first estimating the k leading eigenvectors,

$$\mathbf{C}E_k = \lambda_k E_k \tag{A2}$$

where E_k the kth eigenvector of length S, and λ_k the corresponding eigenvalue. The scores of the kth principal component are given by

$$A_k = \mathbf{F}E_k. \tag{A3}$$

and k leading A_k can be interpreted as a proxy for the characteristic patterns underlying the mean seasonal cycles across space. Figure 1 visualizes the three leading principal components as an RGB-color composite, revealing a distinct map of European phenological regions.

Third, the question is how to identify regions of similar phenology in this continuous space spanned by the principal components. One could use, for instance, some clustering algorithm. However, given the high density of spatial points and the continuous sampling, an equivalent approach is to choose an equidistant grid in the space of the principal components. We choose a very dense grid, such that each cell is as wide as 4% of the range of the first PC. We then define an FAPAR anomaly threshold as a predefined quantile based on the distribution of FAPAR values separately for each grid cell and its 26 neighbours in the space of the leading 3 PCs. This threshold is assigned to all points in the respective grid cellrepresented herein. This threshold is assigned to the all points represented therein. Figure A1 illustrates this approach in detail.

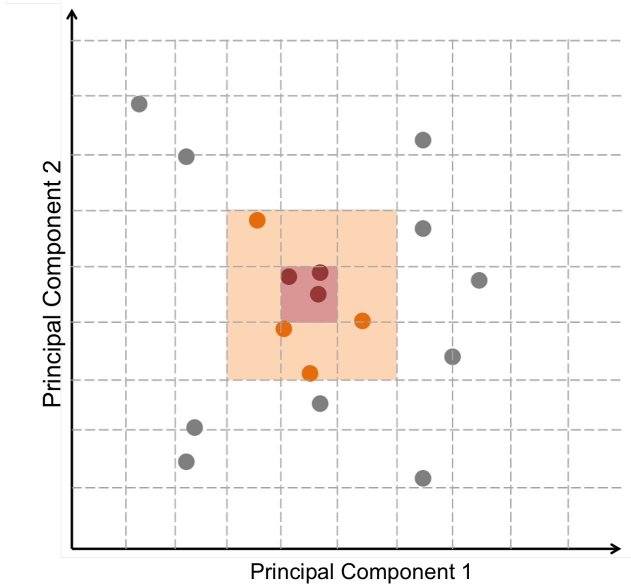

Figure A1. Illustration of identification of regions with similar threshold: We define a grid in the space of the leading PCs (geographically shown in Fig. 1), where each mesh width corresponds to 4% of the total min-max range of the first PC. We assign percentile thresholds as calculated over a $3 \times 3 \times 3$ set of mesh elements (shown in orange) and assign these percentiles to the central dots (shown in red). For the sake of clarity, we illustrate the approach only in the space of the leading two PCs.

We have now proposed a FAPAR threshold for each point and can map this threshold back to the geographical space by remapping each point to the known geographical coordinates u, v. This is shown in Fig. 2.

Appendix B: Spatiotemporal correlations

Fig. 5 reveals a strong discrepancy between theoretical and observed detection probability. Here we investigate this discrepancy further. We generated Gaussian data but introduced varying spatiotemporal correlation structures of different degrees. We followed the approach suggested by **??** to simulate data with a power law power spectrum of some prescribed exponential spectral decay. The method combines an approach for generating spatial fields of a desired correlation structure that likewise have a similar temporal correlation. The idea is that the Fourier coefficients of some artificial data (white noise) are forced to decay as a power law function across frequencies i.e. proportionally to $f^{-\beta}$. An inverse transformation to space yields a correlated data field. If we choose $\beta = 0$, it corresponds to uncorrelated, $\beta = -\frac{3}{5}$ to moderately correlated, and $\beta = -\frac{8}{5}$ to highly correlated data. These artificial datasets are visualized in Fig. B1g-i. We used a simplified event search radius ($z = 1$, $\tau = 1$) and investigate two cases:

1. Ignoring the time domain: In this case, the empirically identified detection rates correspond exactly to the theoretical detection probabilities. This finding reveals that the spatial correlation structure does not explain a deviation from the theoretically expected pattern (compare appendix Figs. B1a—c). This is explained by the fact that, although patterns of extreme anomalies might be correlated in space, the tower placement is still random and for sufficiently sparse networks and relatively contiguous landscapes (i.e. only small edges, no islands, etc.) it has no effect.

2. Considering spatial and temporal correlations: In this case we find a tendency towards lower detection probabilities. This effect becomes more pronounced with larger extremes and spatiotempoal autocorrelation (see appendix Fig. B1d-f) due to a stronger tendency for large spatio-temporal extremes to occur away from the domain's boundaries, thus any tower that is randomly placed close to a boundary would have a disproportionately low chance of detecting large extremes.

However, the approximation of the expected probabilities for the small events is still inconsistent with our empirical finding (recall Fig. 5). Hence, we repeat the artificial experiment using the exact algorithmic settings applied to the FAPAR data: we allow for a tolerance radius ($z \gg 1$, $\tau = 1$) to identify each extreme by a given tower. Again we distinguish the two cases:

1. Ignoring the time domain: Using a large search radius for detecting extremes (which is clearly necessary in real and e.g fragmented landscapes) leads to increased event detection rates. This effect can lead to higher detection rates that exceed the simple statistical expectations as derived from the binomial distribution by several orders of magnitude in the case of small extremes (see appendix Figs. B2a—c).

2. Considering the full spatiotemporal case reduces the discrepancy slightly (i.e. for large events that would be detected anyway), but still results in an overestimation (see appendix Fig. B2d-f). For very large events, the lines may even cross in the case of strongly autocorrelated data.

These numerical experiments highlight some of the issues that need to be considered in evaluating real networks or quantitative network-design: the phenomena we aim to monitor are highly autocorrelated in time, which leads to considerable edge effects for large events. Therefore, theoretically expected detection rates estimated from the binomial distribution are overly optimistic for large events - unless the effects of autocorrelation and edge effects as a consequence for large events are analytically taken into account.

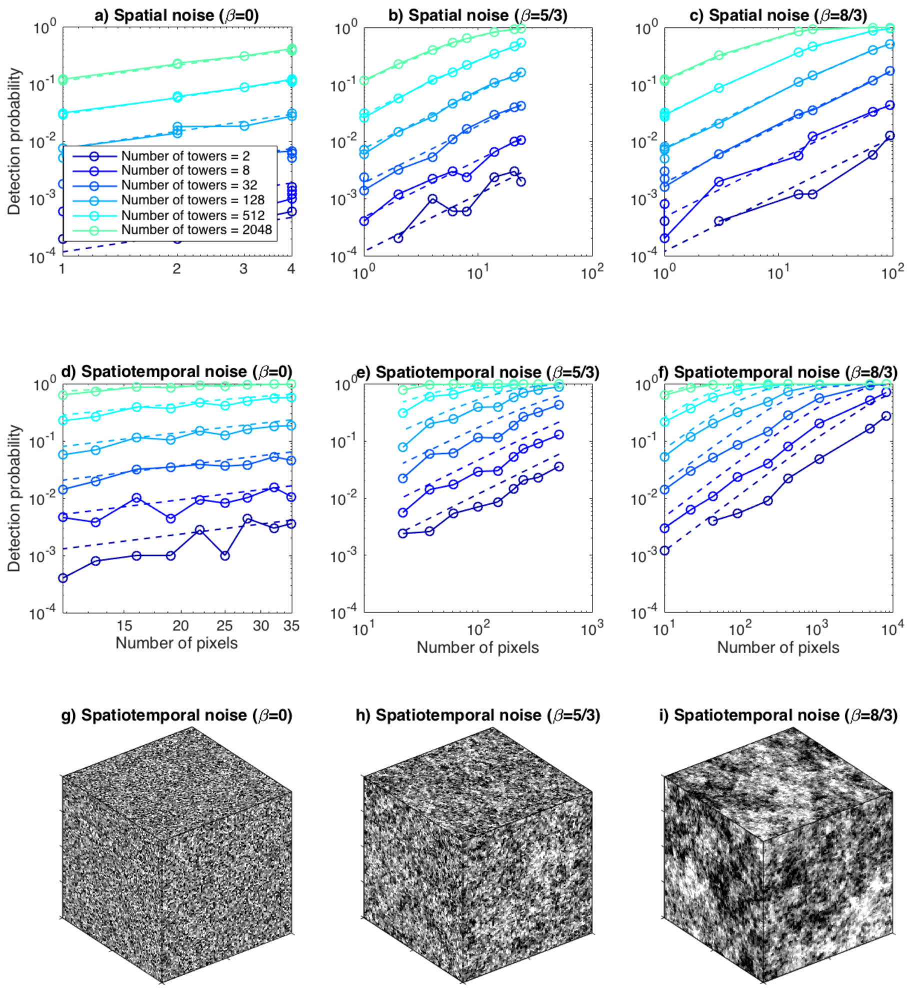

Figure B1. Artificial data example. a) Detection probabilities when ignoring the time domain for varying network sizes. In this case, the empirically identified detection rates correspond exactly to the theoretical detection probabilities. If we induce moderate spatiotemporal correlations in b), and stronger ones in c) we still find an excellent fit to the theoretical expectation because we still have relatively sparse networks and the towers are independent samples of the underlying distribution. If the detection rates over space and time are considered, however, the events are no longer independent due to their temporal autocorrelation, and thus the largest extremes tend to cluster towards the center of the domain. Parts e) and f) show these lower detection rates, and g), f), i) are the data corresponding to results in the columns.

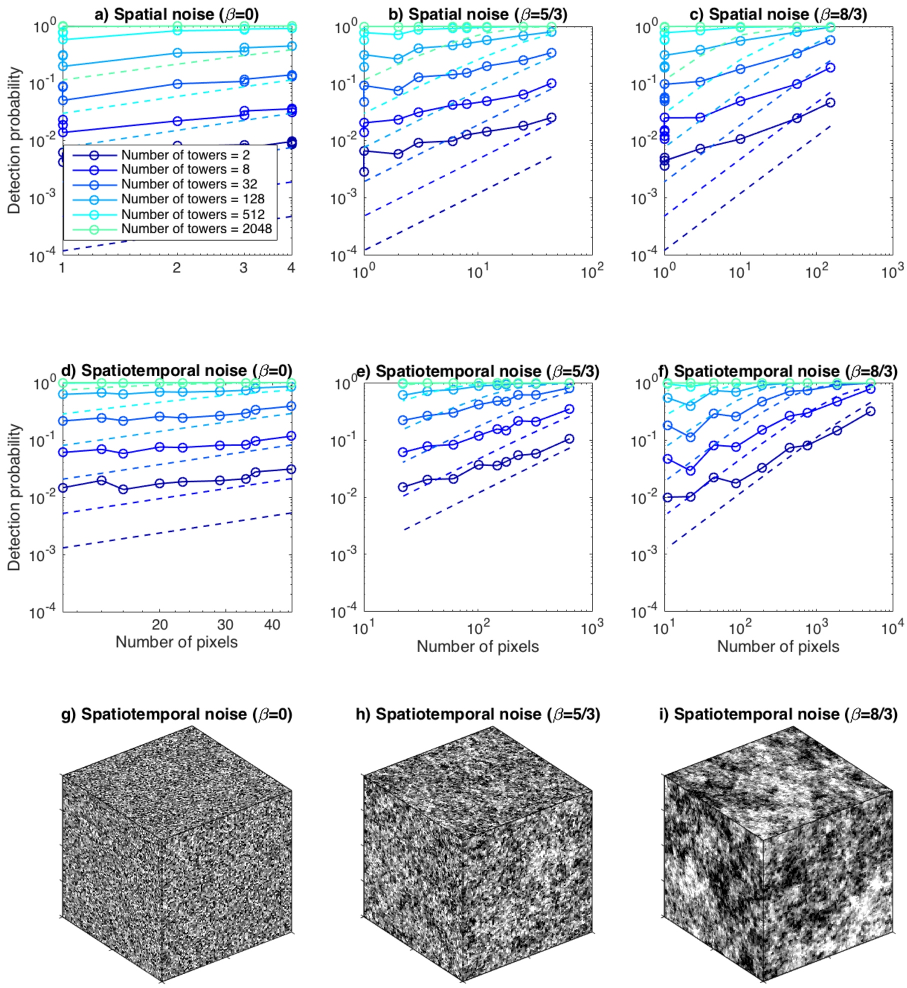

Figure B2. Artificial data example considering the actual event detection algorithm. a) Detection probabilities when ignoring the time domain for varying network sizes. In this case, the empirically identified detection rates dramatically overestimate the theoretical detection probabilities. If we induce moderate spatiotemporal correlations in b), and stronger ones in c) we still find this pattern, but it is less pronounced for the very large events. This shows that having a large footprint for the event detection algorithm leads to an overestimation of the detection rates of small extremes. If the detection rates over space and time are considered, however, the events are no longer independent due to their temporal autocorrelation. Parts e) and f) reveal lower deviations from the expected detection rates, which is a compensating effect of the autocorrelation and event detection method setting. The data corresponding to results in the columns are shown in g), f), and i).

Appendix C: Supplementary figures

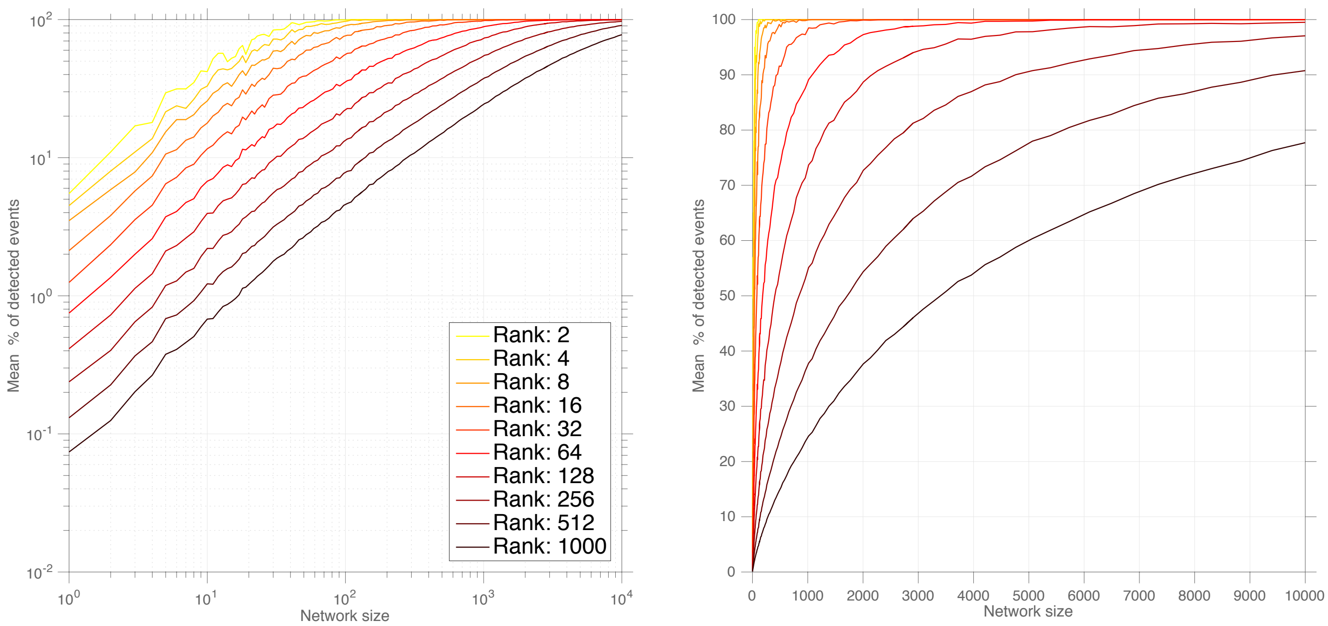

Figure C1. Average detection rates of extremes of given ranks (each line represents the rank of an extreme event) across varying network sizes in logarithmic representation (left panel) and linear representation (right panel). Small ranks indicate large impact extremes that typically also affect large areas (see Fig. 4). The figure shows that detection rates scale with smaller network sizes and then tend to saturate i.e. we find a convergence towards full detection rates.

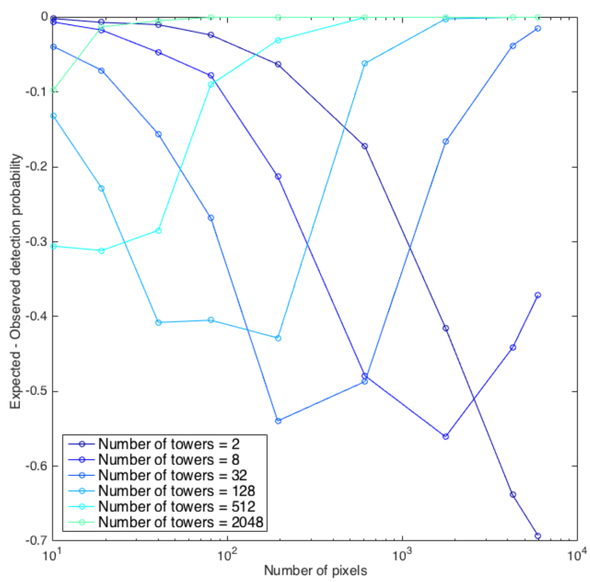

Figure C2. Comparison of the affected area of extremes (Fig 4c) and the theoretical expectation according to a binomial distribution. Our empirical detection probability is lower for the very large extremes, and higher for the small extremes. The problem is more pronounced for small network sizes.

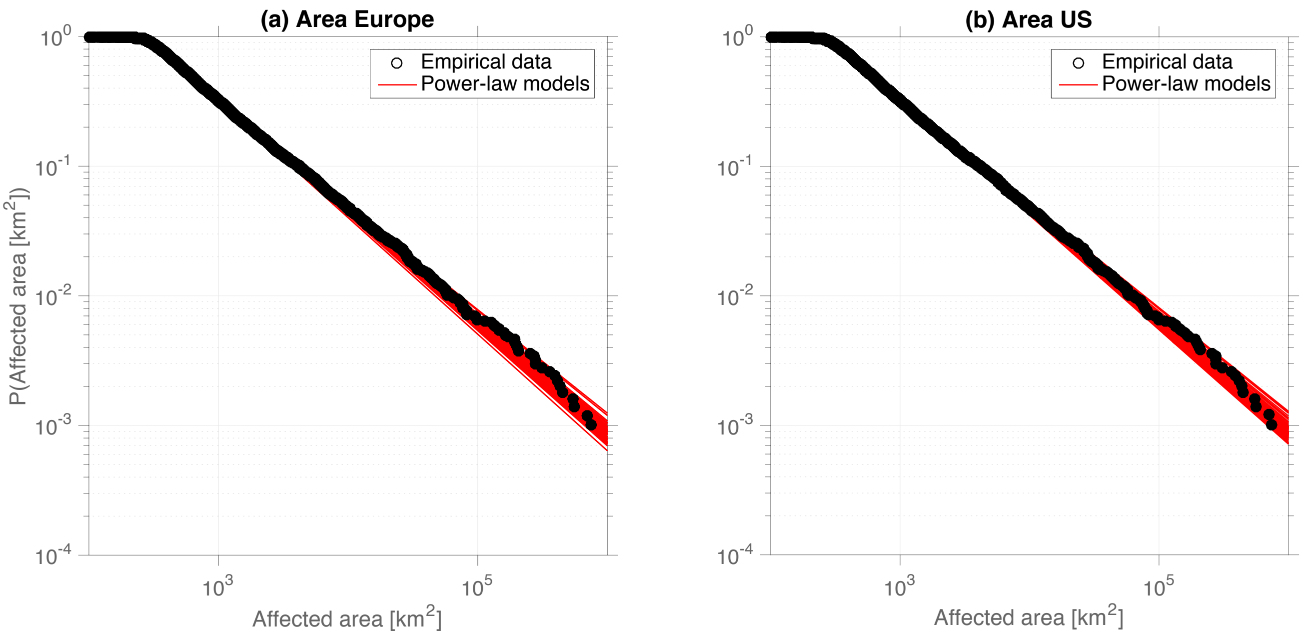

Figure C3. The probability distribution of areas affected by extremes in (a) Europe and (b) the US. The tails of the distributions can be described by power laws. The average scaling exponent for the tails is 1.85 for both cases.

Appendix D: Analogous example for soil moisture

D1 On the ISMN

The approach for testing a network design for its capacity to detect extremes is generic by construction. As an additional demonstration we explore the capacity of the **International Soil Moisture Network ISMN** (http://ismn.geo.tuwien.ac.at/ ?), a steadily growing initiative that comprises collections of soil moisture only. Comparable to FLUXNET there is no specific funding for measurement campaigns, and ISMN crucially depends on the contributions of historical observations by the respective communities.

Methodology

Direct observations of soil moisture from satellites are available (?), but these data still suffer from concatenating different data sources. And in fact these transitions make the data set very problematic for detecting extremes – or in other words, extreme event detection may identify the data merging edges. Alternatives are classical drought indicators that can be derived from climatological data only. Here, we rely on the Standardized Precipitation Index (SPI) for detecting extreme events as extracted from SPI and compare it to a random network of the same size (Fig. D1). The SPI is extracted following standard methodology (?) from monthly ERA-Interim rainfall data (?), using a 3-monthly aggregation window over the 1979-2011. We us the SPI only for illustration purposes until more robust EO for soil moisture become available, i.e. we assume that low SPI values are proxies for low soil moisture contents.

Further, a local 10th percentile threshold is applied on the SPI time series to flag dry events with subsequent detection of the large connected events. The choice of the local threshold is consistent with the typical meteorological/climatological use of SPI time series. Hence, in contrast to biophysical applications as presented in the main part of the paper, global or regional thresholds might not be physically meaningful for evaluating the local impacts of climate variables. Since meteorological reanalyses typically operate at much coarser resolution than EO data sets, for the analogous analysis presented here both the spatial and temporal search space are chosen to comprise only the spatially and temporally adjacent voxel (i.e. z = 0.5° and tau= 1 month in the SPI dataset).

To evaluate the ISMN, all station locations and the periods of active data sampling of each station were used for spatio-temporal intersection with the SPI extremes in two different setups: Firstly, we consider all stations active only in periods when these stations were collecting data (‘dynamic’ network); and secondly, a ‘static’ (counterfactual) situation is taken into account, where all stations are taken as active throughout the entire ERA-Interim period. The comparison was restricted to Europe due to data availability (i.e. most regional networks that form ISMN are operated in Europe (?)).

Results

If we consider the full spatiotemporal intersection we find that only the first five SPI extremes would have affected areas where the ISMN has stations (Fig. D1, red line). Higher ranked extremes are less likely of being detected. An annual random site placement (gray lines) would have been more efficient in identifying the extremes. In fact the current geographical coordinates u,v would have only reached the potential of a random network if they had been operated without ceasing over the entire monitoring period (blue lines). But that would have implied much more measurement years than the random site placement. For very high ranks of extremes (the very small events) the continuously operated real-world network would have outperformed the random network. These results are consistent with the results shown in the main paper.

An interesting feature of ISMN is that the network has changed its structure over the last decades to a very large extent. In the eighties, all station locations are confined to Eastern Europe (Fig. D2, upper panel). In the last decade, Western European station networks became active, but both the number and data availability from East European stations was severely reduced (Fig. 11, lower panel). This change in network design materializes strongly in the spatial locations of the detected events: While in the eighties most extremes in Eastern Europe where 'seen' by at least one tower and the detection rates in West Europe were poor, this pattern is reversed in the last decade (Fig. D2). Further, both decades highlight that a static random tower placement is more efficient than the current network, which is explicable by the high degree of site clustering. The importance of maintaining continuous observation alive becomes even more evident if one analyzes the network development over time in more detail (Fig. D3). In conclusion, the complementary analysis presented here substantiates the main paper in that the consideration both the spatial location and the availability of historical data is a crucial element to reconstruct the impacts of extreme events in the recent past.

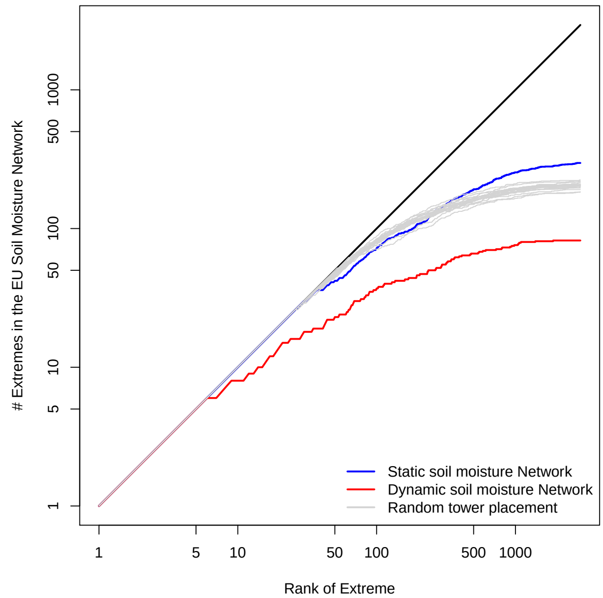

Figure D1. International Soil Moisture Network and its capacity to detect SPI extremes in Europe. Again red line shows the reduction of detection capacity due to inactive towers. Randomly placing observation years in space and time leads to higher detection rates for large extremes, and lower rates for small extremes.

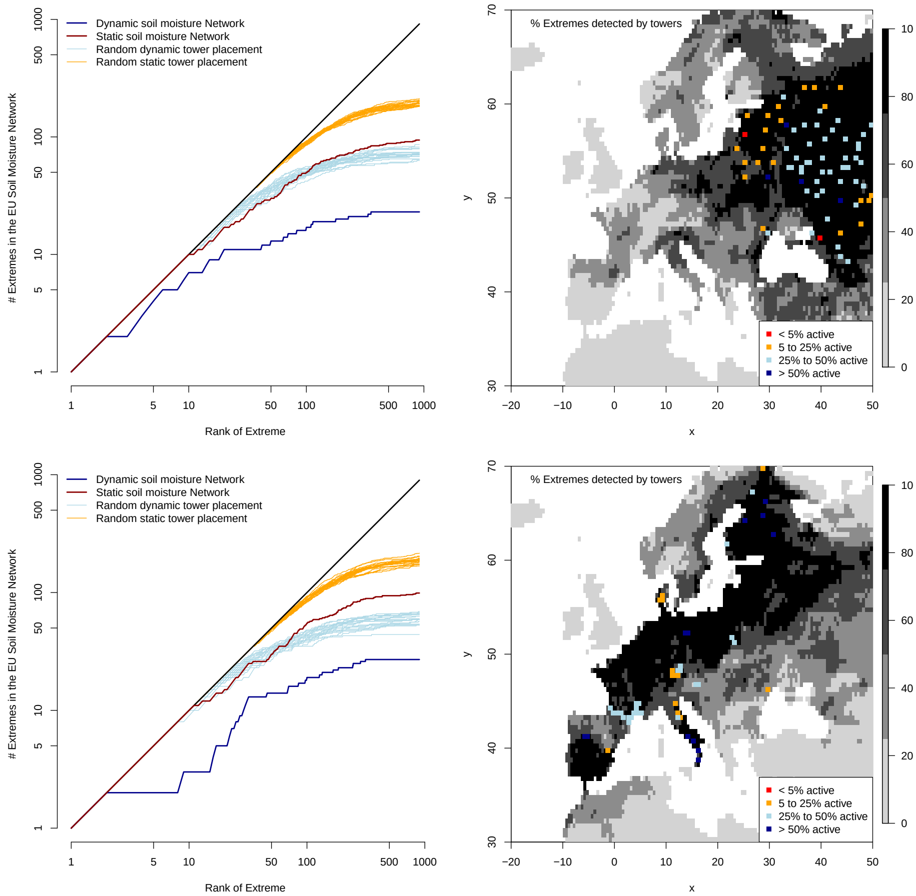

Figure D2. International Soil Moisture Network and its capacities to detect SPI extremes in Europe vs. a random network for the 1980ies (upper row) and 2000s (bottom row).

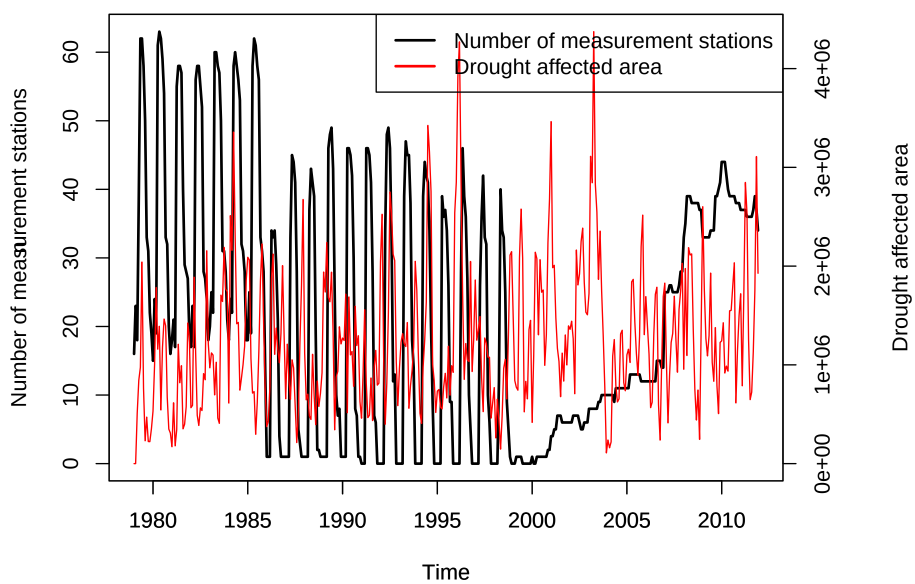

Figure D3. Number of stations in the International Soil Moisture Network over time confronted with drought affected area.

References

Anyamba, A., Small, J. L., Britch, S. C., Tucker, C. J., Pak, E. W., Reynolds, C. A., Crutchfield, J., and Linthicum, K. J.: Recent Weather Extremes and Impacts on Agricultural Production and Vector-Borne Disease Outbreak Patterns, Plos ONE, 9, e92 538, doi:https://doi.org/10.1371/journal.pone.0092538, 2014.

Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., and Vesala, T.: Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology, Advances in Ecological Research, 30, 113–175, 2000.

Aubinet, M., Vesala, M., and Papale, D., eds.: Eddy Covariance –A Practical Guide to Measurement and Data Analysis, Springer, 2012.

Babst, F., Poulter, B., Bodesheim, P., Mahecha, M. D., and Frank, D. C.: Improved tree-ring archives will support earth-system science, Nature Ecology & Evolution, 1, 0008, doi:10.1038/s41559-016-0008, http://dx.doi.org/10.1038/s41559-016-0008, 2017.

Balddocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – the state and future of the eddy covariance method, Global Change Biology, 20, 3600–3609, 2014.

Baldocchi, D.: Turner Review No. 15. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems, Australian Journal of Botany, 56, 1–26, doi:10.1071/BT07151, 2008.

Bastos, A., Gouveia, C. M., Trigo, R. M., and Running, S. W.: Analysing the spatio-temporal impacts of the 2003 and 2010 extreme heatwaves on plant productivity in Europe, Biogeosciences, 11, 3421–3435, 2014.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate, Science, 329, 834–838, doi:10.1126/science.1184984, 2010.

Carvalhais, N., Reichstein, M., Collatz, G. J., Mahecha, M. D., Migliavacca, M., Neigh, C. S. R., Tomelleri, E., Benali, A. A., Papale, D., and Seixas, J.: Deciphering the components of regional net ecosystem fluxes following a bottom–up approach for the Iberian Peninsula, Biogeosciences, 7, 3707–3729, doi:10.5194/bg-7-3707-2010, 2010.

Chen, X., Long, D., Hong, Y., Liang, S., and Hou, A.: Observed radiative cooling over the Tibetan Plateau for the past three decades driven by snow cover-induced surface albedo anomaly, Journal of Geophysical Research: Atmospheres, 122, 6170–6185, doi:10.1002/2017JD026652, http://dx.doi.org/10.1002/2017JD026652, 2017JD026652, 2017.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, Nature, 437, 529–533, doi:10.1038/nature03972, 2005.

Clauset, A. and Woodbard, R.: Estimating the historica and future probabilities of large terrorist events, The Annals of Applied Statistics, 7, 1838–1865, 2013.

Clauset, A., Shalizi, C. R., and Newman, M. E. J.: Power-law distribution in empirical data, SIAM Review, 51, 661–703, doi:10.1137/070710111, 2009.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., Jackson, T., and Jackson, T.: The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements, Hydrology and Earth System Sciences, 15, 1675–1698, 2011.

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., Zamojski, D., Cordes, C., Wagner, W., and Drusch, M.: Global Automated Quality Control of In situ Soil Moisture data from the International Soil Moisture Network, Vadose Zone Journal, 12, doi:10.2136/vzj2012.0097, 2013.

Flach, M., Gans, F., Brenning, A., Denzler, J., Reichstein, M., Rodner, E., Bathiany, S., Bodesheim, P., Guanche, Y., Sippel, S., and Mahecha, M.: Multivariate Anomaly Detection for Earth Observations: A Comparison of Algorithms and Feature Extraction Techniques, Earth System Dynamics - Discussions, doi:10.5194/esd-2016-51, 2017.

Frank, D., Reichstein, M., Bahn, M., Frank, D., Mahecha, M. D., Smith, P., Thonike, K., van der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J. G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S. I., Walz, A., Wattenbach, M., Zavala, M. A., and Zscheischler, J.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, Global Change Biology, 21, 2861–2880, 2015.

Global Terrestrial Observing System: Terrestrial essential climate variables for assessment, mitigation and adaptation (GTOS–52, edited by R. Sessa and H. Dolman), Tech. rep., Food and Agricultural Organiazation of the U. N. Rome., Rome, 2008.

Hargrove, W. W. and Hoffman, F. M.: Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions, Environmental Management, 34, 39–60, 2004.

Hartmann, H., Adams, H. D., Anderegg, W. R. L., Jansen, S., and Zeppel, M. J. B.: Research frontiers in drought-induced tree mortality: crossing scales and disciplines, New Phytologist, 205, 965–969, doi:10.1111/nph.13246, http://dx.doi.org/10.1111/nph.13246, 2014-18763, 2015.

Houborg, R., Fisher, J. B., and Skidmore, A. K.: Advances in remote sensing of vegetation function and traits, International Journal of Applied Earth Observation and Geoinformation, 43, 1–6, 2015.

Huesca, M., Merino-de Miguel, S., Eklundh, L., Litago, J., Cicuéndez, C., Rodríguez-Rastrepo, M., L., U. S., and Palacios-Ortueta, A.: Ecosystem functional assessment based on the "optical type" concept and self-similarity patterns: An application using MODIS-NDVI time series autocorrelation, International Journal of Applied Earth Observation and Geoinformation, 43, 132–148, 2015.

IPCC: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, Tech. rep., Field, C. B. and V. Barros and T. F. Stocker and D. Qin and D. J. Dokken and K. L. Ebi and M. D. Mastrandrea and K. J. Mach abd G.-K. Plattner and S. K. Allen and M. Tignor and P. M. Midgley, Cambridge University Press, Cambridge, 2012.

Jung, M., Verstraete, M., Gobron, N., Reichstein, M., Papale, D., Bondeau, A., Robustelli, M., and Pinty, B.: Diagnostic assessment of European gross primary production, Global Change Biology, 14, 2349–2364, doi:10.1111/j.1365-2486.2008.01647.x, 2008.

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001–2013, doi:10.5194/bg-6-2001-2009, 2009.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Niovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, doi:10.1038/nature09396, 2010.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land–atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, Journal of Geophysical Research–Biogeosciences, 116, G00J07, doi:10.1029/2010JG001566, 2011.

Kaminski, T. and Raynner, P. J.: Assisting the Evolution of the Observing System for the Carbon Cycle through Quantitative Network Design, Biogeosciences Discussions, doi:10.5194/bg-2017-168,, 2017.

Kaminski, T., Scholze, M., and Houweling, S.: Quantifying the Benefit of A-SCOPE Data for Reducing Uncertainties in Terrestrial Carbon Fluxes in CCDAS, Tellus B, 62, 784–796, doi:10.1111/j.1600-0889.2010.00483.x, http://journals.sfu.ca/coaction/index.php/tellusb/article/view/16634, 2010.

Kao, R. H., Gibson, C. M., Gallery, R. E., Meier, C. L., Barnett, D. T., Docherty, K. M., Blevins, K. K., Travers, P. D., Azuaje, E., Springer, Y. P., Thibault, K. M., McKenzie, V. J., Keller, M., Alves, L. F., Hinckley, E.-L. S., Parnell, J., and Schimel, D.: NEON terrestrial field observations: Designing continental-scale, standardized sampling, Ecosphere, 3, 115, 2012.

Keller, M., Schimel, D. S., Hargrove, W. W., and Hoffman, F. M.: A continental strategy for the National Ecological Observatory Network, Frontieres in Ecology and the Environment, 6, 282–284, 2008a.

Keller, M., Schimel, D. S., Hargrove, W. W., and Hoffman, F. M.: A continental strategy for the National Ecological Observatory Network, Frontiers in Ecology and the Environment, 6, 282–284, 2008b.

Knorr, W., Gobron, N., Scholze, M., Kaminski, T., Schnur, R., , and Pinty, B.: Impact of terrestrial biosphere carbon exchanges on the anomalous CO_2 increase in 2002–2003, Geophyical Research Letters, 34, L09 703, doi:10.1029/2006GL029019, 2007.

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe, M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active microwave satellite–based retrievals, Hydrology and Earth System Sciences, 15, 425–436, doi:10.5194/hess-15-425-2011, http://www.hydrol-earth-syst-sci.net/15/425/2011/, 2011.

Lloyd-Hughes, B.: A spatio–temporal structure–based approach to drought characterisation, International Journal of Climatology, 32, 406–418, doi:10.1002/joc.2280, 2012.

Ma, X., Huete, A., Moran, S., Ponce-Campos, G., and Eamus, D.: Abrupt shifts in phenology and vegetation productivity under climate extremes, Journal of Geophysical Research: Biogeosciences, 120, 2036–2052, doi:10.1002/2015JG003144, http://dx.doi.org/10.1002/2015JG003144, 2015JG003144, 2015.

McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., and Nilsson, S.: Comparison of four global FAPAR datasets over Northern Eurasia for the year 2000, Remote Sensing of Environment, 114, 941–949, doi:10.1016/j.rse.2009.12.009, 2010.

McDowell, N. G., Coops, N. C., Beck, P. S. A., Chambers, J. Q., Gangodagamage, C., Hicke, J. A., Huang, C.-Y., Kennedy, R., Krofcheck, D. J., Litvak, M., Meddens, A. J. H., Muss, J., Robinson Negrón-Juarez, R., Peng, C., Schwantes, A. M., Swenson, J. J., Vernon, L. J., Williams, A. P., Xu, C., Zhao, M., Running, S. W., and Allen, C. D.: Global satellite monitoring of climate-induced vegetation disturbances, Trends in Plant Science, 20, 114 – 123, doi:http://dx.doi.org/10.1016/j.tplants.2014.10.008, http://www.sciencedirect.com/science/article/pii/S1360138514002726, 2015.

McKee, T. B., Doesken, N. J., Kleist, J., et al.: The relationship of drought frequency and duration to time scales, in: Proceedings of the 8th Conference on Applied Climatology, vol. 17, pp. 179–183, American Meteorological Society Boston, MA, USA, 1993.

Monteith, J. l.: Climate and Efficiency of Crop Production in Britain, Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 281, 277–294, 1977.

Nasahara, K. N. and Nagai, S.: Review: Development of an in situ observation network for terrestrial ecological remote sensing: the Phenological Eyes Network (PEN), Ecological Research, 30, 211–223, doi:10.1007/s11284-014-1239-x, http://dx.doi.org/10.1007/s11284-014-1239-x, 2015.

Newman, M. E. J.: Power laws, Pareto distribution and Zipf's law, Contemporary Physics, 46, 323–351, doi:10.1080/00107510500052444, 2005.

Nicolai-Shaw, N., Zscheischler, J., Hirschi, M., Gudmundsson, L., and Seneviratne, S. I.: A drought event composite analysis using satellite remote-sensing based soil moisture, Remote Sensing of Environment, doi:http://dx.doi.org/10.1016/j.rse.2017.06.014, 2017.

Niu, S., Luo, Y., Li, D., Cao, C., Xia, J., Li, J., and smith, M. B.: Plant growth and mortality under climatic extremes: An overview, Environmental and Experimental Botany, 98, 13–19, 2014.

Oliphant, A. J.: Terrestrial Ecosystem-Atmosphere Exchange of CO_2, Water and Energy from FLUXNET; Review and Meta-Analysis of a Global in-situ Observatory, Geography Compass, 6, 689–705, 2012.

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., Kiely, G., Lasslop, G., Mahecha, M. D., Margolis, H., Merbold, L., Montagnani, L., Moors, E., Olesen, J. E., Reichstein, M., Tramontana, G., van Gorsel, E., Wohlfahrt, G., and Ráduly, B.: Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks, Journal of Geophysical Research: Biogeosciences, 120, 1941–1957, doi:10.1002/2015JG002997, http://dx.doi.org/10.1002/2015JG002997, 2015JG002997, 2015.

Pfeifer, M., Disney, M. I., Quaife, T., and Marchant, R.: Terrestrial ecosystems from space: a review of earth observation products for macroecology applications, Global Ecology and Biogeography, 21, 603–624, 2012.

Pinty, B., Lavergne, T., Widlowski, J.-L., Gobron, N., and Verstraete, M. M.: On the need to observe vegetation canopies in the near–infrared to estimate visible light absorption, Remote Sensing of Environment, 113, 10–23, doi:10.1016/j.rse.2008.08.017, 2009.

Pinty, B., Andredakis, I., Clerici, M., Kaminski, T., Taberner, M., Verstraete, M. M., Gobron, N., Plummer, S., and Widlowski, J.-L.: Exploiting the MODIS albedos with the two-stream Inversion Package (JRC-TIP): 2. Fractions of transmitted and absorbed fluxes in the vegetation and soil layers, Journal of Geophysical Research – Atmospheres, 116, 201, doi:10.1029/2010JD015373, 2011.

Rammig, A., Wiedermann, M., Donges, J. F., Babst, F., von Bloh, W., Frank, D., Thonicke, K., and Mahecha, M. D.: Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity, Biogeosciences, 12, 373–385, doi:10.5194/bg-12-373-2015, http://www.biogeosciences.net/12/373/2015/, 2015.

Rayner, P., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial Carbon fluxes from a Carbon Cycle Data Assimilation System (CCDAS), Global Biogeochemical Cycles, 19, 20 PP, doi:doi:10.1029/2004GB002254, http://www.agu.org/pubs/crossref/2005/2004GB002254.shtml, 2005.

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D. C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., and Wattenbach, M.: Climate extremes and the carbon cycle, Nature, 500, 287–295, doi:10.1038/nature12350, 2013.

Reyer, C. P., Leuzinger, S., Rammig, A., Wolf, A., Bartholomeus, R. P., Bonfante, A., de Lorenzi, F., Dury, M., Gloning, P., Abou Jaoudé, R., Klein, T., Kuster, T. M., Martins, M., Niedrist, G., Riccardi, M., Wohlfahrt, G., de Angelis, P., de Dato, G., François, L., Menzel, A., and Pereira, M.: A plant's perspective of extremes: terrestrial plant responses to changing climatic variability, Global Change Biology, 19, 75–89, doi:10.1111/gcb.12023, http://dx.doi.org/10.1111/gcb.12023, 2013.

Richardson, A. D., Klosterman, S., and Toomey, M.: Near-surface sensor-derived phenology, chap. 22, pp. 413–430, Phenology: An Integrative Environmental Science, Springer, 2013.

SanClements, M., Metzger, S., Luo, H., Pingintha-Durden, N., Zulueta, R. C., , and Loescher, H. W.: The National Ecological Observatory Network (NEON): Providing free long-term ecological data on a continental scale, iLEAPS newsletter, Special issue on Environmental Research Infrastructures,, 23–26, 2015.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., and Roy, D.: First operational BRDF, albedo nadir reflectance products from {MODIS}, Remote Sensing of Environment, 83, 135 – 148, doi:http://dx.doi.org/10.1016/S0034-4257(02)00091-3, http://www.sciencedirect.com/science/article/pii/S0034425702000913, the Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring, 2002.

Schimel, D., Pavlik, R., Fischer, J., Asner, G., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K., and Cox, P.: Observing terrestrial ecosystems and the carbon cycle from space, Global Change Biology, 32, 1762–1776, 2015.

Schwalm, C. R., Williams, C. A., Schaefer, K., Baldocchi, D., Black, T. A., Goldstein, A. H., Law, B. E., Oechel, W. C., Paw U, K. T., and Scott, R. L.: Reduction in carbon uptake during turn of the century drought in western North America, Nature Geosciences, 5, 551–556, 2012.

Seixas, J., Carvalhais, N., Nunes, C., and Benali, A.: Comparative analysis of MODIS–FAPAR and MERIS–MGVI datasets: Potential impacts on ecosystem modeling, Remote Sensing of Environment, 113, 2547–2559, doi:10.1016/j.rse.2009.07.018, 2009.

Sippel, S., Forkel, M., Rammig, A., Thonicke, K., Flach, M., Heimann, M., Otto, F. E. L., Reichstein, M., and Mahecha, M. D.: Contrasting and interacting changes in simulated spring and summer carbon cycle extremes in European ecosystems, Environmental Research Letters, 2, 075 006, doi:10.1088/1748-9326/aa7398., 2017.

Smith, M. D.: An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research, Journal of Ecology, 99, 656–663, 2011.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, doi:10.5194/bg-13-4291-2016, http://www.biogeosciences.net/13/4291/2016/, 2016.

Ustin, S. L. and Gamon, J. A.: Remote sensing of plant functional types, New Phytologist, 186, 795–816, doi:10.1111/j.1469-8137.2010.03284.x, 2010.

Venema, V., Ament, F., and Simmer, C.: A Stochastic Iterative Amplitude Adjusted Fourier Transform algorithm with improved accuracy, Nonlinear Processes in Geophysics, 13, 321–328, doi:10.5194/npg-13-321-2006, http://www.nonlin-processes-geophys.net/13/321/2006/, 2006a.

Venema, V., Theis, S., and Simmer, C.: Online generation of temporal and spatial fractal red noise, Geophysical Research Abstracts, 8, 09 460, 2006b.

Verstraete, M. M., Gobron, N., Aussedat, O., Robustelli, M., Pinty, B., Widlowski, J.-L., and Taberner, M.: An automatic procedure to identify key vegetation phenology events using the JRC–FAPAR products, Advances in Space Research, 41, 1773–1783, doi:10.1016/j.asr.2007.05.066, 2008.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., and Wang, Y.-P.: Improving land surface models with FLUXNET data, Biogeosciences, 6, 1341–1359, doi:10.5194/bg-6-1341-2009, http://www.biogeosciences.net/6/1341/2009/, 2009.

Wingate, L., Ogée, J., Cremonese, E., Filippa, G., Mizunuma, T., Migliavacca, M., Moisy, C., Wilkinson, M., Moureaux, C., Wohlfahrt, G., Hammerle, A., Hörtnagl, L., Gimeno, C., Porcar-Castell, A., Galvagno, M., Nakaji, T., Morison, J., Kolle, O., Knohl, A., Kutsch, W., Kolari, P., Nikinmaa, E., Ibrom, A., Gielen, B., Eugster, W., Balzarolo, M., Papale, D., Klumpp, K., Köstner, B., Grünwald, T., Joffre, R., Ourcival, J.-M., Hellstrom, M., Lindroth, A., Charles, G., Longdoz, B., Genty, B., Levula, J., Heinesch, B., Sprintsin, M., Yakir, D., Manise, T., Guyon, D., Ahrends, H., Plaza-Aguilar, A., Guan, J. H., and Grace, J.: Interpreting canopy development and physiology using the EUROPhen camera network at flux sites, Biogeosciences Discussions, 12, 7979–8034, doi:10.5194/bgd-12-7979-2015, http://www.biogeosciences-discuss.net/12/7979/2015/, 2015.

Xiao, J., Chen, J., Davis, K. J., and Reichstein, M.: Advances in upscaling of eddy covariance measurements of carbon and water fluxes, Journal of Geophysical Research – Biogeosciences, 117, G00J01, doi:10.1029/2011JG001889, 2012.

Zscheischler, J., Mahecha, M. D., Harmeling, S., and Reichstein, M.: Detection and attribution of large spatiotemporal extreme events in Earth observation data, Ecological Informatics, 15, 66–73, doi:10.1016/j.ecoinf.2013.03.004, 2013.

Zscheischler, J., Mahecha, M. D., von Buttlar, J., Harmeling, S., Jung, M., Rammig, A., Randerson, J. T., Schölkopf, B., Seneviratne, S. I., Tomelleri, E., Zaehle, S., and Reichstein, M.: Few extreme events dominate global interannual variability in gross primary production, Environmental Research Letters, 9, 035 001, doi:10.1088/1748-9326/9/3/035001, 2014a.

Zscheischler, J., Reichstein, M., Harmeling, S., Rammig, A., Tomelleri, E., and Mahecha, M. D.: Extreme events in gross primary production: a characterization across continents, Biogeosciences, 11, 2909–2924, doi:10.5194/bg-11-2909-2014, http://www.biogeosciences.net/11/2909/2014/, 2014b.