

Multivariate Hydrometeorological
Extreme Events and their Impacts
on Vegetation -
Potential Methods and Applications

Dissertation
(kumulativ)

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat)

vorgelegt dem Rat der Chemisch-Geowissenschaftlichen Fakultät der
Friedrich-Schiller-Universität Jena

von Milan Flach (M. Sc.)

geboren am 08.07.1988 in München

II

Gutachter:

1. Prof. Dr. Alexander Brenning
Institut für Geographie, Friedrich-Schiller-Universität Jena
2. Prof. Dr. Miguel D. Mahecha
Remote Sensing Center for Earth System Research, Universität Leipzig

Tag der Verteidigung: 06. November 2020

Contents

Summary	VII
Zusammenfassung	XI
1 Introduction	1
1.1 Motivation	1
1.1.1 Impacts of droughts and heatwaves on humans and society . .	3
1.1.2 Impacts of droughts and heatwaves on the carbon cycle	4
1.1.3 The role of vegetation during droughts and heatwaves	5
1.1.4 Extreme events and climate change	7
1.1.5 Towards a multivariate perspective on extreme events	9
1.2 Objectives and structure of the thesis	10
2 Background	13
2.1 On the definition of extreme events	13
2.2 Overview on multivariate detection concepts	16
2.2.1 Statistical process monitoring	16
2.2.2 Novelty detection	17
2.2.3 Anomaly detection	18
3 Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques	21
4 Contrasting biosphere responses to hydrometeorological extremes: revisiting the 2010 western Russian heatwave	43
5 Vegetation modulates the impact of climate extremes on gross primary production	65
6 Discussion	87
6.1 Alternative pathways	87
6.2 Sensitivity of the results	88
6.3 Timing of droughts and heatwaves	89

6.4	Lagged effects in forests	90
6.5	A more general view on impacts of droughts and heatwaves	91
6.6	Climate extremes beyond droughts and heatwaves	92
6.7	Differences between the multivariate and univariate perspective	92
6.8	Outlook	95
6.8.1	Further potentials of the multivariate perspective	95
6.8.2	Novel climates	97
7	Concluding summary	99
	Acknowledgments	103
	Übersicht der Manuskripte	121
	Erklärung zu den Eigenanteilen	123
	Selbständigkeitserklärung	127
S1	Supplementary material	129
S1.1	Multivariate perspective compared to single variables	129
S1.2	An exemplary typology of extreme events	129
S1.3	Sensitivity of the regionalisation procedure	131
S2	Supplementary material of chapter 3	133
S3	Supplementary material of chapter 5	137

Summary

Motivation. Climate extreme events such as the 2018 drought and heatwave in Germany often affect the well-being of humans and society. In Germany, increased oil prices due to the low water levels during the 2018 event as well as potentially lagged impacts through bark beetle attacks on forests in 2019 attracted considerable media attention.

In general, droughts can lead to lower drinking water quality and scarcity of water, which can directly affect humans or trigger armed conflict and violence. Harvest losses due to droughts and heatwaves can increase food prices and lead to hunger or even mass migration due to the loss of income in agriculture, highlighting just a few of the many possible impacts of such events on humans and society.

Furthermore, droughts and heatwaves are known for their severe impacts on vegetation and the global carbon cycle. Although terrestrial ecosystems are usually a carbon sink, they can turn into a carbon source during and after climate extreme events. In severe cases, the impacts of these events on terrestrial ecosystems can undo years of carbon sequestration.

In the context of climate change, the negative impacts of climate extreme events have to be taken into account. The number of heatwaves is expected to increase, as well as the number of droughts in some regions. More frequent climate extreme events with negative impacts on vegetation and the terrestrial carbon cycle can even partly offset human efforts to reduce carbon emissions by frequently turning some ecosystems into carbon sources.

A key aspect in the context of impacts of droughts and heatwaves on the carbon cycle is the role of different plant species and vegetation types at different scales. The physiological mechanisms during droughts and heatwaves are differing on a species level, but also on a larger scale e.g. different rooting depths between grasslands, forests and crops suggest differences in the associated responses to climate extremes. These differences in carbon uptake of different vegetation types have not yet been systematically evaluated for droughts and heatwaves on a global scale.

Furthermore, it should be noted that these climate extreme events are inherently multivariate. It is well known that the capacity of hot air for storing water is higher than the capacity of cold air for storing water. Extreme events that have had a strong

impact in the past were frequently concurrent droughts and heatwaves, like Europe in 2003, Russia in 2010, or the US in 2012. It is expected that these *compounding* droughts and heatwaves will increase in many regions in the future.

The aforementioned multivariate nature of climate extreme events such as droughts and heatwaves requires a multivariate perspective on climate extreme events. Common peak over threshold detection of extreme events in climate science and related disciplines uses the marginal distribution of the variables and does not consider covariation among the variables, potential non-linearities or can not be applied to high-dimensional settings. Thus, multivariate extreme event detection schemes have to be adapted and transferred from other disciplines that provide a multivariate perspective on climate extreme events.

Objectives. The overall objective of my thesis is to *improve the detection and understanding of climate extremes and their impacts on vegetation by facilitating a broader multivariate perspective that complements previous approaches to detect extreme events.*

1. As multivariate extreme event detection schemes are rarely used in climate science and related disciplines, I will first transfer and adapt methods from other disciplines developed for the detection of multivariate anomalies (chapter 3). The objective of this study is to evaluate which *combination of multivariate anomaly detection algorithm and feature extraction is best suited for detecting anomalous events.*
2. One particularly well-performing algorithm from the artificial experiment is then chosen to detect events in variables describing hydrometeorology and the exchange of matter between the atmosphere and the terrestrial surface (chapter 4). The objective is to *evaluate whether a broader multivariate perspective facilitates our understanding of extreme events and their impacts by revealing previously overlooked facets.* This is done using the example of a well studied extreme event of the past: the Russian heatwave in 2010.
3. One result of the analysis of the Russian heatwave is that forests and crops are associated with different and largely contrasting responses during the Russian heatwave. Chapter 5 evaluates whether different responses between vegetation types are just a singular case study or are common during droughts and heatwaves globally. The objective is to *evaluate the importance of different vegetation types on shaping the impact of climate extremes relative to other factors,* which is still highly uncertain on the global scale.

Methods and Main Results. In order to address the first objective, I implemented and tested several combinations of potential multivariate detection algorithms with preprocessing and feature extraction steps (such as dimensionality reduction) on artificial data that mimic various types of anomalous conditions, including extreme

events (chapter 3). I have identified one of the detection schemes as outperforming other approaches and being able to handle the covariation between the variables, non-linearities and high-dimensional settings in a more accurate way. The detection scheme consists of (i) subtracting the seasonality and (ii) a principal component analysis to account for the linear covariation and to reduce the dimensionality. It (iii) uses kernel density estimators in the reduced principal component space to detect deviations from the general multivariate and potentially non-linear distribution.

Second, this detection scheme is used on two sets of variables: one set of variables describing hydrometeorology and one set of variables of fluxes between the atmosphere and the terrestrial land surface (chapter 4). I chose one particularly well-studied case of an extreme event, the 2010 Russian heatwave, in order to evaluate the results and to compare them to existing studies. A major, existing finding is that, depending on the researcher's perspective, the affected area or volume of the so-called Russian heatwave is very different and strongly dependent on the focus on different variables. This mismatch between different affected areas of the Russian heatwave can be easily remedied by the proposed multivariate extreme event detection scheme, complementing earlier studies on this event. Furthermore, it is shown that northern forest ecosystems are associated with enhanced productivity during the Russian heatwave. In contrast, agricultural systems are strongly affected by the heatwave. Taking into account several confounding factors, it is shown that the vegetation type is indeed the most important factor for this contrasting effect.

Third, the role of different vegetation types during droughts and heatwaves is globally assessed between 2003 and 2018 by applying the developed detection scheme to hydrometeorological variables (temperature, radiation, water availability). For this evaluation, droughts and heatwaves that occur during the growing season from this set of extreme events are selected. The associated productivity during the droughts and heatwaves is modeled using boosted regression trees. The main result is that besides the factors background climate and duration in the statistical model, the vegetation type is particularly important for the direction of the associated productivity response. Forests are more often associated with immediate enhanced productivity during droughts and heatwaves than agriculture and other ecosystems. Thus, contrasting vegetation responses in case studies like the Russian heatwave 2010 are not just a single occurrence but can be observed frequently in global estimates of gross primary productivity. Although vegetation type related differences in the response to climate extremes are plausible to a certain degree due different water management and (micro-)climate, these results also point to potential biases with respect to remote sensing derived estimates of gross primary productivity.

General Conclusions and Outlook. Overall the thesis introduces a detection scheme for multivariate extreme events and applies it in the context of droughts and

heatwaves and their impacts on vegetation and the carbon cycle. It demonstrates that consistent application of this multivariate detection scheme can complement our knowledge of specific cases of extreme events as well as the impact of these events on the carbon cycle and the modulating role of different vegetation types during extreme events globally. However, the focus on immediate productivity during droughts and heatwaves requires a thorough assessment of lagged effects of extreme events in future studies. Further ground-based observations would greatly reduce the uncertainty of the findings which largely rely on remote sensing-based retrievals of productivity. Future studies should improve these remote-sensing derived retrievals of productivity with respect to their sensitivity to extreme events and specific vegetation types. In general, the proposed methodology of this thesis can be easily extended to detect any type of extreme events and to evaluate changes in the future composition of multivariate climate extreme events.

Zusammenfassung

Motivation. Klimaextremereignisse, wie die Trockenheit und Hitzewelle in Deutschland 2018, beeinflussen das Wohlbefinden von Menschen in unserer Gesellschaft. Die gestiegenen Ölpreise fanden beispielsweise aufgrund des niedrigen Wasserstandes während der Trockenheit und Hitze im Jahr 2018 in Deutschland, sowie wegen dem Baumsterben aufgrund von Borkenkäferbefall im Jahr 2019 in den Medien große Beachtung.

Trockenheit kann im Allgemeinen zu verringerter Trinkwasserqualität und -knappheit führen, was sich direkt auf die Gesundheit des Menschen auswirken kann oder sogar bewaffnete Konflikte und Gewalt nach sich ziehen kann. Ernteverluste aufgrund von Dürre und Hitzewellen können die Lebensmittelpreise erhöhen. Die daraus resultierenden Einkommensverluste in der Landwirtschaft können Hunger oder sogar Massenmigration auslösen, um nur einige von vielen weiteren möglichen Auswirkungen auf Menschen und die Gesellschaft zu nennen.

Darüber hinaus können Trockenheit und Hitzewellen starke Auswirkungen auf die Vegetation und den globalen Kohlenstoffkreislauf haben. Obwohl terrestrische Ökosysteme normalerweise eine Kohlenstoffs Senke sind, können sie sich während und nach extremen Klimaereignissen in eine Kohlenstoffquelle verwandeln. Damit können besonders starke Klimaextremereignisse die Kohlenstoffaufnahme in terrestrischen Ökosystemen von mehreren Jahren rückgängig machen.

Insbesondere im Zusammenhang mit dem Klimawandel müssen die negativen Auswirkungen der Klimaextremereignisse berücksichtigt werden. Die Anzahl der Hitzewellen wird voraussichtlich zunehmen. Ebenso wird die Anzahl der Trockenheiten in einigen Regionen ansteigen. Häufige Extremereignisse mit negativen Auswirkungen auf die Vegetation und den terrestrischen Kohlenstoffkreislauf können die Anstrengungen des Menschen zur Reduzierung der Kohlenstoffemissionen teilweise aufheben, wenn Ökosysteme regelmäßig in Kohlenstoffquellen verwandelt werden.

Ein entscheidender Aspekt ist die Rolle verschiedener Pflanzenarten und Vegetationstypen auf unterschiedlichen Skalen, die die Auswirkungen von Trockenheiten und Hitzewellen auf den Kohlenstoffkreislauf beeinflussen können. Die physiologischen Mechanismen während Trockenheiten und Hitzewellen unterscheiden sich deutlich auf der Ebene von Pflanzenarten. Aber auch auf größerer Skala sind unterschiedli-

che Reaktionen auf Klimaextremereignisse naheliegend, zum Beispiel aufgrund unterschiedlicher Wurzeltiefen zwischen Grasland, Wäldern und Kulturpflanzen. Diese Unterschiede in der Kohlenstoffaufnahme verschiedener Vegetationstypen sind auf globaler Ebene während Trockenheiten und Hitzewellen bislang nicht systematisch ausgewertet und vollständig verstanden.

Zudem sind Klimaextremereignisse von Natur aus multivariat. Heiße Luft kann mehr Wasser speichern als kalte Luft. Extremereignisse mit starken Auswirkungen traten in der Vergangenheit häufig gleichzeitig auf, wie beispielsweise Trockenheiten und Hitzewellen in Europa 2003, Russland 2010 oder die USA 2012. Es wird erwartet, dass diese kombinierten Trockenheiten und Hitzewellen in Zukunft in vielen Regionen zunehmen werden.

Die zuvor erwähnte multivariate Natur von Klimaextremereignissen erfordert eine multivariate Perspektive auf diese Ereignisse. Bisher findet die Detektion von Extremereignissen in den Klimawissenschaften und verwandten Disziplinen mittels Schwellenwerten statt. Typischerweise werden einzelne Variablen dafür genutzt und keine Kovariation zwischen verschiedenen Variablen oder mögliche Nichtlinearitäten berücksichtigt. Daher müssen multivariate Extremereigniserkennungsschemata aus anderen Disziplinen übertragen und für Klimawissenschaften angepasst werden, um eine multivariate Perspektive auf Klimaextremereignisse zu bieten.

Ziele. Das übergeordnete Ziel meiner Dissertation ist es, die Erkennung und das Verständnis von Klimaextremen und deren Auswirkungen auf die Vegetation zu verbessern, indem eine breitere multivariate Perspektive ermöglicht wird, die bisherige Ansätze zur Erkennung von Extremereignissen ergänzt.

1. Da multivariate Methoden zur Erkennung von Extremereignissen in den Klimawissenschaften und verwandten Disziplinen selten verwendet werden, übertrage und adaptiere ich zunächst Methoden aus anderen Disziplinen, die für die Erkennung multivariater Anomalien entwickelt wurden (Kapitel 3). Das Ziel dieser Studie ist es, *zu evaluieren, welche Kombination von Methoden zur Erkennung von multivariaten Anomalien und Vorverarbeitungsschritten am besten zum Erkennen anomaler Ereignisse geeignet ist.*
2. Ein besonders gut funktionierender Algorithmus auf künstlichen Daten wurde ausgewählt, um Ereignisse in Variablen zu detektieren, die die Hydrometeorologie und den Austausch von Kohlenstoff und Wasser zwischen Atmosphäre und Erdoberfläche beschreiben (Kapitel 4). Ziel ist es dabei, *zu bewerten, ob eine breitere multivariate Perspektive unser Verständnis von Extremereignissen und deren Auswirkungen verbessert, indem zuvor übersehene Facetten aufgedeckt werden.* Dies geschieht am Beispiel eines gut untersuchten Extremereignisses aus der Vergangenheit: die Fallstudie Hitzewelle in Russland 2010.
3. Ein Ergebnis der Fallstudie zur Hitzewelle in Russland 2010 ist, dass Wäl-

der und landwirtschaftliche Flächen mit unterschiedlichen und zum großen Teil gegensätzlichen Reaktionen während der Hitzewelle in Russland reagieren. In Kapitel 5 wird analysiert, ob unterschiedliche Reaktionen zwischen Vegetationstypen nur ein Einzelfall sind oder häufig bei Trockenheiten und Hitzewellen weltweit auftreten. Das Ziel dieser Studie besteht darin, die Bedeutung verschiedener Vegetationstypen für die Beeinflussung der Auswirkungen von Klimaextremereignissen im Verhältnis zu anderen Faktoren zu bewerten, was auf globaler Ebene immer noch sehr unklar ist.

Methoden und Hauptergebnisse. Um das erste Ziel zu erreichen, implementierte und testete ich mehrere Kombinationen von potenziellen multivariaten Detektionsalgorithmen und kombinierte sie mit Vorverarbeitungs- und Merkmalsextraktionsschritten (z. B. Dimensionsreduktion) in einem extra dafür erstellten Datensatz. Dieser Datensatz imitiert verschiedene Arten anomaler Bedingungen, einschließlich extremer Ereignisse (Kapitel 3). Ein Detektionsschema kann besonders gut anomale Bedingungen erkennen und dabei mit der Kovariation zwischen verschiedenen Variablen, Nichtlinearitäten und vielen Variablen umgehen. Das Detektionsschema besteht aus (i) Subtrahieren der Saisonalität, (ii) einer Hauptkomponentenanalyse, um die lineare Kovariation zu berücksichtigen und die Dimensionalität zu reduzieren, und (iii) einer Kerndichteschätzung im reduzierten Hauptkomponentenraum, um Abweichungen von der allgemeinen multivariaten und möglicherweise nicht-linearen Struktur der Daten zu erkennen.

Im zweiten Schritt wird dieses Detektionsschema für zwei Variablensätze verwendet. Ein Variablensatz beschreibt die Hydrometeorologie, und ein Variablensatz die Stoffflüsse zwischen der Atmosphäre und der Landoberfläche (Kapitel 4). Um die Ergebnisse zu bewerten und mit vorhandenen Studien zu vergleichen entschied ich mich für einen besonders gut untersuchten Fall eines Extremereignisses, die Hitzewelle in Russland 2010. Ein Hauptergebnis ist, dass abhängig von der Perspektive des Forschenden das betroffene Gebiet oder Volumen der sogenannten Hitzewelle in Russland sehr unterschiedlich ist. Es hängt stark vom Fokus der verschiedenen Variablen ab. Diese Unterschiede zwischen verschiedenen betroffenen Gebieten der russischen Hitzewelle kann durch das vorgeschlagene multivariate Detektionsschema leicht behoben werden und ergänzt frühere Studien zur Hitzewelle in Russland. Darüber hinaus wird gezeigt, dass in borealen Wäldern eine erhöhte Produktivität während der Hitzewelle in Russland zu beobachten ist. Landwirtschaftliche Systeme hingegen sind stark von der Hitzewelle betroffen. Unter Berücksichtigung mehrerer Störfaktoren wird gezeigt, dass der Vegetationstyp tatsächlich der wichtigste Faktor für diese gegensätzliche Reaktion ist.

Im dritten Schritt wird die Rolle verschiedener Vegetationstypen während Trockenheiten und Hitzewellen zwischen 2003 und 2018 global analysiert. Dafür wird

das entwickelte Detektionsschema auf hydrometeorologische Variablen (Temperatur, Strahlung, Wasserverfügbarkeit) angewendet. Aus den so identifizierten Extremereignissen werden Trockenheiten und Hitzewellen ausgewählt, die während der Vegetationsperiode stattfinden. Die Produktivität von Ökosystemen während Trockenheit und Hitzewellen wird unter Verwendung von boosted regression trees modelliert. Folgendes Hauptergebnis kristallisiert sich dabei heraus: Neben den Faktoren Klima und Dauer des Extremereignisses bestimmt der Vegetationstyp entscheidend, ob sich die Produktivität steigert oder im Vergleich zu langjährigen Mittel reduziert. Wälder sind viel häufiger mit einer Steigerung der Produktivität während Dürren und Hitzewellen assoziiert als die Landwirtschaft und andere Ökosysteme. Daher sind gegensätzliche Reaktionen wie bei der Hitzewelle in Russland 2010 nicht nur ein Einzelfall, sondern können häufig in globalen Schätzungen der Primärproduktivität beobachtet werden. Produktivitätsunterschiede zwischen unterschiedlichen Vegetationstypen sind durch unterschiedliches Wassermanagement und (Micro-)Klima bis zu einem gewissen Grad plausibel, deuten hier aber auch auf potentielle Probleme in Schätzungen der Produktivität mithilfe von Fernerkundung hin.

Allgemeine Schlussfolgerungen und Ausblick. Insgesamt wird in der Arbeit ein Detektionsschema für multivariate Extremereignisse vorgestellt. Dieses wird im Kontext von Trockenheiten und Hitzewellen sowie deren Auswirkungen auf die Vegetation und den Kohlenstoffkreislauf angewendet. Die Arbeit zeigt, dass eine konsequente Anwendung dieses multivariaten Detektionsschemas unser Wissen über spezifische Fallstudien von Extremereignissen ergänzen kann. Außerdem kann damit die Auswirkung von Extremereignissen auf den Kohlenstoffkreislauf untersucht werden. Die Arbeit ergänzt damit unser Wissen über die modulierende Rolle verschiedener Vegetationstypen während Extremereignissen. Der Fokus liegt hierbei auf die Produktivitätsveränderungen, die während Trockenheiten und Hitzewellen stattfinden. Dies erfordert jedoch auch eine Berücksichtigung der verzögerten Auswirkungen nach den Extremereignissen in weiteren Studien. Weitergehende bodengestützte Beobachtungen würden die Unsicherheit der Ergebnisse stark verringern, da die Ergebnisse größtenteils auf Schätzungen der Produktivität auf der Grundlage der Fernerkundung beruhen. Zukünftige Studien könnten diese auf Fernerkundung basierenden Schätzungen der Produktivität, im Hinblick auf Ihre Sensitivität für Extremereignisse und unterschiedliche Vegetationstypen, verbessern. Im Allgemeinen kann die vorgeschlagene Methodik leicht erweitert werden, um jede Art von Extremereignissen zu erkennen und Änderungen in der zukünftigen Zusammensetzung multivariater Klimaextremereignisse zu analysieren.

Chapter 1

Introduction

1.1 Motivation

2018 was a year of 'exceptional' weather, as stated by the vice president of the German Weather Service (Deutscher Wetterdienst, DWD) at the end of the year (DWD, 2018). According to their evaluation, 2018 was the hottest year since the start of the country-wide weather observations in 1881. Additionally, it was the sunniest and also one of the driest years in Germany (except southern Bavaria, Beenen et al., 2018). In terms of the multivariate distribution of temperature and precipitation anomalies, 2018 was clearly an exceptionally hot and dry year in that country (Fig. 1.1). This coincidence of several variables being far away from normal variability (temperature, precipitation, and radiation as reported by the DWD for 2018) seems to be a feature of high-impact extremes, e.g. like Europe 2003 or Russia 2010 (Fischer & Knutti, 2013). It already indicates that the monitoring of climate extreme events may be a multivariate task.

Importantly, climate extreme events, like in 2018, are known to impact society. For 2018, media reported wheat harvest losses of 25%, as well as low water levels in rivers in Germany. These low water levels led to the lowest hydroelectric power production for more than a decade. Furthermore, they hampered inland waterway transport which affected e.g. regional oil prices in the country (Beenen et al., 2018). This year (2019), tree mortality in Germany gained a lot of media attention (DPA, 2019; Köppe, 2019; Weiß, 2019). These reports of tree mortality in 2019 may potentially be related to the 2018 drought and heat event, as insect infestations are known to affect forests in the year after the extreme event (Rouault et al., 2006).

Recent scientific literature shows that the year 2018 was characterized by a climatic dipole with hot and dry conditions north of the Alps but opposite conditions (cool and moist) in the Mediterranean (Buras et al., 2019). Furthermore, in 2018 several concurrent heat events were affecting the Earth north of 30° latitude. These kinds of spatial concurrent events would very likely not have occurred without human-

Wie außergewöhnlich war das Jahr 2018?

Abweichung Temperatur und Niederschläge 1881 - 2018 für Deutschland

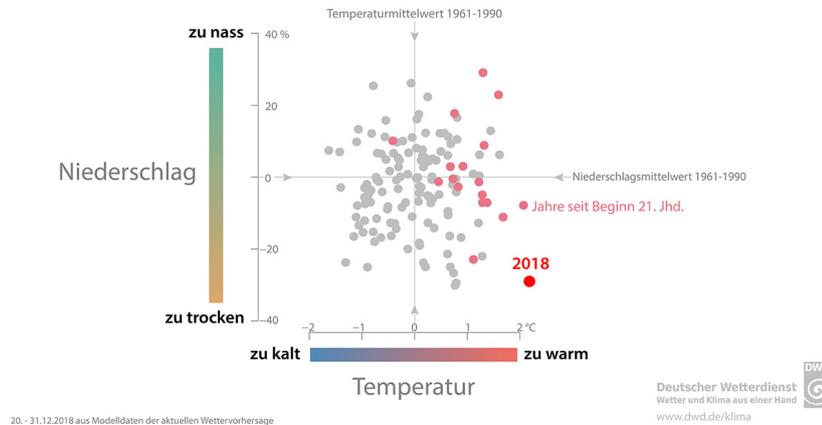


Figure 1.1 – ‘How exceptional was the year 2018?’ Deviation of temperature and precipitation between 1881 - 2018 in Germany. Red (blue) colors indicate too hot (cold) temperatures, green (orange) colors indicate too wet (dry) years in terms of precipitation. Figure from DWD, 2018.

induced climate change (Vogel et al., 2019). In general, it is expected that climate change leads to an increase in the frequency of heatwaves (Seneviratne et al., 2012), and compounding hot and dry events in some regions (Zscheischler & Seneviratne, 2017). Thus, climate extreme events are gaining more and more relevance in the context of climate change.

Generally, examples of climate extreme events (for a definition of extreme events see Section 2.1) include heavy precipitation, storms, droughts, and heatwaves, among others. Detecting and understanding those (climate) extreme events is crucial for society as they have severe impacts on humans (see Section 1.1.1), which may even be exacerbated when taking place simultaneously as so-called concurrent or compound events (AghaKouchak et al., 2014; Zscheischler et al., 2018). The overarching aim of the present thesis is to contribute to this inherently multivariate task of monitoring and understanding climate extreme events as well as their impacts and to *improve the detection and understanding of climate extremes and their impacts on vegetation by facilitating a broader multivariate perspective which complements previous approaches to detect extreme events.*

Understanding and detecting climate extreme events is important as climate extremes are directly or indirectly affecting humans and society (Section 1.1.1). They are also projected to increase in frequency and intensity in some regions of the world in future climate (Section 1.1.4, Easterling et al., 2000; Meehl & Tebaldi, 2004). However, one particular focus of this thesis is to detect climate extreme events and to evaluate their impacts on vegetation and the global carbon cycle. Humans are affected by vegetation productivity e.g. through agricultural yields and forestry (Section 1.1.1).

Furthermore, climate extremes are of interest as their associated impacts affect the global carbon cycle (Section 1.1.2, Reichstein et al., 2013), which feeds back into the climate system through changes in land and/or ocean carbon sinks. The carbon cycle-climate feedback is a crucial source of uncertainty in climate change projections (Section 1.1.4, Friedlingstein et al., 2014). Another focus is on two specific types of climate extreme events, namely droughts and heatwaves, as those two are expected to have the most severe impacts on the productivity of vegetation (Zscheischler et al., 2014b).

1.1.1 Impacts of droughts and heatwaves on humans and society

Extremes events, in general, can have severe impacts on human lives, which is one motivation to study extreme events. Here, I want to highlight the impacts of droughts and heatwaves on humans in particular.

Heatwaves. Heatwaves can affect humans in various ways. Observed impacts of heatwaves on humans include direct health impacts (increased mental health problems, Hansen et al., 2008, increased morbidity, Åström et al., 2011), or indirect health impacts via respiratory problems due to increased surface ozone (Filleul et al., 2006), or via lowering the protein content of crop yields (Dwivedi et al., 2013). Heatwaves can lead to stock mortality, e.g. of broilers (Vale et al., 2010). Heatwaves can also affect humans through affecting infrastructure directly (melting of asphalt, Reeves et al., 2010, bridge cracking, Zhou et al., 2016, airplane groundings, Smoyer-Tomic et al., 2003), or indirectly through power outages (e.g. due to high energy demand of air conditioning, Miller et al., 2008, or reduced generator cooling, De Bono et al., 2004). Reported economic impacts of heatwaves include impacts on tourism (Buckley & Foushee, 2011), or lowered labour productivity through increased absenteeism (Zander et al., 2015). Heatwaves may even enhance cases of suicide and aggressive behaviour (Berry et al., 2009). They can also lead to mass migration due to losses of crops and farming income in certain areas of the world (Mueller et al., 2014; Nawrotzki et al., 2015). The variety of impacts of heatwaves on humanity shows that there should be an inherent interest for society to detect and understand those kind of events.

Droughts. Observed impacts of droughts are similarly diverse and severe. Shortage (Shen et al., 2007), contamination (Benotti et al., 2010; Vliet & Zwolsman, 2008) and decreasing quality of drinking water can directly or indirectly affect human health and lead to diarrhoea, cholera, or dysentery (Calow et al., 2010). Increased dust emissions can impact human health as well (Miri et al., 2007; Prospero & Lamb, 2003). Virus outbreaks e.g. of the West Nile virus (Epstein, 2001), and chikungunya (Gould & Higgs, 2009) are associated with droughts if followed by flooding. Food quality

is affected by decreased protein content in crops (Dwivedi et al., 2013). Increased livestock mortality (Begzsuren et al., 2004; Terrence McCabe, 1987), and shellfish mortality (Wetz & Yoskowitz, 2013) have been reported. Agricultural losses, e.g. due to exceeding physiological limits in plants, can lead to starvation and economic losses affecting e.g. wheat prices globally (Wegren, 2010). Losses in forestry e.g. though massive dead wood due to insect attacks (Rouault et al., 2006) also lead to economic losses or may affect wood prices for energy consumption. Infrastructure during droughts can be affected by low water levels in rivers, hampering shipping (Van Lanen et al., 2016), locally increased oil prices are reported e.g. for 2018 (Hanrahan, 2018). Changing water levels affect tourism (Scott & Lemieux, 2010). Droughts lead to increased fire risk (Brando et al., 2014; Westerling et al., 2006), which destroy homes, or can increase migration from rural areas (McLeman & Smit, 2006; Meze-Hausken, 2000). Conflicts about water rights and access can be enhanced by droughts (AghaKouchak et al., 2015; Gleick, 2014) and droughts can trigger insurgency, violence (Hendrix & Salehyan, 2012; Uexkull et al., 2016) or armed conflicts in general (Mach et al., 2019). Hence, detecting and understanding droughts and heatwaves is important for early warning systems or impact-mitigation strategies.

Known case studies of drought and heatwave impacts show that it is often discussed or depending on the (impact) perspective whether the event was perceived as a drought or a heatwave. However, from a meteorological perspective, many famous large scale climate events are both hot and dry (Barriopedro et al., 2011; Ciais et al., 2005; Wolf et al., 2016). One such example is the so called European heatwave in the summer of 2003 which was perceived as a heatwave in the public and similarly in the scientific literature (Black et al., 2006; Coumou & Rahmstorf, 2012; Fischer et al., 2007; Schär & Jendritzky, 2004). This perception is probably related to the human perception of heat, and the observed mortality among elderly people associated with the heatwave (Garcia-Herrera et al., 2010). However, as carbon cycle impacts of the 2003 heatwave are more related to dry conditions than hot conditions (Granier et al., 2008; Reichstein et al., 2007), it is predominantly called a drought rather than a heatwave in the context of carbon cycle impacts (Frank et al., 2015; Reichstein et al., 2013). The impacts of droughts and heatwaves on the carbon cycle are the topic of Section 1.1.2.

1.1.2 Impacts of droughts and heatwaves on the carbon cycle

Another motivation to study extreme events is their impact on the terrestrial carbon cycle. In particular extreme temperature during heatwaves and the lack of water during droughts have impacts on vegetation productivity.

Extreme temperature during heatwaves leads to an increase of the atmospheric water demand (vapour pressure deficit), which first of all increases evapotranspira-

tion. To counteract the atmospheric demand of water and to prevent dry down, plants are capable of closing their stomata which leads to reduced gross primary productivity (the uptake of CO_2 by ecosystems). Furthermore, plants respiration may be enhanced during times of increased temperature (Bastos et al., 2014). Additionally, extreme temperatures can disrupt enzymatic activity needed for photosynthesis and subsequently change growth patterns (Larcher, 2003).

Similarly, the lack of water during droughts leads to drought stress in plants as soon as the evaporative demand is larger than the available water from soil. Gross primary productivity gets reduced through decreasing stomatal and mesophyll conductance, and enzymatic concentration and activity (Chaves et al., 2009; Keenan et al., 2010). In contrast to heatwaves, respiration rates during droughts are rather decreased (Bastos et al., 2014).

The loss of carbon on a large scale is reported in both droughts and heatwaves. For example, carbon losses for the European heatwave 2003 are estimated to offset three to five years of terrestrial net carbon uptake in the region (Ciais et al., 2005). For comparison, this is as much as half of the anthropogenic emissions of the European Union member states in 2015 (Sippel et al., 2018).

Secondary effects of droughts and heatwaves can have similarly strong effects on the carbon cycles as the event itself. During the Russian Heatwave 2010 fires are estimated to emit 70 Tg C into the atmosphere (Guo et al., 2017), which is of similar magnitude (90 Tg C) as the estimates of direct carbon losses in vegetation (Bastos et al., 2014). Additionally, other secondary effects like pest or pathogen outbreaks can be triggered by climate extremes as well and may even exceed carbon losses of the event itself (Frank et al., 2015; Rouault et al., 2006).

Droughts and heatwaves are strongly interlinked. During high temperatures, air can store more water than during low air temperature, i.e. the atmospheric demand for water is larger during heatwaves. This phenomenon leads to a faster drying out of soil moisture and to fewer rainfall than usual (Trenberth et al., 2013; Williams, 2012). Additionally, evaporative cooling is low during heat events (Teuling et al., 2010) and heat is further enhanced if soil moisture is limited (Hirschi et al., 2010; Seneviratne et al., 2006). Apart from this positive feedback loop of drought and heat events, dry and hot air can be transported and thus propagate to other neighbouring or teleconnected regions (Miralles et al., 2019). These processes show that droughts and heatwaves cannot be evaluated independently from each other (Mueller & Seneviratne, 2012), and inherently require a multivariate view on climate extremes (Section 1.2).

1.1.3 The role of vegetation during droughts and heatwaves

When motivating impacts of droughts and heatwaves on the carbon cycle, one has to consider that these impacts are largely mediated through different vegetation types.

Three main vegetation types are particularly important, as they cover more than half of the terrestrial surface on Earth: About 20%-28% of the Earth terrestrial surface are covered by forests (Erb et al., 2007; Feng et al., 2016; Keenan et al., 2015; Li et al., 2018), 10%-13% are covered by crops (Erb et al., 2007; Li et al., 2018; Monfreda et al., 2008; Portmann et al., 2010), and 24-36% are covered by grasslands and savannahs (Erb et al., 2007; Li et al., 2018). Uncertainties in the numbers arise e.g. from different data sources, timing of the analysis, spatial resolutions, methods to estimate the numbers, and also definitions of the vegetation types. In the context of the global carbon cycle, forest ecosystems play a major role. Gross primary productivity is globally estimated to be $122 \text{ Pg C year}^{-1}$. Forests make up about half ($59 \text{ Pg C year}^{-1}$) of the global gross primary productivity. For comparison, crops account for $15 \text{ Pg C year}^{-1}$, and grasslands and savannahs for $40 \text{ Pg C year}^{-1}$, despite grasslands and savannahs are covering equally large or even larger areas than forests (estimation of gross primary productivity from Beer et al., 2010).

In general impacts of climate extremes may differ among different vegetation types. It can be reasonably assumed that a climate extreme with a given magnitude hitting two different vegetation types at the same time will have a different impact on forests, crops, or grasslands and savannahs (Frank et al., 2015), e.g. due to vegetation type specific responses to climate extremes. Reasons for these differences may be related to physiological differences, (climatic) history and adaptation, or management.

The vulnerability of crops to climate extremes is of key interest. May be not due to the role of crops for the global carbon cycle, but more due to their importance for food production and food security. Vulnerability of crops highly depends on the timing, i.e the growth stage (Velde et al., 2011). Agricultural systems are highly managed systems: most crops are annually planted and then harvested. Thus, impacts of climate extremes can be managed, e.g. through (short term) irrigation, changes in timing of seeding or harvest, and other (more long-term) management strategies (Lobell et al., 2012; Pereira et al., 2002). Lagged effects in agricultural systems are rather not expected to play a role due to resetting the conditions through a mostly annual harvest (Frank et al., 2015).

For forests, large impacts of climate extremes on the carbon cycle are expected due to their global importance for the carbon cycle. However, the impacts on productivity in forests relative to other vegetation types are still uncertain. The degree of isohydricity differentiates the vulnerability of forest ecosystems to climate extremes on a species level (Roman et al., 2015; Ruehr et al., 2015; Yi et al., 2017). On an ecosystem level, forests have access to deeper soil water than other vegetation types (Fan et al., 2017; Yang et al., 2016), as well as a different water management, i.e. a higher water use efficiency for productivity compared to other ecosystems (Heerwaar-

den & Teuling, 2014; Teuling et al., 2010). Micro-climate in forests may also lead to different impacts of extremes events on forests, i.e. in forest generally lower temperatures and more humidity in the air are expected compared to adjacent ecosystems (Chen et al., 1993). Forests may also adapt to historic climate conditions, or weaken their investment tissue maintenance and defence (Doughty et al., 2015). These differences between species and ecosystem scale may translate into different impacts of droughts and heatwaves on forest ecosystems, which needs to be further scrutinized - in particular systematically on a global scale.

1.1.4 Extreme events and climate change

Climate change is another motivation to study climate extreme events: Small changes in the mean and the variance lead to much stronger changes in the tails of a variable's distribution, i.e. the changes in the extremes are expectedly large (Katz & Brown, 1992; Nicholls & Alexander, 2007). Then, changing patterns in climate extremes may induce sensitive feedback mechanisms which can further exacerbate climate change (Reichstein et al., 2013).

Generally, the statement of intensified frequencies and intensities of climate extremes may be oversimplified (Sippel et al., 2018). For heat extremes, the statement may be justified (Coumou & Robinson, 2013; Fischer & Knutti, 2014), and the frequency of heatwaves even increased globally during a pause in global mean temperature increase (the climate hiatus) (Seneviratne et al., 2014). However, droughts require a more nuanced description. As more water can be stored in the atmosphere with increasing temperatures, a quicker establishment of more intense (Trenberth et al., 2013) and longer lasting droughts (Fischer & Knutti, 2014) can indeed be expected. However, a higher variability in precipitation patterns can lead to strong regional differences of drought patterns (Pendergrass et al., 2017), and soil moisture can further complicate the patterns via inducing persistence and (mostly positive) feedback mechanisms of precipitation (Seneviratne et al., 2010). Thus, whereas the statement may be justified generally for heatwaves, changes in drought patterns are more complex and depend on the region of interest.

However, marine and atmospheric circulation patterns are also changing. For example the loss of sea ice due to climate change is associated with changes in large scale atmospheric circulation patterns (Overland & Wang, 2010). This process is taking place faster than expected (Dai et al., 2019). Additionally, the weakening of ocean circulation patterns, e.g. due to freshwater inflows (Stouffer et al., 2006), has a strong influence on climate (Minobe et al., 2008). Hence, the influence of changing large scale marine or atmospheric circulation patterns can be stronger than changes in local or regional climate.

The role of vegetation and the terrestrial carbon cycle within the process of climate

change is still under discussion and is also a source of uncertainty in climate change projections (Friedlingstein et al., 2014). Terrestrial carbon uptake may be increased due to elevated CO_2 . However, the effect saturates with time (Cao & Woodward, 1998). Furthermore, carbon uptake is limited by nitrogen and is thus enhanced by atmospheric nitrogen deposition globally. Other nutrient limitations may counteract this process (Zaehle & Dalmonech, 2011). Longer summers in high latitude regions may be another factor leading to a global increase in carbon uptake (Black et al., 2000; Myneni et al., 2001; Nemani et al., 2003). Thus, many factors point towards increased carbon uptake with climate change.

The role of climate extremes is crucial in this context. Although plant responses to elevated CO_2 may counteract droughts (Swann et al., 2016), increased carbon uptake with elevated CO_2 is reduced under more frequent extreme weather conditions (Obermeier et al., 2016). Furthermore, many carbon cycle models are discussed to underestimate (Schewe et al., 2019) or overestimate (Ukkola et al., 2016) the impacts of climate extremes. For example carbon cycle feedback to extremes and related insect outbreaks are ignored in climate models (Kurz et al., 2008). The occurrence of more frequent large scale climate extremes, like the European heatwave 2003, can turn some ecosystems into carbon sources instead of sinks (Ciais et al., 2005), or simply offset human efforts to reduce carbon emissions (Sippel et al., 2018). Additionally, the risk of concurrent (compounding) droughts and heatwaves in the future is increased due to the dependency structure of temperature and precipitation (Zscheischler & Seneviratne, 2017), which calls for a multivariate perspective on climate extremes (Zscheischler et al., 2018). This approach is also reflected in the objectives of this thesis (Section 1.2).

For forests in particular, large net impacts of climate extremes on the carbon cycle are expected (Frank et al., 2015). These are expected due to the global importance of forest ecosystems for the carbon cycle, their large carbon stocks of forest ecosystems, and potentially long-lasting legacy effects of forest dieback after droughts (Frank et al., 2015). However, their vulnerability to future droughts and heatwaves is still discussed (Allen et al., 2015). Some studies point towards a low vulnerability of forests and suggest that forests are capable of dealing with droughts in future hotter climate and may even benefit from global warming. In contradiction, other studies point towards a high vulnerability of forests to droughts and heatwaves, which will increase forest mortality and exceed critical physiological thresholds (Allen et al., 2015). Although evidence in previous research appears to be contradictory, forests with access to water may be rather resilient to drought and heat in the short term. In contrast, lagged responses or sustained drought and heat, may lead to tree mortality in the long term. Thus, the vulnerability of forests to droughts and heatwaves in a future climate still needs to be assessed systematically. Analysing vegetation

responses in current climate conditions may already improve our understanding of the vulnerability of forests to future conditions.

1.1.5 Towards a multivariate perspective on extreme events

Extreme events in climate science and related studies of ecosystems and the exchange of matter between the atmosphere and the terrestrial land surface are commonly detected as peak over threshold in marginal distributions of one variable (Section 2.1) or in absolute univariate indices defined in a purely climatological framework (Sillmann et al., 2013). If more than one variable is of interest, a frequently used approach is to detect peak over threshold extremes in each marginal distribution separately. In such a way detected extreme events in different variables can be connected via a logical *OR*, or a logical *AND* connection. While the former technique detects extreme events in any of the variables under scrutiny, it is often not explicitly applied, but implicitly e.g. by comparing maps of from different (climate) variables, like temperature, precipitation, or soil moisture (Bastos et al., 2014; Ciais et al., 2005; Reichstein et al., 2007). The latter technique is usually referred to as coincidences or co-exceedances (Donges et al., 2016; Donges et al., 2011; Guanche et al., 2016; Rammig et al., 2015; Siegmund et al., 2016; Zscheischler & Seneviratne, 2017; Zscheischler et al., 2015). It is used for the detection of compound extremes events, i.e. multivariate extreme events which occur simultaneously in more than one variable. However, one should be aware that both techniques are not considering the multivariate covariance structure between the variables of interest in the detection process.

In contrast to the previously mentioned approaches, the copula approach explicitly includes the covariance structure among different variables. Copulas are joint distribution functions of the marginal distributions, which are transformed to be uniform. This approach is used to find a multivariate parametric description of non-normally distributed variables. It is successfully applied in hydrology (Genest et al., 2007; Renard & Lang, 2007), meteorology and climate science (De Michele, 2003; Schoelzel & Friedrichs, 2008; Vannitsem, 2006; Vrac & Naveau, 2007) and received more attention within the discussion on compound events (Bevacqua et al., 2017; Zhou et al., 2019; Zscheischler & Seneviratne, 2017). The strength of copulas is clearly to obtain a parametric description of the multivariate distribution, although one has to choose the family of the copula used (Schoelzel & Friedrichs, 2008). However, copulas are restricted to low dimensional settings (two or three variables) Mikosch:2006kv, Mikosch:2006hn, which is one reason they are not further used in this thesis.

Hence, it seems there is a need for extreme event detection methods in climate sciences and related studies, which consider a multivariate covariance structure. This method should be able to deal with highly correlated potentially dependent vari-

ables and should be applicable to high dimensional data. Transfer and adaptation of promising methods from other disciplines is further developed in this thesis.

1.2 Objectives and structure of the thesis

The overarching aim of my thesis is to *improve the detection and understanding of climate extremes and their impacts on vegetation by facilitating a broader multivariate perspective which complements previous approaches to detect extreme events*. Therefore, the thesis will provide some background information about anomaly and extreme event detection in chapter 2 and the following main parts.

First, the detection of climate extremes is an inherently multivariate task (Section 1.1), but multivariate approaches to detect climate extremes are rarely used for Earth observations and lacking a common comparison (Chapter 1.1). Therefore, in chapter 3 I adopt and develop generic approaches in order to detect multivariate anomalies of different types in artificial data. The artificial data is built to mimic different properties and types of anomalies in Earth observations. It serves as a test bed to evaluate the capability of a wide selection of multivariate algorithms, preprocessing steps, and feature extraction steps to detect different types of anomalies in the artificial data. The objective of this part is to evaluate which *combination of multivariate anomaly detection algorithm and feature extraction is best suitable for detecting anomalous events*. I select one of the best performing algorithm and feature extraction combinations to detect extreme events in data. Extreme events in the artificial experiment are represented as a few time steps lasting "BaseShift" in the artificial experiment.

Second, in chapter 4 I further develop and apply the selected combination of preprocessing, feature extraction, and multivariate algorithm to detect extreme events in hydrometeorological and biospheric data. The objective is to test the algorithm on Earth observation data and to *evaluate whether a broader multivariate perspective facilitates our understanding of extreme events and their impacts by revealing previously overlooked facets*. Therefore, I focus on the 2010 Russian heat wave, an extreme event that is well-known from the literature. A key result of the study is that an objective detection algorithms which detects multivariate extreme events in any set of variables relative to their normal conditions can bridge the gap between different views on the very same extreme event, i.e. the Russian heatwave 2010 (e.g. a climate or hydrometeorological view and a biospheric view). Applying the algorithm in a rigorous (objective) manner shows that during the Russian heatwave 2010, forest ecosystems in higher latitudes are associated with enhanced primary productivity whereas in contrast southern agricultural systems reduce their productivity tremendously.

Third, in chapter 5 I further built on these results to evaluate whether this kind of contrasting response during the Russian heatwave in 2010 was just a singular case or can be found frequently for extreme droughts and heatwaves, when applying the developed extreme event detection scheme globally. The role of vegetation types in mediating impacts of droughts and heatwaves on productivity is still highly uncertain on a global scale (Section 1.1.3). The more general objective of this study is to *evaluate the importance of different vegetation types on shaping the impact of climate extremes relative to other factors*. To elaborate on this issue further, I focus on relative droughts, heatwaves and combined droughts and heatwaves during growing season as they are typically affecting primary productivity most negatively. The droughts, heatwaves and combined droughts and heatwaves can be detected by the previously developed detection scheme reliably. However, as the generic algorithm detects any kind of extreme event, the selection of droughts and heatwaves takes place after the detection scheme.

In the overarching discussion (Chapter 6), I discuss possible alternative pathways, further potentials of the developed methodology and possible future research directions. I will also return to the objectives of the thesis in the concluding summary (Chapter 7).

Chapter 2

Background

2.1 On the definition of extreme events

There is no common objective definition of extreme events. In the public, extreme events in weather are usually perceived as periods in which the variable of interest (e.g. temperature, precipitation) is not in the range of normal variability which would be expected at this time and region. In a scientific context, Seneviratne et al., 2012 defines an extreme event to be a value near the upper (or lower) range of observed values in a climate or weather variable. More statistically, extreme events can be defined as values that occur in the tails of a probability distribution (Ghil et al., 2011; Kantz, 2005). Common of the above mentioned definitions is that extreme events cannot be defined objectively, as the definitions include a subjective step of defining (i) the reference of normal variability and (ii) a threshold to discriminate between normal variability and the extreme values.

Several approaches of choosing a reference of normal variability exist. One approach particularly common in climate science is to choose a reference of normal variability based the past observations, e.g. often a climatic mean and standard deviation within a period of 30 years (Coumou & Robinson, 2013; Hansen et al., 2012; Huntingford et al., 2013). Especially for few observations (e.g. 30 yearly observations), the estimation of the mean and standard deviation within the reference period leads to an overestimation of extreme events in the out-of-reference period. One can adjust for this systematic bias (Sippel et al., 2015). Another approach is to estimate the mean and standard deviation within the period of interest (Zscheischler et al., 2014a; Zscheischler et al., 2013). This approach is not affected by the above mentioned bias of overestimating extreme events. It rather underestimates the number of extreme events as the mean and standard deviation are strongly influenced by extreme observations. Generally using robust estimates of the mean and standard deviation, like the median and the median absolute deviation is influenced less by outliers or extreme observations (Leys et al., 2013; Rousseeuw & Hubert, 2011). This

thesis builds upon these robust statistical measures. Additionally, a recently developed approach relies on robustness through spatial replicates (Mahecha et al., 2017). The idea is to define the reference of normal variability not solely based on the temporal information of one time series, but to include information of other similar time series (or grid cells, in spatio-temporal gridded data sets) as spatial replicates in the estimation of the reference of normal variability. Similar time series in this context may be time series of similar seasonal patterns, but other features are possible as well. The thesis is build upon and further develops this idea (Chapter 4, 5). Apart from defining a robust reference of normal variability, the latter idea can also be used to define more robust regional thresholds to discriminate between normal variability and extremes events.

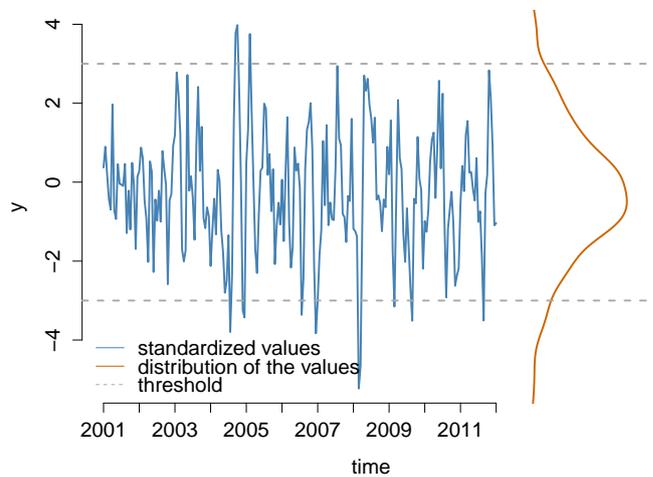


Figure 2.1 – Detecting extreme events with a peak over threshold.

One key choice for detecting extreme events is the choice of the threshold to discriminate between extreme events and normal variability. A threshold is often chosen in the variable of interest, defining extreme events as observations which are above (or below) the defined threshold (Figure 2.1). In climate science, extreme events are often defined as observations exceeding two or three standard deviations (Coumou & Robinson, 2013; Hansen et al., 2012; Huntingford et al., 2013). This approach assumes a gaussian distributions of the underlying variable which may not always be the case. Choosing a certain percentile in the observed values is more appropriate in those cases. Common choices are 1%, 5% or 10% (McPhillips et al., 2018; Seneviratne et al., 2012). In this thesis, a percentile of 5% is chosen to detect extreme events. However, as the results may be sensitive to this key choice, a sensitivity analysis of the results with respect to other choices of the threshold is performed (Chapter 4). Recently, a cross-disciplinary review for defining extreme events concluded that they cannot present a unifying threshold for defining extreme

events. The threshold may be depending on event types, the system under scrutiny, geography, society, and the goal of the study. However, they identified that it is important to discuss the threshold and to explicitly name the used definition of extreme events (McPhillips et al., 2018).

Apart from extreme events in climatological or hydrometeorological variables, extreme events may be defined differently from a biospheric perspective. In the biosphere, extreme events may not only include extremeness in the (climatic) driving variable, but also in the biospheric response variable (e.g. some vegetation index), as a climate extreme may not always be relevant for the biosphere or a biospheric event is not exclusively restricted to extreme conditions in the hydrometeorological drivers. (Smith, 2011) defines extreme events as 'episode or occurrence in which a statistically rare or unusual climatic period alters the ecosystem structure and/or function well outside the bounds of what is considered typical or normal variability'. In this context, a backward assessment of detecting extreme impacts rather than extreme events in the driving variables and to evaluate the conditions under which these impacts occurred is another straightforward approach (Zscheischler et al., 2013). This approach leads to useful insights, although the focus on extreme impacts is questioned in an cross-disciplinary context as it is biased towards negative (or positive) impacts. A standalone backward assessment does not consider similar conditions in the drivers which did not lead to extreme impacts e.g. in more resilient systems (McPhillips et al., 2018). This critique can be addressed if the backward assessment is accompanied by a forward assessment of extremes in the drivers.

Extreme events may not only occur in one variable, but can occur simultaneously or successively in multiple variables (AghaKouchak et al., 2014; Griffin & Anchukaitis, 2015). This is usually discussed under the term 'compound events'. Apart from extreme events in multiple variables, compound events also include conditions in the variables which may not be extreme in the marginal distribution of the variable, but lead to extreme events or impacts when combined (Leonard et al., 2014; Seneviratne et al., 2012). It is currently under debate, whether compound events are events which lead to societal or environmental risk (Zscheischler et al., 2018), or are more generally speaking events for which more than one variable is involved (Buttlar et al., 2018; Zscheichler et al., 2014; Zscheischler & Seneviratne, 2017; Zscheischler et al., 2014c). In this thesis, the latter definition is adapted. More generally speaking, multivariate extreme events are defined here as extreme events which include more than one variable in the detection process. This definition includes but is not exclusively restricted to compound events.

2.2 Multidisciplinary overview on potential multivariate detection concepts

Multivariate extreme event detection is not commonly applied in climate and related disciplines. Used techniques so far are not taking the covariance structure of the data into account, are not suitable for non-linearities or restricted to few variables (Section 1.1.5). Therefore, I review existing techniques which can be adapted and used in climate science, and related studies on fluxes between the atmosphere and biosphere in the following.

2.2.1 Statistical process monitoring

One promising concept for multivariate event detection evolves from industrial process monitoring. There, several sensors monitor the processes taking place within the production chain in time. An alarm is raised as soon as the production process gets 'out of control'. This concept is classically called statistical process control or process monitoring (more recently) (Santos-Fernandez, 2013). Process monitoring in industry is explicitly built to deal with multivariate and highly correlated variables, similar to the case in earth system sciences.

Multivariate process monitoring is known in industry since Harold Hotelling published a multivariate extension of the traditional Shewart's control chart in 1947, which is now known as Hotelling's T^2 (Lowry & Montgomery, 1995). It ought to be one of the most widely used multivariate process monitoring techniques in industry (Bersimis et al., 2007). Hotelling's T^2 is based on the Mahalanobis distance to the mean of the data, thus assumes a multivariate normal distribution (Figure 2.2). Many modifications and extensions to Hotelling's T^2 have been developed. A slight modification is used to detect outliers in hyperspectral images (Bauer & Bauer, 2015; Smetek & Bauer, 2007). One famous and also widely used extension is the multivariate exponentially weighted moving average which includes temporal information through exponential weighting before applying Hotelling's T^2 (Lowry & Woodall, 1992). Another extension is the fast minimum covariance determinant estimator which uses a subset of the observations to robustly estimate the mean and covariance (Rousseeuw & Van Driessen, 1990).

The concept of traditional statistical process control splits the data in two parts: 'in-control' data is used for the first establishment of the algorithms in Phase I. In the second step (Phase II) the data stream is continuously monitored with the parameters and control limits obtained from Phase I (Santos-Fernandez, 2013). The procedure of splitting the data in two parts is similar to using a reference period in climate science. However, one difference is that Phase I data should be based on 'in-control' data, i.e. the estimation of the mean or the covariance matrix are not confounded by

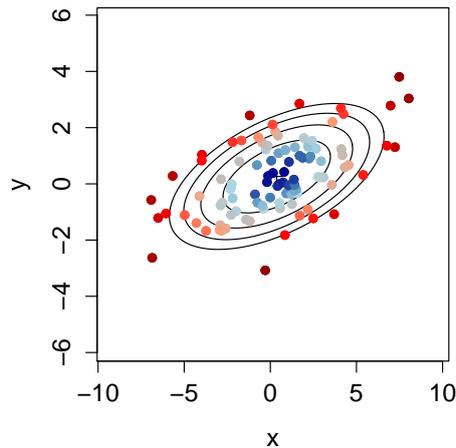


Figure 2.2 – A random sample of multivariate normal distributed data to illustrate Hotelling’s T^2 . Colors denote Hotellings’s T^2 scores, black lines are isolines of similar mahalanobis distsance to the mean. Figure modified after Santos-Fernandez, 2013

outliers or extreme observations like in climate sciences.

2.2.2 Novelty detection

Another concept is called novelty detection. Novelty detection tries to detect data during testing which differs from the data which was available during training (Pimentel et al., 2014). Thus, the concept as such is very well suited to detect instances of data for which a previous model was not constructed for. Common application encompass image processing and classification, healthcare informatics, credit card fraud detection, fault detection in industrial networks, sensor networks, and text mining (see (Pimentel et al., 2014) and references therein).

Conceptually, there is a major difference to the application in earth system sciences for multivariate extreme event detection in this thesis. Multivariate extreme event detection in earth system sciences does not aim at detecting differences between a training data set and a testing data set, because a training data set without extreme events cannot be found with certainty. Thus, a perfect novelty detection algorithm would not detect those constellations of variables during multivariate extreme events which are already contained in the training data. Nevertheless, we evaluate the performance of two different methods developed in the field of novelty detection (Chapter 3). One commonly used method in this context is support vector data description, which also allows for some outliers in the training data (Tax & Duin, 2004). The other more recently developed method in the field of image classification is based on the kernel null foley sammon transform (Bodesheim et al., 2013).

In contrast to the supervised novelty detection which detect differences between training and testing data set, a more conceptually related technique might be the

field of unsupervised anomaly detection.

2.2.3 Anomaly detection

Closely related to novelty detection is the field of anomaly detection. Anomaly detection aims detecting patterns in the data that do not belong to the 'normal' behavior. It is often used interchangeably with the term outlier detection. The main difference to novelty detection is that anomaly detection is also possible in an unsupervised way, i.e. there is not necessarily a 'normal' training data set like in novelty detection (Chandola et al., 2009).

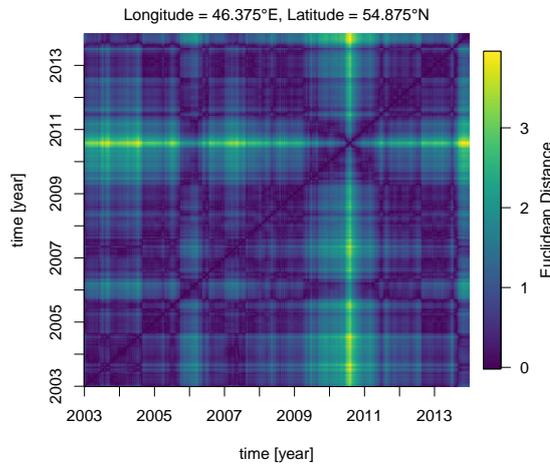


Figure 2.3 – Pairwise distances of observations at different time instances in the three dimensional space of temperature, surface moisture and radiation deviations from the mean seasonal cycle; depicted is one pixel affected by the Russian Heatwave 2010, for more information on variables see Chapter 5.

Many techniques in anomaly detection are based on the idea that the variable space can be condensed by computing distances between different observations (Chandola et al., 2009; Pimentel et al., 2014). These pairwise distances are used to compute an anomaly score (Figure 2.3), e.g. based on the k-nearest neighbor(s) (Harmeling et al., 2006; Ramaswamy et al., 2000), or based on the number of observations falling within a hypersphere (Faranda & Vaienti, 2013; Knorr et al., 2000). Both approaches can be seen as estimates of the inverse density of a given data point (Chandola et al., 2009), i.e. observation with higher distances to their neighbors, or few other observations within a certain hyperball are being considered anomalous or outliers. Thus, another way to obtain an anomaly score is to use density estimates of the data. Following up on this idea, one can directly transform pairwise distances into a measure of similarity or density by centering gaussian kernels around each observation. This technique is known as kernel density estimation or parzen window estimator (Parzen, 1962), and also frequently used for anomaly detection (Harmeling et al., 2006; Pimentel et al., 2014; Yeung & Chow, 2002).

In the particular case of time series, anomalous observations often form block-like structures in the matrix of pairwise distances, i.e. rectangles in the matrix which differ between 'anomalous' parts and 'normal' parts of the data (see e.g. 2.3). This finding at the start of this thesis led to the parallel development of an algorithm in computer science, which is based on measuring the kullback leibler divergence between two data distributions (Barz et al., 2017; Rodner et al., 2016).

Apart from the above mentioned techniques to detect anomalies, outliers or novelties, many other techniques and specific modifications of the above mentioned techniques exist (for an overview see (Chandola et al., 2009; Pimentel et al., 2014)). A comparison of different detection algorithms as done in Chapter 3 can never be 'complete' by definition as new algorithms are being constantly developed. Also frequently used for anomaly detection are classification based techniques (e.g. k-means clustering) and parametric approaches (e.g. gaussian mixture models) (Chandola et al., 2009; Pimentel et al., 2014). Both are not used in this thesis as they require a parameter for the number of clusters or the number of gaussian mixtures, respectively, which is not possible to obtain in a generic automatic way without training or without a reference period containing solely 'normal' data. Thus this thesis is built predominantly on unsupervised and non-parametric techniques to detect anomalies as outlined before.

Chapter 3

Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques

This chapter is originally published in

Flach, M., Gans, F., Brenning, A., Denzler, J., Reichstein, M., Rodner, E., Bathiany, S., Bodesheim, P., Guaniche, Y., Sippel, S., & Mahecha, M. D. (2017): Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques, Earth Syst. Dynam., 8, 677–696, doi: 10.5194/esd-8-677-2017.



Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques

Milan Flach¹, Fabian Gans¹, Alexander Brenning^{2,4}, Joachim Denzler^{3,4,5}, Markus Reichstein^{1,4,5}, Erik Rodner^{3,4}, Sebastian Bathiany⁶, Paul Bodesheim¹, Yanira Guancho^{3,4}, Sebastian Sippel¹, and Miguel D. Mahecha^{1,4,5}

¹Max Planck Institute for Biogeochemistry, Department Biogeochemical Integration, P.O. Box 10 01 64, 07701 Jena, Germany

²Friedrich Schiller University Jena, Department of Geography, Jena, Germany

³Friedrich Schiller University of Jena, Department of Mathematics and Computer Sciences, Computer Vision Group, Jena, Germany

⁴Michael Stifel Center Jena for Data-driven and Simulation Science, Jena, Germany

⁵German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany

⁶Wageningen University, Department of Environmental Sciences, Wageningen, the Netherlands

Correspondence to: Milan Flach (milan.flach@bgc-jena.mpg.de)

Received: 22 October 2016 – Discussion started: 2 November 2016

Revised: 15 June 2017 – Accepted: 2 July 2017 – Published: 8 August 2017

Abstract. Today, many processes at the Earth’s surface are constantly monitored by multiple data streams. These observations have become central to advancing our understanding of vegetation dynamics in response to climate or land use change. Another set of important applications is monitoring effects of extreme climatic events, other disturbances such as fires, or abrupt land transitions. One important methodological question is how to reliably detect anomalies in an automated and generic way within multivariate data streams, which typically vary seasonally and are interconnected across variables. Although many algorithms have been proposed for detecting anomalies in multivariate data, only a few have been investigated in the context of Earth system science applications. In this study, we systematically combine and compare feature extraction and anomaly detection algorithms for detecting anomalous events. Our aim is to identify suitable workflows for automatically detecting anomalous patterns in multivariate Earth system data streams. We rely on artificial data that mimic typical properties and anomalies in multivariate spatiotemporal Earth observations like sudden changes in basic characteristics of time series such as the sample mean, the variance, changes in the cycle amplitude, and trends. This artificial experiment is needed as there is no “gold standard” for the identification of anomalies in real Earth observations. Our results show that a well-chosen feature extraction step (e.g., subtracting seasonal cycles, or dimensionality reduction) is more important than the choice of a particular anomaly detection algorithm. Nevertheless, we identify three detection algorithms (k -nearest neighbors mean distance, kernel density estimation, a recurrence approach) and their combinations (ensembles) that outperform other multivariate approaches as well as univariate extreme-event detection methods. Our results therefore provide an effective workflow to automatically detect anomalies in Earth system science data.

1 Introduction

The Earth system can be conceptualized as a system of highly interconnected subsystems (e.g., atmosphere, biosphere, hydrosphere, lithosphere). Each of these subsystems can be monitored and characterized by multiple variables. Technological progress over the past decades has led to a boost in satellite technologies (Pfeifer et al., 2011; Nagendra et al., 2013) as well as ground station development and routine monitoring (Baldocchi et al., 2001; Dorigo et al., 2011; Ciaia et al., 2014). Additionally, advanced computational methods efficiently integrate remote sensing and in situ information to routinely derive novel data products (e.g., Beer et al., 2010; Jung et al., 2011; Tramontana et al., 2016). One key scientific challenge is co-interpreting these multiple views of the Earth system, in particular to address the impacts of changes in the climate system, the land use system, and other transformations.

Of particular importance is the analysis of extreme events like droughts, fires, heat waves, or floods, which are expected to change in a future climate (Kharin et al., 2013). One matter of concern is changes in hydrometeorological extremes that may translate into anomalies in vegetation dynamics, or extremes in vegetation dynamics that might result from slight changes in climatological conditions or human intervention and that can have severe consequences for vegetation and the carbon cycle (Easterling et al., 2000; Meehl and Tebaldi, 2004; Seneviratne et al., 2012; Reichstein et al., 2013). Apart from natural events, one also aims to detect events that are a direct consequence of human interference, e.g., detecting deforestation activities is required to assess the compliance with laws or agreements on forest conservation and climate change.

The flood of observational data is accompanied by a similar increase in data from Earth system models (Overpeck et al., 2011). As large numbers of data are difficult to handle and to translate into quantities of human interest, it can be easy to overlook events of particular importance. For example, using a simple semiautomatic detection scheme to identify abrupt climate shifts in simulations of future climate, Drijfhout et al. (2015) found a number of abrupt events that had previously been overlooked in simulations.

In observations, anomalous events are often detected using extreme event detection methods suitable for univariate data streams (e.g., Alexander et al., 2006; Rahmstorf and Coumou, 2011; Zhou et al., 2011; Donat et al., 2013; Lehmann et al., 2015). Univariate extreme event detection can also be used to infer knowledge about underlying drivers of extremes (Zscheischler et al., 2014a); it is particularly valid when the variable of interest is either of specific importance or integrates a wide array of relevant processes. However, some information might only be inferred when taking the multivariate combination of several data streams into account (Vicente-Serrano et al., 2010; Seneviratne et al., 2012; Fischer, 2013; Zscheischler et al., 2015). For instance,

a significant fraction of events of carbon extremes in Europe is not associated with univariate climate extremes (Zscheischler et al., 2014b). Earth observations (EOs) are multivariate and naturally characterized by strong dependencies and correlations in space, time, and across dimensions (Leonard et al., 2013). We assume that any suitable anomaly detection algorithm needs to consider these data properties. By considering multivariate constellations for anomaly detection, it might become possible to gain further information, i.e., about anomalies that cannot be detected with univariate extreme event detection methods (for a review of approaches see, e.g., Ghil et al., 2011).

Multivariate approaches in geoscience make use of anomalies occurring simultaneously in multiple data streams, often referred to as coincidences or co-exceedances (e.g., Donges et al., 2011b; Rammig et al., 2015; Zscheischler et al., 2015; Donges et al., 2016; Guaniche et al., 2016; Siegmund et al., 2016). An alternative is the copula approach introduced to the field by Schoelzel and Friedrichs (2008) and Durante and Salvadori (2010). However, the copula approach is limited so far to two or three simultaneous data streams (Mikosch, 2006), which makes it unsuitable for high-dimensional data as used in this paper.

Interestingly, there are multiple industrial applications that likewise require anomaly detection. In this context, anomaly detection has become a standard procedure in the wake of Harold Hotelling's publication of the T^2 control chart in 1947 (Hotelling, 1947; Lowry and Woodall, 1992). Consider, for instance, several sensors observing some industrial production chain. These (potentially correlated) sensor data streams can be monitored with a statistical process control (SPC) algorithm (Lim et al., 2014; Ge et al., 2013; Lowry and Montgomery, 1995). The basic idea is to raise an alarm as soon as an anomaly according to the SPC is detected, meaning that the production chain is out of control. Despite the obvious analogy, the ideas of SPC are largely unknown in the geoscience community to the best of our knowledge. Conceptually, the industrial application is equivalent to the idea of monitoring environmental variables. However, data differ. EOs exhibit strong (potentially nonlinear) dependencies among the variables; seasonal cycles are typically present in both temporal mean and variance. The variables may also encode dynamic feedbacks and abrupt transitions. EOs are possibly more strongly corrupted by noise compared to industrial applications. Furthermore, industrial applications are typically less affected by low-frequency variability than EOs. The most problematic aspect when considering SPC concepts in Earth system sciences is, however, defining states of normality.

The objective of this study is to provide an overview and comparison of anomaly detection algorithms and their combination with feature extraction techniques for identifying multivariate anomalies in EOs. Spatiotemporal EOs are therefore stored in the Earth system data cube, which is a four-dimensional array of latitudes, longitudes, time, and dif-

ferent measurement variables. To detect multivariate anomalies in EOs, we define an anomaly to be any consecutive spatiotemporal part of the data cube that differs with respect to the mean, the variance, the amplitude of the seasonal cycle, or trends from the normal rest of the data cube. We adapt algorithms from SPC and novelty detection. The study is structured as follows: first, we create a series of artificial Earth system data cubes that try to mimic a series of real world features (in terms of multiple variables, seasonal cycles, and correlation structure, etc.). We are aware that these artificial data cubes are not real simulations of Earth system data cubes. However, relying on artificial data in this paper is motivated by the fact that a meaningful quantitative evaluation of unsupervised anomaly detection algorithms and feature extraction techniques in real Earth observation data is difficult due to the lack of ground-truth data (Zimek et al., 2012). Second, we use these artificial data to evaluate the capability of different algorithms to detect multivariate anomalous events, including compound events (e.g., events in which none of the single variables are extreme, but their joint distribution is anomalous and might lead to an extreme impact) (Seneviratne et al., 2012; Leonard et al., 2013). Specifically, we evaluate the performance of the algorithms in detecting multivariate changes in the mean (comparable to an extreme event), the amplitude of the annual cycle, the variance, and the onset of trends. Using the artificial dataset as a test bed we apply various feature extraction schemes (Sect. 3.1), several detection algorithms (Sect. 3.2), and combinations of detection algorithms (ensembles, Sect. 3.4) to compare their performance in identifying anomalous events (Sect. 3.3). From this comparison we select suitable combinations of feature extraction (Sect. 4.1) and a few algorithms (Sect. 4.2) as well as ensembles of algorithms (Sect. 4.3) as the best ones applicable to EOs, including suggestions for their specific usage (Sect. 5).

2 Experimental setup

2.1 Generation principle of the artificial data

Ground truth for detecting anomalies in multivariate data is rare, in particular for detecting anomalies in real EOs. Thus, we generate artificial data that represent common properties of EOs, including anomalies. In particular, we focus on the existence of seasonality, correlations among variables, and non-Gaussian distributions. Data generation assumes that each subsystem of the Earth has uncorrelated intrinsic properties, i.e., it is dominated by a few independent components. Consequently, generating these independent components (which cannot directly be monitored) is the first step. We then derive variables that contain elements of all independent components and correspond to the observable measurements as a set of correlated variables (Fig. 1).

More precisely, as a basic version we create three independent components for the artificial data, each consisting of a signal (Gaussian, $SD = 1.0$) that includes seasonality in

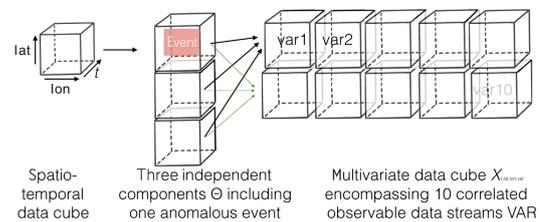


Figure 1. Combination of three independent component cubes to derive 10 correlated variables X as observable measurements. The anomalous event is propagated into some variables of X .

some cases (Sect. 2.2). Anomalous events are induced in one of the independent components for which we track the exact spatiotemporal location. These three independent components are then weighted with randomly generated linear (or nonlinear, Sect. 2.3) weights to create a set of 10 correlated variables, which represent the artificial data cube, i.e., try to mimic observable measurements. We add some additional measurement noise (Gaussian, $SD = 0.3$) to the data cube. For more technical details of this generation scheme we refer the reader to the Appendix A.

Our standard data cube $X_{t_i,j,\text{lat},\text{lon},\text{var}}$ encompasses $t_i, j = 1, \dots, T$ time steps ($T = 300$) corresponding to a 6.5-year time series of satellite images in 8-day intervals, $\text{lat} = 1, \dots, \text{LAT}$ latitudes ($\text{LAT} = 50$), $\text{lon} = 1, \dots, \text{LON}$ longitudes ($\text{LON} = 50$), and $\text{var} = 1, \dots, \text{VAR}$ data streams, or variables ($\text{VAR} = 10$).

2.2 Generating anomalous events

Anomalous events are introduced in the independent components only and then propagated from the independent component to some of the variables in the data cube with random weights. The anomalies are contiguous in space and time. The center of the anomaly is assigned randomly. The challenge is to detect the propagated anomaly through the unsupervised algorithms, i.e., without using the information about the spatiotemporal location of the anomaly. With this data cube generation scheme, we can generate anomalies by controlling the type of the anomalous event (event type), the magnitude of the anomalous event, and the spatiotemporal location.

We create four data cubes using the following temporary event types:

- a shift in the baseline, i.e., shift of the running mean of a time series (BaseShift) (Fig. 2a). This event type is closely related to extremes in real world EOs.
- an onset of a trend in the time series (TrendOnset) (Fig. 2b).
- a change in the amplitude of the mean seasonal cycle of a time series (MSCChange) (Fig. 2c), which might

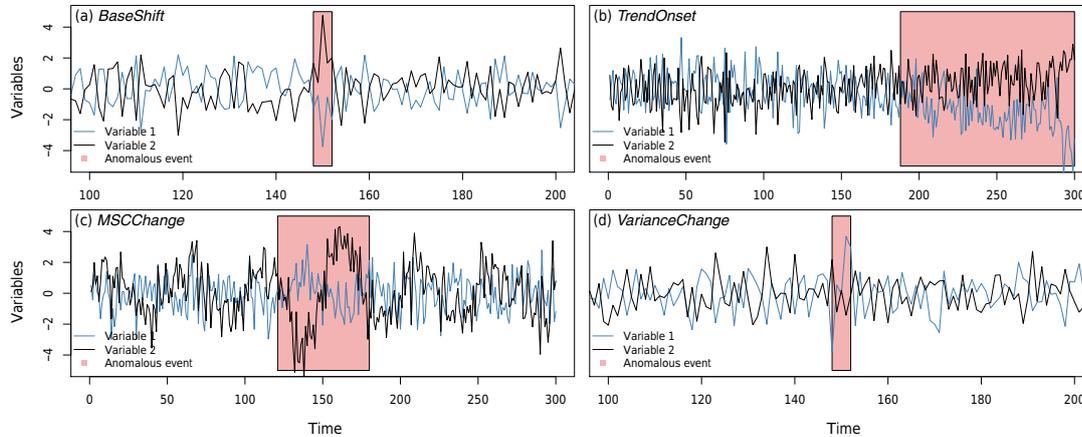


Figure 2. Visualization of the four different event types (a–d) with two variables along time $t_{i,j} = 1, \dots, T$ ($T = 300$). The two variables contain an anomalous event (red shape) that is propagated through the underlying independent components with randomly drawn weights within the generation process of the variables. For illustration purposes two variables are shown for one specific magnitude of the anomaly. The artificial data farm encompasses 10 variables and anomalous events of 20 different magnitudes ranging from very subtle to exceptionally high changes.

happen in the real world carbon cycle as a response to combined drought–heat waves (Ciais et al., 2005).

- d. a change in the variance of the time series (VarianceChange) (Fig. 2d), e.g., in temperature (Huntingford et al., 2013).

2.3 Additional data properties

Apart from the basic data cubes, we want to test the influence of certain data properties on the anomaly detection algorithm. In order to do so, we create data cubes, each with one added data property, i.e., we increase the number of independent components (MoreIndepComponents) or use a squared dependency among independent components (NonLinearDep) instead of a linear one. Furthermore, typical EO variables are often driven by extrinsic forcings, i.e., the Earth’s solar system orbit, rotation, and axis tilt. Thus, we add a seasonal cycle modifying the signal (SeasonalCycle). In a global context, the mean is rarely constant; we therefore introduce a linear latitudinal trend into the baseline (LatitudinalGradient). In the basic case, the signal of our independent components follows a Gaussian distribution. In the more complicated versions, we also implement alternative scenarios with Laplacian (doubly exponentially) distributed signals (LaplacianNoise) and signals that exhibit spatiotemporal correlation with red noise (CorrelatedNoise). Signal-to-noise ratio is 0.3 in the basic version, one additional data property increases the signal-to-noise ratio to 1.0 (NoiseIncrease). Also, the shape and duration of anomalous events differ. We double (LongExtremes) or reduce the temporal duration of the

anomalous events (ShortExtremes) and change the spatial shape from rectangular to randomly affecting neighboring grid cells (RandomWalkExtreme).

2.4 Experiment design

Each data cube with a specific type of the event is generated 20 times, each time with a different magnitude of the anomalous event (Appendix A). We introduce 10 spatially contiguous anomalous events into the independent components, with a spatial extent of 20 latitude and longitude steps each. Each event has a temporal extent of five time steps (which would be equivalent to 40 consecutive anomalous days in a 6.5-year record). Our total number of anomalies equals about 3 % of the total data cube, which we consider to be a realistic scenario (comparable to Zscheischler et al., 2014a), for example. Some latitudes and longitudes do not exhibit any anomaly by design. The algorithms (Sect. 3.2) are expected to be able to deal with parts of the data cube that do not exhibit anomalies at all, as this is also very likely to happen for applications in real EOs.

Our experiment comprises 36 different event-type combinations of data properties, each repeated 20 times with varying event magnitudes (Appendix A). The entire set of artificial data cubes consists of 720 data cubes, corresponding to ≈ 87 GB of data¹.

¹Code to reproduce the data farm is provided in the Data Availability section.

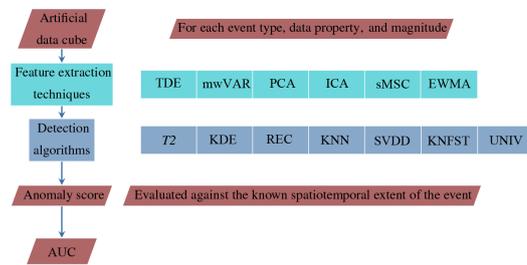


Figure 3. Data processing for detecting multivariate anomalies. We extract relevant features from each artificial data cube before applying the detection algorithms. The detection algorithms output some anomaly score, which we evaluate against the known extent of the event using the area under the curve (AUC). Feature extraction elements on the right-hand side are understood as options and can be combined with each other.

3 Workflows to detect anomalies

The idea of this study is to elaborate workflows that contain both data preprocessing via feature extraction and algorithms for the detection of anomalous events (Fig. 3). In the following we introduce these two elements separately and explain the performance evaluation strategy afterwards.

3.1 Feature extraction

Feature extraction is a process to derive information from the data and condense it into nonredundant characteristic patterns. This may facilitate data interpretation (van der Maaten, 2009). In our study the aim is to maximize the detection of anomalous events by providing relevant features. Feature extraction is often an element of data preprocessing. A very simple form of feature extraction could be to subtract the mean seasonal cycle. We consider the anomaly time series to be the extracted feature in this case. Here, we concentrated mainly on feature extraction methods that are used in the context of classical multivariate SPC (Lowry and Montgomery, 1995), data-based process monitoring in industry (Ge et al., 2013), and univariate extreme event detection. The following feature extraction methods are used in this study:

Subtracting the median seasonal cycle (sMSC) is one way to deseasonalize time series. Deseasonalization may be instrumental in detecting anomalous events across different seasons. The remaining part of the time series is often referred to as anomalies in the climatological sense. These anomalies are used here as an input feature. Please note, that the climatological anomalies are only the difference between the mean behavior and thus are not to be mixed up with anomalies (strange or rare regions in the data, closely related to extreme events) as detected through the (multivariate) anomaly detection algorithms (Sect. 3.2).

Computing the moving window variance (mwVAR) is a popular technique for detecting trends in the variance in univariate time series (e.g., Huntingford et al., 2013). We choose a window size of 10 and compute the variance in the running window along the time series of each variable. We use the estimates of the mwVAR time series as a feature to detect multivariate anomalies in the variance.

Time delay embedding (TDE) increases the feature vector Y_t with time-delayed vectors ($Y_t = (X_t - 0\tau, X_t - 1\tau, X_t - (m - 1)\tau)$) to include temporal context information. In the univariate case, this approach ideally creates an image of the attractor of a dynamical system (Takens, 1980). In high-dimensional multivariate data applications it is used to include information of the dynamics in the feature vector (e.g., Koçak et al., 2004; Ge et al., 2013; Smets et al., 2009). Critical hyperparameters are the time delay τ and the number of dimensions m . We fix m to 3 (corresponding to the number of independent components within the data farm creation) and τ to 6, which is a compromise between the typical choice of the first zero crossing of the temporal autocorrelation function or the first local minimum of the mutual information (Webber Jr. and Marwan, 2015) (here: 11.5 corresponding to one-quarter of the annual cycle with 46 time steps) and an accurate temporal detection (requires small τ).

Principal component analysis (PCA) is a data rotation, used to find an orthogonal (uncorrelated) subspace of the data of $n_{PC} \leq \text{VAR}$ variables (Von Storch and Zwiers, 2001). We choose n_{PC} such that at least 95 % of the variance in the original data cube is explained. By assuming a homogeneous covariance structure within the entire data cube, we perform the PCA globally, i.e., with the same rotation matrix for all grid cells. The combination of TDE and PCA is sometimes referred to as dynamic PCA when considering subsequent lags in the time series (Lee et al., 2004).

Independent component analysis (ICA) can be regarded as a nonlinear alternative to PCA; it has become a standard technique of data-based process monitoring. We use one ICA variant that tries to separate different sources of data by maximizing the negentropy, a measure of non-Gaussianity of the data (Hyvärinen and Oja, 2000)². We apply ICA globally to each data cube. The hyperparameter is the number of independent components (sources). We choose the number of independent components to be equal to n_{PC} (see PCA) for consistency reasons (Majeed and Avison, 2014).

Exponentially weighted moving average (EWMA) is one way of reducing the noise of the time series and taking temporal information into account. It is common in the context of classical multivariate SPC to detect only “significant” outliers (Lowry and Woodall, 1992). The multivariate feature time series Y is computed recursively as

$$Y_{t_i} = \lambda X_{t_i} + (1 - \lambda)Y_{t_i-1}. \quad (1)$$

²We use the fastICA algorithm implemented in the Julia package `MultivariateStats.jl` (<https://github.com/JuliaStats/MultivariateStats.jl>).

The hyperparameter λ determines the degree of exponential weighting between 1 (no weighting) and zero (common choice $0.1 \leq \lambda \leq 0.3$; Santos-Fernández, 2013). We stay in this range with $\lambda = 0.15$.

There is of course a multitude of alternative approaches available in the literature, but we focus on the previously summarized ones as they are widely used and efficiently implemented. Furthermore, different feature extraction methods can also be combined (Fig. 3). As the number of possible combinations is considerably large, we focus here on dimensionality reduction techniques (ICA, PCA) combined with some EWMA to reduce the noise level afterwards. Depending on the event type and data properties, additionally removing seasonality (sMSC) or including the variance mVAR seems to be straightforward. Information about the dynamics (TDE) can be included before applying dimensionality reduction techniques to keep the dimensionality of the system as low as possible. In the following, combinations are noted in the order in which they were applied (e.g., PCA_EWMA means first applying PCA, then applying EWMA to the PCA features). In some cases this might lead to non-commutative combinations, especially for nonlinear feature extraction techniques (ICA, TDE).

3.2 Anomaly detection algorithms

We use several detection algorithms that we implemented in the Julia package `MultivariateAnomalies` (<https://github.com/milanflach/MultivariateAnomalies.jl>). Some anomaly detection algorithms require the estimation of parameters (details are given below for each algorithm separately). In that case we fix the model parameters for the entire data cube. We estimate model parameters (σ , ε , Q , μ ; see below) and train the models themselves (support vector data description, kernel Null Foley–Sammon Transform, KNFST; see below) based on a random subsample of 5000 data points obtained from the entire data cube. To account for variability in the model parameter estimation, we resample three times. More resampling is not affordable due to high computational costs of processing the large number of data cubes. However, very little random variability is observed with this sample size for the best algorithms. Thus, we consider a resampling of three times to be sufficient for a first attempt to account for variability in the parameterization. The following algorithms are investigated for anomaly detection.

Univariate approach (UNIV) is a simple approach to define extremes in univariate data by identifying all points above (or below) a certain quantile. This so-called “peak-over-threshold” approach can be transferred to deal with multiple univariate data streams. In this case, one would consider a data point to be extreme if one or several of the univariate variables are above (or below) a certain quantile threshold of the marginal distributions of each single variable. (here: globally) (e.g., Ledford and Tawn, 1996; Bae et al., 2003; Donges et al., 2016). Applications of the so-

called co-occurrence or coincidence analysis can be found in Donges et al. (2011b), Rammig et al. (2015), Zscheischler et al. (2015), Guanche et al. (2016), and Siegmund et al. (2016). For comparing the algorithms, we are interested in the information that at least one variable is above a certain threshold. We compute this information for different thresholds (in terms of quantiles of the marginal distributions between 0.0 and 1.0, accuracy 0.01) to get a score, i.e., a ranking of the extremeness of the data points.

Hotelling’s T^2 ($T2$) computes the squared Mahalanobis distance of each data point X_t to its temporal mean μ weighted with the covariance matrix Q (Hotelling, 1947):

$$(X_t - \mu)' Q^{-1} (X_t - \mu). \quad (2)$$

A crucial prerequisite is the estimation of the covariance matrix Q , which is estimated from the random subsample of 5000 data points. Combining the feature extraction EWMA with $T2$ equals the traditional multivariate exponential weighted moving average (Lowry and Woodall, 1992; Lowry and Montgomery, 1995).

Apart from computing weighted distances to the mean (like $T2$), it is also possible to compute pairwise Euclidean distances in variable space $d(X_{t_i}, X_{t_j})$ between vectors X_{t_i} and X_{t_j} of time steps t_i and t_j for all possible time steps $t_i, t_j = 1 \dots T$. The resulting matrix D with $D_{ij} = d(X_{t_i}, X_{t_j})$ is often referred to as distance matrix or dissimilarity matrix. For real world data, variables have to be standardized with care before computing the distance matrix (Sect. 5). However, in the artificial data used the variables are already comparable by construction; thus, standardization is not needed. The following algorithms are based on pairwise distances.

K -nearest neighbors (KNNs) can be used for anomaly detection by considering the mean distance to the k -nearest neighbors (k -nearest neighbors Gamma, KNN-Gamma) and the length of the mean of the vectors pointing from X_{t_i} to its k -nearest neighbors (k -nearest neighbors Delta, KNN-Delta) (Harmeling et al., 2006; Ramaswamy et al., 2000). With the latter approach KNN-Delta also considers the direction of the neighbors, i.e., has higher values in case its nearest neighbors are pointing in one direction, which is in contrast to the directionless distance of KNN-Gamma. We fix the hyperparameter k at 10 after carefully trying different choices for k without seeing major effects on preliminary results. Furthermore, we exclude trivial temporal autocorrelations by excluding five neighboring time steps ($abs(t_i - t_j) \geq 5$) to also be nearest neighbors.

Recurrences (REC). Within the framework of the theory of nonlinear dynamical systems, each state of a dynamical system will revisit a particular region in its phase space, if waiting for a sufficiently long time (Poincaré, 1890). These dynamics can be visualized in the recurrence plot and are quantified with several metrics usually referred to as recurrence quantification analysis (Marwan et al., 2007). It seems straightforward to use the concept of recurrence analysis to detect states in a dynamical system that are considered to be

rare or unusual. Faranda and Vaienti (2013) used the concept of recurrences and combined it with extreme value theory. We want to use a more general approach without binning the time series. We count the number of observations ζ falling into a certain ε ball in a system of multiple variables, condensed by their distance $d(\mathbf{X}_{t_i}, \mathbf{X}_{t_j})$:

$$\zeta(\mathbf{X}_{t_i}) = \sum_{j=1}^T \Phi(\varepsilon - d(\mathbf{X}_{t_i}, \mathbf{X}_{t_j})). \quad (3)$$

$\Phi(z)$ is the Heaviside function, coding the distances to binary values ($\Phi(z) = 0$ if $z < 0$, $\Phi(z) = 1$ otherwise). An ε hyperball containing only few recurrent observations is considered to be rare in comparison to the majority of ζ values. We compute $1 - \zeta \cdot T^{-1}$ to get anomaly scores, which are more likely to be an anomaly for high score values. $\zeta \cdot T^{-1}$ is known as local recurrence rate or degree density in recurrence analysis (Marwan et al., 2007; Donner et al., 2010) (Donges et al., 2012). ε is the crucial hyperparameter, defining the radius of the ball. Typical choices of ε in recurrence analysis are quantiles of the distribution of elements of the distance matrix, e.g., 5 or 10 % (Donges et al., 2011a; Flach et al., 2016). As we are not interested in small-scale variations in REC, but more in major anomalies we estimate ε as median of the distance matrices in the random subsample. This choice turned out to be the optimal choice (in terms of maximizing the area under the curve, AUC; Sect. 3.3) for ε in a small simulation, varying the thresholds between the 5 and 95 % quantiles of the element of the distance matrix (Supplement Fig. S1). We exclude five neighboring time steps to be counted as recurrences (similar to KNN). KNN has similarities to REC, as one could also choose a data-adaptive k such that $\zeta = k$.

The distance matrix D can be transformed into a kernel matrix $K = \exp(-0.5 \cdot D \cdot \sigma^{-2})$, i.e., by computing pairwise dissimilarities using Gaussian kernels centered on each data point.

Kernel density estimation (KDE) is a standard technique for estimating densities based on column means of the kernel matrix K (Parzen, 1962). The bandwidth σ of the kernel is a hyperparameter. We estimate σ by using the median of the temporal distance matrix on the random subsample, which is a common choice (Schölkopf and Smola, 2001; Schölkopf et al., 2015).

Support vector data description (SVDD) models the distribution of the training data with an enclosing hypersphere in a high-dimensional kernel feature space (Tax and Duin, 2004). As usual a kernel matrix of the random subsample is used for training. Although being a rather simple data description, a hypersphere in the kernel feature space can result in complex nonlinear decision boundaries in the original space of predictor variables if a nonlinear kernel function is used. In addition to the σ hyperparameter of the kernel function (see KDE), the SVDD approach has a parameter called outlier ratio ν (fixed to 0.2). The outlier ratio ν controls the number of training samples that can be located outside of the hyper-

sphere to prevent overfitting. As anomaly score for testing, its distance to the center of the hypersphere in the kernel feature space is computed. Testing requires pairwise similarities between test and training samples. For performance reasons in terms of computation time, we used the implementation by (Chang and Lin, 2013) of the one-class support vector machine (Schölkopf et al., 2001), which is an alternative formulation that leads to identical data descriptions as SVDD in our setup.

kernel Null Foley–Sammon Transform (KNFST) maps the training data into a so-called null space, in which the training samples have zero variance, i.e., all training samples are mapped to the same point called the target value (Bodesheim et al., 2013). Nonlinearity is incorporated by using a kernel matrix containing pairwise similarities of the training samples (training on the random subsample as for SVDD). Since all training samples are represented by a single target value in the one-dimensional null space, the anomaly score of a test sample is the absolute difference between its projection in the null space and this target value. The projection of the test sample requires pairwise similarities to the training samples. Compared to SVDD no parameters need to be tuned except for σ of the kernel function that is fixed to the same values for all kernel methods.

3.3 Ranking of the workflows

Given the large number of potential combinations of feature extraction and anomaly detection algorithms, we need an objective criterion to compare the performances of the numerous possible workflows. We use the area under the receiver operator characteristics curve (AUC) as our measure of detection skill for a specific event type (Fawcett, 2006). The AUC is based on the fraction of events that are correctly detected (true positives) and the fraction of detections among all non-events (false positives), for all possible decision thresholds that could be applied to scores produced by the algorithms. AUC values of 0.5 would be achieved by random detection, and values below 0.5 indicate that a lower score is more likely assigned to (true) anomalies than to non-anomalies.

For each data cube with a given event magnitude and event type we compute the AUC for each data property, feature extraction, and algorithm combination. This leads to an entire catalogue of possible combinations, namely 1.27×10^5 (4 event types, 20 event magnitudes, 11 data properties, 18 feature extraction combinations, 8 algorithms). The number of combinations strongly requires simplification to infer knowledge about which combination is advisable to use. Hence, we focus on events of magnitudes typically detected in real world data i.e., deviations from the mean (extremes) larger than 2 SD (e.g., temperature extremes in Hansen et al., 2012), a relative increase or decrease in the mean annual cycle amplitude of 25 % (which might happen in the carbon cycle after combined drought and heat waves (Ciais et al., 2005) or in

the Arctic due to abrupt sea ice losses (Bintanja and van der Linden, 2013; Bathiany et al., 2016), for example) or an increase in the signal variance of 25 % (e.g., in temperature; Huntingford et al., 2013).

One way of summarizing the results of such a large number of combinations is treating the AUC values as the outcomes of an experiment in which the different design decisions (e.g., feature extraction techniques, anomaly detection algorithms) are the experimental factors. As a control treatment we introduce the simplest possible approach to detecting the anomaly: UNIV approach on the selected event type, without any further data properties (e.g., short extremes or increased measurement noise) on the event type and without prior feature extraction. In order to assess the (averaged) effect of each experimental factor, we fit a linear mixed-effect model (Pinheiro et al., 2016) to the AUC data (fixed effects: data properties, feature extraction, anomaly detection algorithms; random effect: magnitude of the event). This model's coefficients express the overall effect of a factor level with respect to the control while averaging over all other experimental factors. They are considered to be significant for $p < 0.01$.

Additionally, we compute the resampling variation in parameter estimation of the anomaly detection algorithms (RVP) as mean difference of the maximum AUC and minimum AUC for each resampling $i = 1 \dots 3$ (Sect. 3.2).

$$\text{RVP}_{\text{algorithm}} = \text{mean}(\max(\text{AUC}_{\text{comp, feat, magn, event}, i}) - \min(\text{AUC}_{\text{comp, feat, magn, event}, i})) \quad (4)$$

3.4 Ensembles of anomaly detection algorithms

Summarizing the output of several anomaly detection algorithms is one way to create more robust results (Thompson, 1977). For better comparability of the algorithms' outputs, we rank them by computing the percentiles of the algorithm scores. These are then aggregated into ensemble scores by computing the minimum (consensus voting), the mean (balanced voting), or the maximum (risky voting) of the scores of selected well-performing algorithms (e.g., Aggarwal, 2012; Zimek et al., 2013).

4 Results and discussion

In the following, we present the performance of the workflows in subsections corresponding to feature extraction techniques (Sect. 4.1), anomaly detection algorithms (Sect. 4.2), and ensembles of detection algorithms (Sect. 4.3). Specifically, we present the AUC difference to the UNIV control, i.e., the output of the linear mixed-effect model on the experimental factors feature extraction and detection algorithm (Fig. 4). The corresponding tables present the estimates as well as the RVP (Tables 1, 2). Apart from the model the full range of AUC values with respect to different event magni-

tudes, data properties, and event types is presented in Appendix B, Fig. B1.

4.1 Feature extraction techniques

Feature extraction techniques are often more important than the detection algorithm itself (Fig. 4). However, we find that choosing a suitable feature extraction technique largely depends on the event type of interest. Therefore, the feature extraction techniques are presented for different event types separately.

BaseShift Shifts in the baseline are simulated to mimic extreme events. Increasing the magnitude (in terms of standard deviations) of a BaseShift makes it easier to detect the event (Fig. B1). Dimensionality reduction (via PCA or ICA) is a crucial feature extraction technique step as it derives meaningful uncorrelated subsets of the data (Fig. 4a). The combination of dimensionality reduction with some temporal smoothing (EWMA) does not exhibit better overall performance (Fig. 4a) as it fails for ShortExtremes due to oversmoothing. Nevertheless, EWMA can improve the detection rate for special cases, i.e., long events (LongExtremes) and high signal-to-noise ratios (NoiseIncrease) (Fig. B1).

TrendOnset Results look very similar to those of BaseShift, except that temporal smoothing with EWMA has a stronger positive effect than for BaseShift. This may be related to the fact that events for TrendOnset are longer than those for BaseShift. Since the algorithms used in this work are not specifically designed to detect the onset of linear trends, we speculate that their capability to detect such anomalies may be related to their ability to detect base shifts. While algorithms specifically designed to detect changes in trends (e.g., Forkel et al., 2013) were not included in our work due to our focus on more generic types of anomalies, such specialized algorithms may perform better for this particular class of anomaly.

MSCChange In the detection of MSCChange, most feature extraction algorithms showed some skill in the detection of an amplitude increase, while only a subset of these also succeeded in detecting decreases in amplitude (Fig. B1). We focus on the latter ones, which have one step in common: they subtract the median seasonal cycle before applying the detection algorithm (sMSC) (Fig. 4c). In line with the results for TrendOnset and BaseShift, temporal smoothing in combination with dimensionality reduction improves detection by a large margin (PCA_sMSC_EWMA). Furthermore, accounting for temporal dynamics with a TDE is even more suitable (TDE_PCA_sMSC_EWMA).

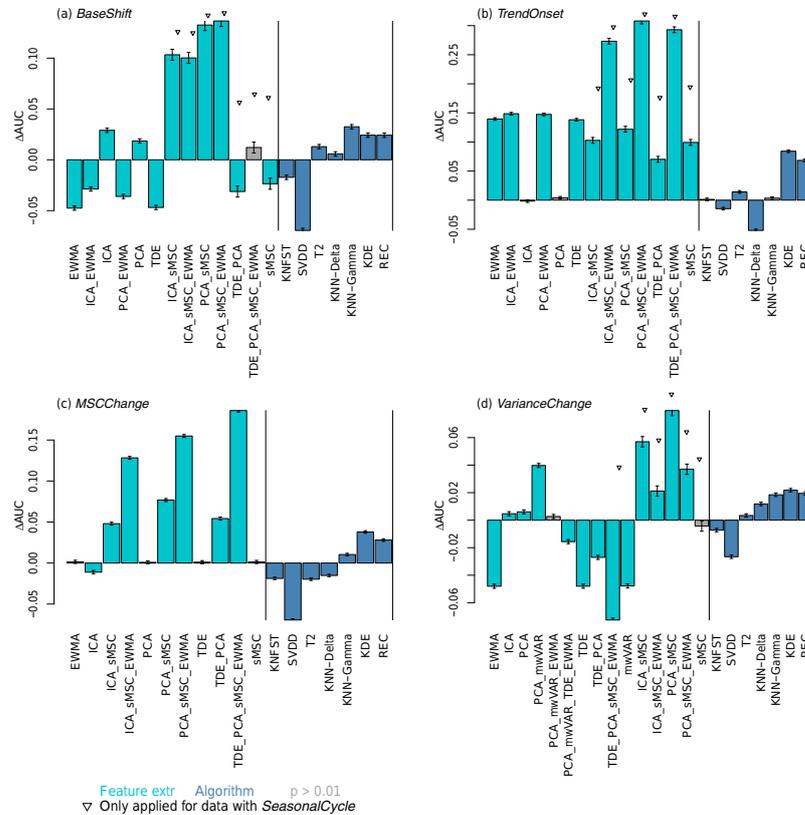


Figure 4. AUC difference with respect to the UNIV control in the experimental factors feature extraction and detection algorithm for the event types (a–d).

VarianceChange The algorithms used are hardly able to detect any decrease in variance (Fig. B1). This may be due to an overwriting of the decrease in signal variance with the independent noise since we are using a signal-to-noise ratio of 0.3. Thus, we exclude a decrease in the variance from the evaluation of the detection algorithms compared to the control. The detection of an increase in the variance can be improved by a combination of dimensionality reduction and variance in a moving window (PCA_mwVAR) (Fig. 4d). Using the variance in a moving window is a popular approach (Huntingford et al., 2013) although it has to be applied with care when used in conjunction with normalization procedures (Sippel et al., 2015).

SeasonalCycle Seasonality occurs in most EOs. Not accounting for the seasonal cycle has a negative impact on the AUC (Appendix B, Fig. B2a, b, d). However, if we subtract the median cycle within the feature ex-

traction step (PCA_sMSC_EWMA; Fig. 4a, b, d), we can almost account for the negative AUC impact of the seasonal cycle, as in our experimental setting anomalous events do occur independently of seasonality. However, depending on the research question, independence of seasonality might not always be the case: some EOs may depend on vegetation activity, for example, which results in a strong dependence on seasonality.

4.2 Performance of multivariate anomaly detection algorithms

In contrast to the investigated combinations of feature extraction methods, we can identify three of the tested algorithms performing on average almost equally well for most event types given a suitable feature extraction as discussed before (Sect. 4.1).

Table 1. Average AUC difference of the anomaly detection algorithms from the UNIV control for each event type.

	KNFST	SVDD	$T2$	KNN-Delta	KNN-Gamma	KDE	REC
BaseShift	−0.017	−0.069	0.013	0.006	0.032	0.024	0.024
TrendOnset	0.001	−0.015	0.014	−0.052	0.003	0.084	0.068
MSCChange	−0.023	−0.072	−0.023	−0.019	0.007	0.039	0.029
VarianceChange	−0.007	−0.027	0.003	0.012	0.018	0.022	0.019
Mean	−0.012	−0.046	0.002	−0.013	0.015	0.042	0.035
RVP	0.007	0.111	0.003			0.000	0.001

KDE and REC These techniques exhibit overall the highest AUC and lowest RVP (Table 1). Their estimated mean differences are rather small since REC can be considered as a binary form of the KDE. As REC uses a threshold ε for defining the hyperball of recurrences, the results can exhibit slightly higher AUC than KDE (Fig. S1). However, with REC the caveat is that the parameter ε is not necessarily optimally chosen.

KNN In most of the cases, KNN-Gamma performance is better than the UNIV control, but only as good as the UNIV control for detecting TrendOnset. This may be due to the fact that for TrendOnset, the mean distance to the KNN does not change, unless considering a very large number of KNN values or excluding a large fraction of temporally near data points to be within the KNN. When excluding TrendOnset, the mean performance increases to 0.019, which is comparable to KDE and REC. In addition, we observe even superior performance of KNN-Gamma compared to KDE and REC for difficult data properties (e.g., MoreIndepComponents, CorrelatedNoise; Fig. S2). In contrast, KNN-Delta does not yield high AUC, probably because we do not construct anomalies in the data cube explicitly with a direction that is accounted for by KNN-Delta (mean length of the vectors to its KNN). The finding that simple algorithms like KNN-Gamma (or KDE, $T2$) are very competitive, if not favorable algorithms, is in line with results of Harmeling et al. (2006), Killourhy and Maxion (2009), and Ding et al. (2014) for various data sets.

KNFST and SVDD These techniques perform on average worse than or equally as well as UNIV. Also, the RVP is highest among the algorithms (Table 1). It has already been reported that SVDD can exhibit remarkable fluctuations in the results for sample sizes smaller than 1000 data points (Ding et al., 2014). However, we use 5000 points for training. Thus, we suggest that the fluctuations are due to the fact that SVDD and KNFST use a training set that is chosen at random and may itself contain anomalies. In the current setting the size of the training sample (5000) is rather small compared to the spatiotemporal size of the data cube (750 000), and it does not seem to be sufficient to train these algo-

gorithms on the data cube. Increased sample sizes, however, would heavily increase memory demand and computing time, rendering kernel algorithms computationally inapplicable. Furthermore, Ding et al. (2014) shows that the sample size has a remarkable effect for SVDD (better performance for larger sample sizes). However, even with very large sample sizes SVDD still performs worse than KNN in the setting of Ding et al. Training and testing SVDD on each pixel does also not improve the results as the number of anomalies differs between different pixels in our setting. Training and testing SVDD on each pixel assumes the same number of anomalies in each pixel (constant outlier ratio assumed by the fixed ν parameter), which is contrary to the generation of the artificial data farm.

We explicitly do not want to state that KNFST and SVDD are generally worse algorithms, i.e., they are just not built for these massive numbers of data. KNFST and SVDD outperform other algorithms in very different settings (novelty detection in images) (Bodesheim et al., 2013).

$T2$ exhibits good performance for detecting starting trends and shifts in the mean. However, it also exhibits the third largest RVP (Table 1), indicating that the estimation of the covariance matrix may be sensitive to random variation in the data. Nevertheless, the RVP is still far better than for SVDD. The robust estimation of the mean and covariance matrix might be a difficult task (Smetek and Bauer, 2007; Rousseeuw and Hubert, 2011) for which rather complex algorithms like the (fast) minimum determinant covariance estimator, which are closely related to $T2$, have been proposed (Rousseeuw and Van Driessen, 1990). Furthermore, $T2$ assumes a multivariate Gaussian distribution and linear dependencies among the variables. Thus, it is not preferable for the data properties NonLinearDep and CorrelatedNoise unless combined with a nonlinear feature extraction technique like ICA (Fig. B1).

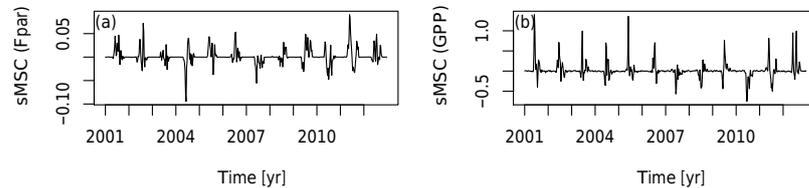


Figure 5. The residual time series obtained by subtracting the median seasonal cycle from (a) the fraction of absorbed photosynthetic radiation (Fpar) and (b) gross primary productivity (GPP) at northern latitudes exhibit heteroscedastic patterns.

4.3 Ensembles

The selection of algorithms for computing the ensemble is a compromise between accurate detection of and diversity amongst the selected algorithms (Zimek et al., 2013). We select the four best algorithms (KDE, REC, KNN-Gamma, T_2 ; referred to together as 4b) and the three best distance-based algorithms (KDE, REC, KNN-Gamma; referred to together as 3d) for computing their ensembles. We assume that this choice accounts for accuracy (best algorithms selected) as well as for diversity (different algorithms selected).

Overall, ensemble building improves the anomaly detection rate. The mean AUC of each of the ensemble members (3d: +0.030, 4b: +0.023) is lower than the AUC of the ensemble, regardless of whether the maximum or the mean is used for ensemble aggregation. Minimum aggregation of ensemble members, however, performs worse than the individual ensemble members REC and KDE. Using the maximum or mean yields consistently higher AUC than using the minimum (Table 2). The superior performance of the maximum choice compared to the minimum indicates that single algorithms overlook more often anomalous events than raising false alarm. Nevertheless, the maximum has the caveat that even a single algorithm may cause a false alarm (Zimek et al., 2013), e.g., due to a poor parameterization or inadequate assumptions about properties of the data. Thus, a more balanced voting procedure like the mean is the preferable choice and more stable with respect to possible error sources. Among the mean ensembles, the 3d or 4b ensembles perform equally well (0.041 vs. 0.039 ± 0.001 overall) (Table 2).

4.4 Limitations

High dimensionality The utility of distance-based outlier detection algorithms as used in this paper is often questioned in the context of high-dimensional data (Zimek et al., 2012). The “curse of dimensionality” states that the difference between near and far distances diminishes with increasing dimensionality. However, Zimek et al. (2012) showed in the case of KNN that the contrary is true for outliers with fixed magnitude in otherwise uncorrelated data. Dimensionality reduction as crucial feature extraction transforms the data into a few (ideally)

meaningful and uncorrelated variables. Thus, the findings of Zimek et al. (2012) provide strong arguments for applying dimensionality reduction on correlated data. We anticipate that their findings are the reason of the superior performance of dimensionality reduction here.

Heuristic choices Within the parameterization process, several heuristic choices are made. We exclude five time steps to be counted as recurrences or k -nearest neighbors. We fix several parameters, e.g., the number of nearest neighbors is fixed to 10. Also, other parameter choices are rather heuristic (e.g., σ), although commonly used. The artificial data farm’s intrinsic dimension is 3 as it was created from three independent components. Therefore, the embedding dimension m is fixed accordingly, although it can be inferred based on the data by determining the number of false nearest neighbors (Kennel et al., 1992; Hegger et al., 1999). The signal-to-noise ratio of our artificial data farm is 0.3. Furthermore, the choice of the data properties might influence the results for each event type, as the standard deviation of AUC values over all data properties (0.05) is rather large, compared to the average AUC gain of the three best algorithms with respect to the control (+0.03). However, the ordering of the algorithms is also important to derive rankings of algorithms (Hornik and Meyer, 2007). By choosing different subsets of the data properties, we observe that the three best algorithms (KDE, REC, KNN-Gamma) are on top, independently of the chosen data property. Therefore, the data properties might have an influence on the AUC values themselves, but not on the choice of the top three candidates.

5 Remarks on applications for real Earth observations

Our version of the artificial data farm was generated to test different algorithms for their capability to deal with typical properties of EO data. The workflows were chosen to be as generic as possible, and therefore their application to real data with slightly different properties should be made as easy as possible. Nevertheless, several points have to be considered, when applying the algorithms to real EOs.

Table 2. AUC difference of the ensembles of anomaly detection algorithms to the UNIV control. Ensembles are computed out of the four best algorithms (4b, KDE, REC, KNN-Gamma, T_2) and the three best distance-based algorithms (3d, KDE, REC, KNN-Gamma).

	3d max	3d mean	3d min	4b max	4b mean	4b min
BaseShift	0.042	0.037	0.033	0.042	0.038	0.030
TrendOnset	0.059	0.058	0.033	0.060	0.056	0.020
MSCChange	0.033	0.040	0.032	0.033	0.037	0.017
VarianceChange	0.027	0.027	0.025	0.023	0.026	0.022
Mean	0.040	0.041	0.031	0.039	0.039	0.022
RVP	0.001	0.001	0.001	0.001	0.001	0.001

A typical preprocessing of EOs is to center variables to zero mean and standardize to unit variance (also known as z transformation). A standardization of this kind is of key importance in global EOs. Real multivariate observations often have different physical units or ranges, which have to be made comparable before analyzing. However, standardization has to be applied with care. Differences between the mean and variance between geographically distinct or even adjacent grid cells as well as seasonal cycles might corrupt any further analysis. We recommend subtracting the median seasonal cycle before standardization. The median is preferred over the mean as mean seasonal cycles are affected by changes in the amplitude of the cycle. Standardization can be applied globally (i.e., with global spatiotemporal mean and variance), regionally (i.e., with spatiotemporal mean and variance in subregions of the globe), or locally (i.e., with temporal mean and variance in each grid cell). Global standardization might be more robust than local but detects only anomalies in high-variance regions. Local standardization assumes that the number of extreme anomalies is equal in each grid cell, which is a rather strong assumption. Thus, a regional standardization is favorable in regions with similar mean and variance.

Especially variables presenting a signal from the biosphere are known to exhibit heteroscedasticity, e.g., the variance during growing season is substantially larger than during the rest of the year (Fig. 5). Atmospheric variables in high latitudes also show higher variability during the cold season, e.g., temperature variability might be higher over ice (cold season) than over open water (warm season) (Hansen et al., 2012). Specifically for global applications, using estimates of variance or standard deviation locally (in each grid cell) leads to an underestimation of the variance during growing season and thus to an overestimation of anomalies due to standardization, especially in the northern latitudes (Guanche et al., 2016). Thus, we recommend accounting for the heteroscedastic pattern by adjusting the variance during the growing season within similar regions. We also recommend this kind of adjustment for the covariance matrix used, e.g., in T_2 or PCA as well as for the parameterization of KDE or REC.

Furthermore, anomalies are also overestimated when using a reference period for the estimation of the variance (Sippel et al., 2015). However, with 300 observations in 8-day intervals, as used in this study, this issue is expected to be less pronounced than for fewer observations as it scales with the length of the time series. Nevertheless, we rather recommend using estimates of the variance of the entire time series or correcting for the overestimation in the out-of-reference period as shown in Sippel et al. (2015).

Regarding the parameterization process of the algorithms, we use fixed parameters for σ , ε , k , ν , mean vector, and covariance matrix globally on the entire artificial data cube. Local parameterization assumes the same number of anomalies in each region, which is neither suitable for the artificial data by construction nor for real global data. Thus, we recommend parameterizing globally or within similar regions. Classification of the Earth into similar regions and applying multivariate extreme detection in each region will be the subject of future research.

6 Conclusions

Our aim is to identify suitable methods for detecting anomalies in highly multivariate, correlated, and seasonally varying data streams as they are common in Earth system science. In particular, we are interested in detecting shifts in mean (extremes), changes in the amplitude of the seasonal cycle, temporal changes in the variance, and onsets of trends. We test a wide range of workflows (i.e., combining feature extraction techniques and anomaly detection algorithms). All experiments are based on artificial data, designed to mimic real world Earth observations.

We can show that, on average over different anomaly types and data properties, three multivariate anomaly detection algorithms (KDE, REC, KNN-Gamma) outperform univariate extreme event detection as well as other multivariate approaches (mean AUC compared to univariate control: +0.030). Additional slight improvement can be achieved by combining the best algorithms into ensembles using an aggregation by averaging score quantiles (+0.041). In contrast, the tested machine-learning algorithms (SVDD −0.05, KN-FST −0.01) may fail due to overfitting to the training sample.

However, we also find for the considered type of events that including a suitable feature extraction technique in the detection workflow is often more important than the choice of the event detection algorithm itself. However, we find that the feature extraction has to be explicitly designed for the event type of interest, i.e., time delay embedding (for detecting changes in the cycle amplitude) and exponential weighted moving average (for detecting trends and long extremes and removing uncorrelated noise in the signal) increases the detection rate of the anomalous events. Including features of the variance within a moving window works partly for detecting increases in the variance but fails to detect a decrease in the variance due to the relatively high observational noise level. In general, if the data comprise seasonality, subtracting them and using the remaining time series as the input feature is essential. Furthermore, we improve the detection rate of multivariate anomalies in highly correlated data streams by adding a dimensionality reduction method to the workflow (in line with results of Zimek et al., 2012).

The proposed workflows are capable of dealing with common properties of Earth observations like seasonality, non-linear dependencies, and (to a certain degree) non-Gaussian distributions and noise. Nevertheless, they have to be applied with care to Earth observations, i.e., standardization issues along with strong heteroscedastic patterns (e.g., in biosphere variables of northern latitudes) may lead to an overestimation of anomalies. Future work will explore the potential of the identified workflows in rediscovering known and potentially unknown extremes as well as other anomalies in a set of real Earth system science data streams. We anticipate that an automated application of our workflows might enable the establishment of automated Earth system process control in a very generic manner.

Data availability. The artificial data farm can be created after cloning in <https://github.com/CAB-LAB/DataFarm>. Generation is done with the following command within the programming language Julia, version 0.4, using `SurrogateCube`, `SurrogateCube.DataFarm.makeDataFarm(300,50,50,PathToFolder)`.

Appendix A: Technical details on generating the artificial data

Within the generation process, we assume that the signal S is additive to the baseline B . The baseline might represent reoccurring patterns like seasonality or a constant mean. In addition, binary event parameters $ev_{t,lat,lon}$ are introduced, which allow for switching the anomaly on ($ev_{t,lat,lon} \neq 0$) and off ($ev_{t,lat,lon} = 0$) (normality). The event type and magnitude of the event is controlled separately by a parameter for the baseline (k_b), the signal (k_s), and a mean-shift parameter (k_m) scaled with the standard deviation of the data (SD).

$$\Theta_{t,lat,lon} = B_{t,lat,lon} \cdot \mathcal{N}^{(k_b \cdot ev_{t,lat,lon})} + S_{t,lat,lon} \cdot \mathcal{N}^{(k_s \cdot ev_{t,lat,lon})} + k_m \cdot ev_{t,lat,lon} \cdot SD \quad (A1)$$

For a basic version, three independent components $\Theta_{t,lat,lon,var}$ are created with the signal consisting of Gaussian noise (SD = 1). Each component represents intrinsic properties of the Earth system. Furthermore, we assume that properties of the Earth system $\Theta_{t,lat,lon}$ are not measured directly but indirectly via a set of correlated variables, i.e., representing patterns of these intrinsic properties. Hence, these variables propagate anomalous events that occur in one independent component. This set of correlated variables X_{var} is created by weighting the intrinsic properties Θ_{var} with randomly drawn linear (or nonlinear) weights w_j plus additional measurement noise ϵ (Gaussian, SD = 0.3) added to each variable.

$$X_{var} = \sum_{j=1}^{j=3} w_j \cdot \Theta_j + \epsilon \quad (A2)$$

Using this data generation scheme, a standard data cube $X_{t,j,lat,lon,var}$ is created, encompassing 300 time steps (T), 10 temporally correlated variables (VAR), and the total number of latitudes (LAT) and longitudes (LON) fixed to 50 each. We induce anomalous events with a spatial extent of 40 % of the latitude and longitude and 10 events, each with a temporal extent of five time steps. Our total number of anomalies equals about 3 % of the total data cube.

In the basic version we create four data cubes, each with a different temporary event type, including

- shift in the baseline, i.e., shift of the running mean of a time series (BaseShift) (Fig. 2a);
- change in the variance of the time series (VarianceChange) (Fig. 2b);
- change in the amplitude of the mean seasonal cycle of a time series (MSCChange) (Fig. 2c);
- onset of a trend in the time series (TrendOnset) (Fig. 2d).

Regarding the data properties, some of the event type data property combinations are excluded (Table A1). In detail, we do not expect a TrendOnset to infect neighbored cells (TrendOnset plus RandomWalkExtreme) and a TrendOnset can hardly be called a TrendOnset if it encompasses only one time step (ShortExtremes).

Table A1. Parameter settings for the generation of the artificial data farm. Details are given for each event type and data property (in brackets).

	Basic	(Data property)	BaseShift	VarianceChange	MSCChange	TrendOnset
Independent comp.	Θ	3	(MoreIndepComponents)	3 (6)	3 (6)	3 (6)
Dependency (Θ)	Linear (w)	(NonLinearDep squ)	w (squ)	w (squ)	w (sq)	w (sq)
Baseline	B	Const. = c	(SeasonalCycle s , LatitudinalGradient lg)	c (s , lg)	c (s , lg)	c (s , lg)
Signal	S	Gaussian g	(LaplacianNoise l , CorrelatedNoise r)	g (l , r)	g (l , r)	g (l , r)
Variables	VAR	10	10	10	10	10
Noise	ϵ	0.3	(NoiseIncrease)	0.3 (1)	0.3 (1)	0.3 (1)
Events						
Event number	10	(ShortExtremes, LongExtremes)	10 (50, 5)	10 (50, 5)	1	1
Spatial extent	1000		1000	1000	4	1000
Temporal extent	5	(ShortExtremes, LongExtremes)	5 (1,10)	5 (1,10)	92 (46, 184)	150
Magnitudes			$k_m = 0.2-4$	$k_s = -2 : 2$	$k_b = -2 : 2$	$k_m = 0.2-4$
Shape	Rect.	(RandomWalkExtreme rw)	rect (rw)	rect (rw)	rect (rw)	rect

Appendix B: Detailed results

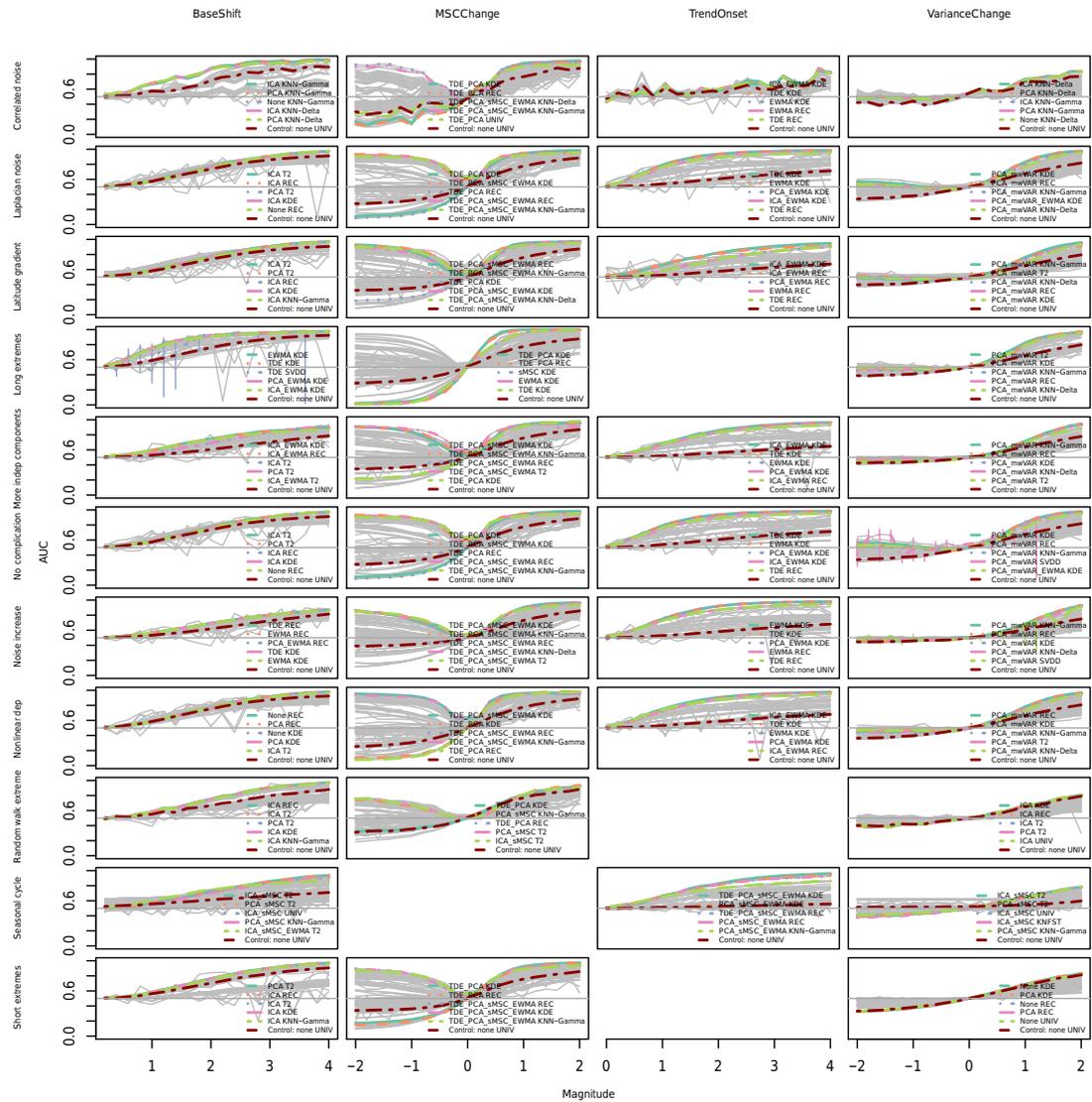


Figure B1. AUC versus event magnitude for all combinations (grey) and the univariate control (red). Columns of the matrix represent different event types; rows represent data properties. Additional colored workflows represent the workflows with the five highest mean values for the magnitudes $> 2SD$ (> 0.6).

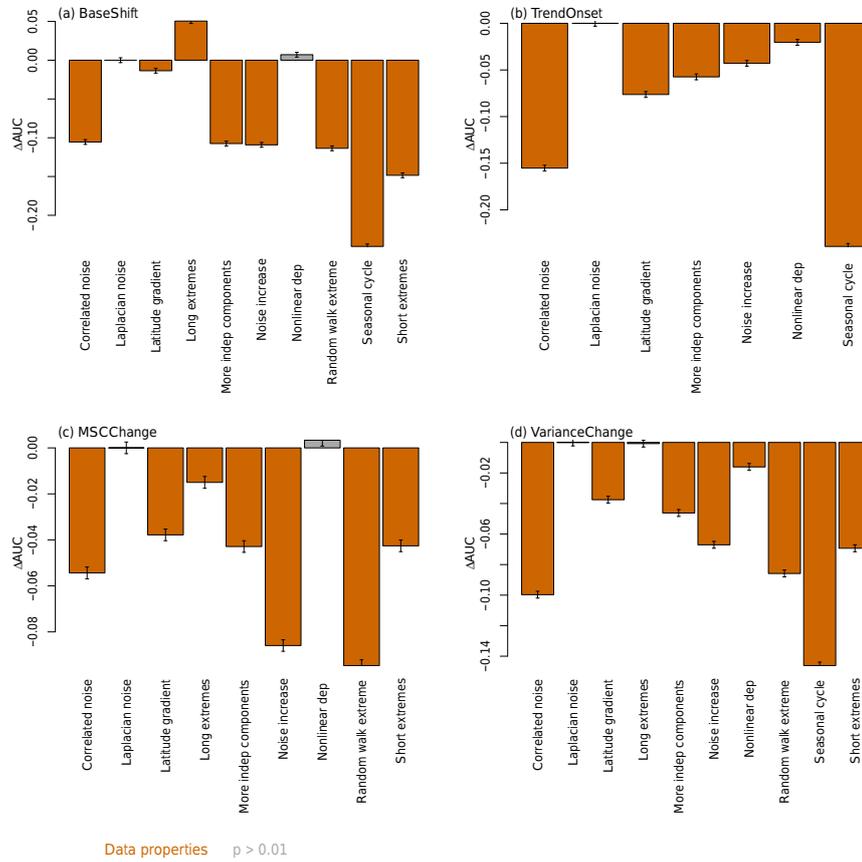


Figure B2. Effect of the data properties on the three best detection algorithms (KDE, REC, KNN-Gamma) presented as AUC difference of the UNIV control for the event types (a–d).

The Supplement related to this article is available online at <https://doi.org/10.5194/esd-8-677-2017-supplement>.

Author contributions. MF and MDM designed the study in collaboration with FG, AB, JD, MR, and ER; MF implemented the algorithms, including contributions from FG, PB, and ER; MF wrote the paper with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research has received funding from the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS), the European Space Agency via the STSE project CAB-LAB, and the BACI project, a European Union's Horizon 2020 research and innovation programme under grant agreement no. 64176. We thank Simone Girst for her kind language check. Reik Donner and one anonymous referee provided valuable suggestions for improvement.

The article processing charges for this open-access publication were covered by the Max Planck Society.

Edited by: Sagnik Dey

Reviewed by: Reik Donner and one anonymous referee

References

- Aggarwal, C. C.: Outlier Ensembles, *ACM SIGKDD Explorations Newsletter*, 14, 49–58, 2012.
- Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L.: Global observed changes in daily climate extremes of temperature and precipitation, *J. Geophys. Res.*, 111, D05109, <https://doi.org/10.1029/2005JD006290>, 2006.
- Bae, K.-H., Karolyi, G. A., and Stulz, R. M.: A New Approach to Measuring Financial Contagion, *Rev. Financ. Stud.*, 16, 717–763, 2003.
- Baldocchi, D., Falge, E., and Wilson, K.: A spectral analysis of biosphere–atmosphere trace gas flux densities and meteorological variables across hour to multi-year time scales, *Agr. Forest Meteorol.*, 107, 1–27, 2001.
- Bathiany, S., Notz, D., Mauritsen, T., Raedel, G., and Brovkin, V.: On the Potential for Abrupt Arctic Winter Sea Ice Loss, *J. Climate*, 29, 2703–2719, 2016.
- Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Altaf Arain, M., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Rouspard, 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, 843–838, 2010.
- Bintanja, R. and van der Linden, E. C.: The changing seasonal climate in the Arctic, *Sci. Rep.*, 3, 1556, <https://doi.org/10.1038/srep01556>, 2013.
- Bodesheim, P., Freytag, A., Rodner, E., Kemmler, M., and Denzler, J.: Kernel Null Space Methods for Novelty Detection, *CVPR*, Portland, Oregon, 3374–3381, 2013.
- Chang, C.-C. and Lin, C.-J.: LIBSVM: A Library for Support Vector Machines, *ACM Transactions on Intelligent Systems and Technology*, 2, 27:1–27:27, 2013.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, 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, 2005.
- Ciais, P., Dolman, A. J., Bombelli, A., Duren, R., Pregon, A., Rayner, P. J., Miller, C., Gobron, N., Kinderman, G., Marland, G., Gruber, N., Chevallier, F., Andres, R. J., Balsamo, G., Bopp, L., Bréon, F.-M., Broquet, G., Dargaville, R., Battin, T. J., Borges, A., Bovensmann, H., Buchwitz, M., Butler, J., Canadell, J. G., Cook, R. B., DeFries, R., Engelen, R., Gurney, K. R., Heinze, C., Heimann, M., Held, A., Henry, M., Law, B., Luyssaert, S., Miller, J., Moriyama, T., Moulin, C., Myeni, R. B., Nussli, C., Obersteiner, M., Ojima, D., Pan, Y., Paris, J.-D., Piao, S. L., Poulter, B., Plummer, S., Quegan, S., Raymond, P., Reichstein, M., Rivier, L., Sabine, C., Schimel, D., Tarasova, O., Valentini, R., Wang, R., van der Werf, G., Wickland, D., Williams, M., and Zehner, C.: Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system, *Biogeosciences*, 11, 3547–3602, <https://doi.org/10.5194/bg-11-3547-2014>, 2014.
- Ding, X., Li, Y., Belatreche, A., and Maguire, L. P.: An experimental evaluation of novelty detection methods, *Neurocomputing*, 135, 313–327, 2014.
- Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Dunn, R. J. H., Willett, K. M., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, A. M. G., Kruger, A. C., Marengo, J., Peterson, T. C., Renom, M., Oria Rojas, C., Rusticucci, M., Salinger, J., Elrayah, A. S., Sekele, S. S., Srivastava, A. K., Trewin, B., Villarreal, C., Vincent, L. A., Zhai, P., Zhang, X., and Kitching, S.: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset, *J. Geophys. Res.-Atmos.*, 118, 2098–2118, 2013.
- Donges, J. F., Donner, R. V., Rehfeld, K., Marwan, N., Trauth, M. H., and Kurths, J.: Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis, *Nonlin. Processes Geophys.*, 18, 545–562, <https://doi.org/10.5194/npg-18-545-2011>, 2011a.
- Donges, J. F., Donner, R. V., Trauth, M. H., Marwan, N., Schellnhuber, H.-J., and Kurths, J.: Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution, *P. Natl. Acad. Sci. USA*, 108, 20422–20427, 2011b.

- Donges, J. F., Heitzig, J., Donner, R. V., and Kurths, J.: Analytical framework for recurrence network analysis of time series, *Phys. Rev. E*, 85, 046105, <https://doi.org/10.1103/PhysRevE.85.046105>, 2012.
- Donges, J. F., Schleussner, C. F., Siegmund, J. F., and Donner, R. V.: Event coincidence analysis for quantifying statistical interrelationships between event time series, *The European Physical Journal Special Topics*, 225, 471–487, 2016.
- Donner, R. V., Zou, Y., Donges, J. F., Marwan, N., and Kurths, J.: Recurrence networks – A novel paradigm for nonlinear time series analysis, *New J. Phys.*, 12, 033025, <https://doi.org/10.1088/1367-2630/12/3/033025>, 2010.
- Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., and Jackson, T.: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, *Hydrol. Earth Syst. Sci.*, 15, 1675–1698, <https://doi.org/10.5194/hess-15-1675-2011>, 2011.
- Drijfhout, S., Bathiany, S., Beaulieu, C., Brovkin, V., Claussen, M., Huntingford, C., Scheffer, M., Sgubin, G., and Swingedouw, D.: Catalogue of abrupt shifts in Intergovernmental Panel on Climate Change climate models, *P. Natl. Acad. Sci.*, 112, E5777–E5786, 2015.
- Durante, F. and Salvadori, G.: On the construction of multivariate extreme value models via copulas, *Environmetrics*, 21, 143–161, 2010.
- Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, T. R. K., and Mearns, L. O.: Climate Extremes: Observations, Modeling, and Impacts, *Science*, 289, 2068–2074, 2000.
- Faranda, D. and Vauti, S.: A new recurrences based technique for detecting robust extrema in long temperature records, *Geophys. Res. Lett.*, 40, 5782–5786, 2013.
- Fawcett, T.: An introduction to ROC analysis, *Pattern Recogn. Lett.*, 27, 861–874, 2006.
- Fischer, E. M.: Robust projections of combined humidity and temperature extremes, *Nature Climate Change*, 3, 126–130, 2013.
- Flach, M., Lange, H., Foken, T., and Hauhs, M.: Recurrence Analysis of Eddy Covariance Fluxes, in: *Recurrence Plots and Their Quantifications: Expanding Horizons*, edited by: Webber Jr., C. L., Ioana, C., and Marwan, N., Springer Proceedings in Physics, Cham, 301–319, 2016.
- Forkel, M., Carvalhais, N., Verbesselt, J., Mahecha, M., Neigh, C., and Reichstein, M.: Trend Change Detection in NDVI Time Series: Effects of Inter-Annual Variability and Methodology, *Remote Sens.*, 5, 2113–2144, 2013.
- Ge, Z., Song, Z., and Gao, F.: Review of Recent Research on Data-Based Process Monitoring, *Ind. Eng. Chem. Res.*, 52, 3543–3562, 2013.
- Ghil, M., Yiou, P., Hallegatte, S., Malamud, B. D., Naveau, P., Soloviev, A., Friederichs, P., Keilis-Borok, V., Kondrashov, D., Kossobokov, V., Mestre, O., Nicolis, C., Rust, H. W., Shebalin, P., Vrac, M., Witt, A., and Zaliapin, I.: Extreme events: dynamics, statistics and prediction, *Nonlin. Processes Geophys.*, 18, 295–350, <https://doi.org/10.5194/npg-18-295-2011>, 2011.
- Guanche, Y., Rodner, E., Flach, M., Sippel, S., Mahecha, M. D., and Denzler, J.: Detecting Multivariate Biosphere Extremes, in: *Proceedings of the 6th International Workshop on Climate Informatics: CI2016*, edited by: Banerjee, A., Ding, W., and Dy, V., NCAR Technical Note NCAR/TN-529+PROC, Boulder: National Center for Atmospheric Research, 9–12, 2016.
- Hansen, J., Sato, M., and Ruedy, R.: Perception of climate change, *P. Natl. Acad. Sci. USA*, 109, E2415–E2423, 2012.
- Harmeling, S., Dornhege, G., Tax, D., Meinecke, F., and Müller, K.-R.: From outliers to prototypes: Ordering data, *Neurocomputing*, 69, 1608–1618, 2006.
- Hegger, R., Kantz, H., and Schreiber, T.: Practical implementation of nonlinear time series methods: The TISEAN package, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 9, 413–435, 1999.
- Hornik, K. and Meyer, D.: Deriving Consensus Rankings from Benchmarking Experiments, in: *Advances in Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization*, edited by: Decker, R. and Lenz, H.-J., Springer, Berlin, Heidelberg, 163–170, 2007.
- Hotelling, H.: Multivariate Quality Control – Illustrated by the Air Testing of Sample Bombsights, in: *Techniques of Statistical Analysis*, edited by: Eisenhart, C., Hastay, M. W., and Wallis, W. A., McGraw-Hill, New York, 111–184, 1947.
- Huntingford, C., Jones, P. D., Livina, V. N., Lenton, T. M., and Cox, P. M.: No increase in global temperature variability despite changing regional patterns, *Nature*, 500, 327–330, 2013.
- Hyvärinen, A. and Oja, E.: Independent component analysis: algorithms and applications, *Neural Networks*, 13, 411–430, 2000.
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneeth, 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, *J. Geophys. Res.*, 116, G00J07, <https://doi.org/10.1029/2010JG001566>, 2011.
- Kennel, M. B., Brown, R., and Abarbanel, H. D. I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Phys. Rev. A*, 45, 3403–3411, 1992.
- Kharin, V. V., Zwiers, F. W., Zhang, X., and Wehner, M.: Changes in temperature and precipitation extremes in the CMIP5 ensemble, *Climatic Change*, 119, 345–357, 2013.
- Killourhy, K. S. and Maxion, R. A.: Comparing Anomaly-Detection Algorithms for Keystroke Dynamics, *IEEE/IFIP International Conference on Dependable Systems & Networks*, 125–134, 2009.
- Koçak, K., Şaylan, L., and Eitzinger, J.: Nonlinear prediction of near-surface temperature via univariate and multivariate time series embedding, *Ecol. Model.*, 173, 1–7, 2004.
- Ledford, A. W. and Tawn, J. A.: Statistics for near independence in multivariate extreme values, *Biometrika*, 83, 169–187, 1996.
- Lee, J.-M., Yoo, C., and Lee, I.-B.: Statistical monitoring of dynamic processes based on dynamic independent component analysis, *Chem. Eng. Sci.*, 59, 2995–3006, 2004.
- Lehmann, J., Coumou, D., and Frieler, K.: Increased record-breaking precipitation events under global warming, *Climatic Change*, 132, 501–515, 2015.
- Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M.: A compound event framework for understand-

- ing extreme impacts, *Wiley Interdisciplinary Reviews: Climate Change*, 5, 113–128, 2013.
- Lim, S. A. H., Antony, J., and Albliwi, S.: Statistical Process Control (SPC) in the food industry – A systematic review and future research agenda, *Trends Food Sci. Tech.*, 37, 137–151, 2014.
- Lowry, C. A. and Montgomery, D. C.: A review of multivariate control charts, *IEE Trans.*, 27, 800–810, 1995.
- Lowry, C. A. and Woodall, W. H.: A Multivariate Exponentially Weighted Moving Average Control Chart, *Technometrics*, 34, 46–53, 1992.
- Majeed, W. and Avison, M. J.: Robust Data Driven Model Order Estimation for Independent Component Analysis of fMRI Data with Low Contrast to Noise, *PLoS ONE*, 9, e94943, <https://doi.org/10.1371/journal.pone.0094943>, 2014.
- Marwan, N., Romano, M. C., Thiel, M., and Kurths, J.: Recurrence plots for the analysis of complex systems, *Phys. Rep.*, 438, 237–329, 2007.
- Meehl, G. A. and Tebaldi, C.: More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century, *Science*, 305, 994–997, 2004.
- Mikosch, T.: Copulas: Tales and facts, *Extremes*, 9, 3–20, 2006.
- Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M., and Mairota, P.: Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats, *Ecol. Indic.*, 33, 45–59, 2013.
- Overpeck, J. T., Meehl, G. A., Bony, S., and Easterling, D. R.: Climate Data Challenges in the 21st Century, *Science*, 331, 700–703, 2011.
- Parzen, E.: On Estimation of a Probability Density Function and Mode, *The Annals of Mathematical Statistics*, 33, 1065–1076, 1962.
- Pfeifer, M., Disney, M., Quaife, T., and Marchant, R.: Terrestrial ecosystems from space: a review of Earth observation products for macroecology applications, *Global Ecol. Biogeogr.*, 21, 603–624, 2011.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team: nlme: Linear and Nonlinear Mixed Effects Models, <http://CRAN.R-project.org/package=nlme>, last access: 15 October 2016, R package version 3.1-128, 2016.
- Poincaré, H.: Sur le problème des trois corps et les équations de la dynamique, *Acta Math.*, 13, 5–7, 1890.
- Rahmstorf, S. and Coumou, D.: Increase of extreme events in a warming world, *P. Natl. Acad. Sci. USA*, 108, 17905–17909, 2011.
- Ramaswamy, S., Rastogi, R., and Shim, K.: Efficient Algorithms for Mining Outliers from Large Data Sets, *SIGMOD Record*, 29, 427–438, 2000.
- 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, <https://doi.org/10.5194/bg-12-373-2015>, 2015.
- 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, 2013.
- Rousseeuw, P. J. and Hubert, M.: Robust statistics for outlier detection, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 73–79, 2011.
- Rousseeuw, P. J. and Van Driessen, K.: A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics*, 41, 212–223, 1990.
- Santos-Fernández, E.: *Multivariate Statistical Quality Control Using R*, vol. 14 of SpringerBriefs in Statistics, 1 Edn., Springer, New York, Heidelberg, Dordrecht, London, 2013.
- Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research – introduction to the copula approach, *Nonlin. Processes Geophys.*, 15, 761–772, <https://doi.org/10.5194/npg-15-761-2008>, 2008.
- Schölkopf, B. and Smola, A.: *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, USA, 2001.
- Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C.: Estimating the Support of a High-Dimensional Distribution, *Neural Comput.*, 13, 1443–1471, 2001.
- Schölkopf, B., Muandet, K., Fukumizu, K., Harmeling, S., and Peters, J.: Computing functions of random variables via reproducing kernel Hilbert space representations, *Stat. Comput.*, 25, 755–766, 2015.
- Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment, in: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (IPCC SREX Report)*, edited by: Field, C., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., Mastrandrea, M., Mach, K., Plattner, G.-K., Allen, S., Tignor, M., and Midgley, Cambridge University Press, Cambridge, 1–582, 2012.
- Siegmund, J. F., Sanders, T. G. M., Heinrich, I., van der Maaten, E., Simard, S., Helle, G., and Donner, R. V.: Meteorological Drivers of Extremes in Daily Stem Radius Variations of Beech, Oak, and Pine in Northeastern Germany: An Event Coincidence Analysis, *Frontiers in Plant Science*, 7, 733, <https://doi.org/10.1093/treephys/21.9.561>, 2016.
- Sippel, S., Zscheischler, J., Heimann, M., Otto, F. E. L., Peters, J., and Mahecha, M. D.: Quantifying changes in climate variability and extremes: Pitfalls and their overcoming, *Geophys. Res. Lett.*, 42, 9990–9998, 2015.
- Smetek, T. E. and Bauer, K. W.: Finding Hyperspectral Anomalies Using Multivariate Outlier Detection, *Proc. 2007 IEEE Aerosp. Conf.*, 1–24, 2007.
- Smets, K., Verdonk, B., and Jordaán, E. M.: Discovering Novelty in Spatio/Temporal Data Using One-Class Support Vector Machines, *Proceeding of International Joint Conference on Neural Networks*, 2956–2963, 2009.
- Takens, F.: Detecting strange attractors in turbulence, in: *Dynamical Systems and Turbulence, Lecture notes in mathematics*, edited by: Rand, D. and Young, L.-S., Springer, Coventry, England, 366–381, 1980.
- Tax, D. M. and Duin, R. P. W.: Support Vector Data Description, *Mach. Learn.*, 54, 45–66, 2004.

- Thompson, P. D.: How to Improve Accuracy by Combining Independent Forecasts, *Mon. Weather Rev.*, 105, 228–229, 1977.
- 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, <https://doi.org/10.5194/bg-13-4291-2016>, 2016.
- van der Maaten, L. J. P.: Feature Extraction from Visual Data, PhD Thesis, Tilburg University, Tilburg, the Netherlands, <http://insy.ewi.tudelft.nl/content/feature-extraction-visual-data>, last access: 23 June 2009.
- Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, *J. Climate*, 23, 1696–1718, 2010.
- Von Storch, H. and Zwiers, F. W.: *Statistical Analysis in Climate Research*, Cambridge Univ. Press, Cambridge, UK, 2001.
- Webber Jr., C. L. and Marwan, N.: Mathematical and Computational Foundations of Recurrence Quantifications, in: *Recurrence Quantification Analysis*, Springer, Cham, Heidelberg, New York, Dordrecht, London, 3–43, 2015.
- Zhou, B., Gu, L., Ding, Y., Shao, L., Wu, Z., Yang, X., Li, C., Li, Z., Wang, X., Cao, Y., Zeng, B., Yu, M., Wang, M., Wang, S., Sun, H., Duan, A., An, Y., Wang, X., and Kong, W.: The Great 2008 Chinese Ice Storm: Its Socioeconomic–Ecological Impact and Sustainability Lessons Learned, *B. Am. Meteorol. Soc.*, 92, 47–60, 2011.
- Zimek, A., Schubert, E., and Kriegel, H.-P.: A survey on unsupervised outlier detection in high-dimensional numerical data, *Statistical Analysis and Data Mining*, 5, 363–387, 2012.
- Zimek, A., Campello, R. J. G. B., and Sander, J.: Ensembles for Unsupervised Outlier Detection: Challenges and Research Questions, *SIGKDD Explorations*, 15, 11–22, 2013.
- Zscheischler, J., Mahecha, M. D., von Buttler, J., Harmeling, S., Jung, M., Rammig, A., Randerson, J. T., Schölkopf, B., Seneviratne, S. I., Tomelleri, E., Zaehle, S., and Reichstein, M.: A few extreme events dominate global interannual variability in gross primary production, *Environ. Res. Lett.*, 9, 035001, <https://doi.org/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, <https://doi.org/10.5194/bg-11-2909-2014>, 2014b.
- Zscheischler, J., Orth, R., and Seneviratne, S. I.: A submonthly database for detecting changes in vegetation-atmosphere coupling, *Geophys. Res. Lett.*, 42, 9816–9824, 2015.

Chapter 4

Contrasting biosphere responses to hydrometeorological extremes: revisiting the 2010 western Russian heatwave

This chapter is originally published in

Flach, M., Sippel, S., Gans, F., Bastos, A., Brenning, A., Reichstein, M., & Mahecha, M. D. (2018): Contrasting biosphere responses to hydrometeorological extremes: revisiting the 2010 western Russian Heatwave, Biogeosciences, 15, 6067–6085, doi: 10.5194/bg-15-6067-2018. (Selected as Highlight Paper of the European Geophysical Union)



Contrasting biosphere responses to hydrometeorological extremes: revisiting the 2010 western Russian heatwave

Milan Flach¹, Sebastian Sippel², Fabian Gans¹, Ana Bastos³, Alexander Brenning^{4,5}, Markus Reichstein^{1,5}, and Miguel D. Mahecha^{1,5}

¹Max Planck Institute for Biogeochemistry, Department of Biogeochemical Integration, P.O. Box 10 01 64, 07701 Jena, Germany

²Norwegian Institute of Bioeconomy Research, Ås, Norway

³Ludwig-Maximilians University, Department of Geography, Munich, Germany

⁴Friedrich Schiller University Jena, Department of Geography, Jena, Germany

⁵Michael Stifel Center Jena for Data-driven and Simulation Science, Jena, Germany

Correspondence: Milan Flach (milan.flach@bgc-jena.mpg.de)

Received: 15 March 2018 – Discussion started: 4 April 2018

Revised: 27 September 2018 – Accepted: 2 October 2018 – Published: 16 October 2018

Abstract. Combined droughts and heatwaves are among those compound extreme events that induce severe impacts on the terrestrial biosphere and human health. A record breaking hot and dry compound event hit western Russia in summer 2010 (Russian heatwave, RHW). Events of this kind are relevant from a hydrometeorological perspective, but are also interesting from a biospheric point of view because of their impacts on ecosystems, e.g., reductions in the terrestrial carbon storage. Integrating both perspectives might facilitate our knowledge about the RHW. We revisit the RHW from both a biospheric and a hydrometeorological perspective. We apply a recently developed multivariate anomaly detection approach to a set of hydrometeorological variables, and then to multiple biospheric variables relevant to describe the RHW. One main finding is that the extreme event identified in the hydrometeorological variables leads to multi-directional responses in biospheric variables, e.g., positive and negative anomalies in gross primary production (GPP). In particular, the region of reduced summer ecosystem production does not match the area identified as extreme in the hydrometeorological variables. The reason is that forest-dominated ecosystems in the higher latitudes respond with unusually high productivity to the RHW. Furthermore, the RHW was preceded by an anomalously warm spring, which leads annually integrated to a partial compensation of 54 % (36 % in the preceding spring, 18 % in summer) of the reduced GPP in southern agriculturally dominated ecosystems.

Our results show that an ecosystem-specific and multivariate perspective on extreme events can reveal multiple facets of extreme events by simultaneously integrating several data streams irrespective of impact direction and the variables' domain. Our study exemplifies the need for robust multivariate analytic approaches to detect extreme events in both hydrometeorological conditions and associated biosphere responses to fully characterize the effects of extremes, including possible compensatory effects in space and time.

1 Introduction

One consequence of global climate change is that the intensity and frequency of heatwaves will most likely be increasing in the coming decades (Seneviratne et al., 2012). Heatwaves co-occurring with droughts form so-called compound events, for which we can expect severe impacts on the functioning of land ecosystems (e.g., primary production, von Buttlar et al., 2018) that may affect human well-being (e.g., via reduced crop yields, health impacts) (e.g., Schefran et al., 2012; Reichstein et al., 2013; Lesk et al., 2016). Investigating historical extreme events offers important insights for deriving mitigation strategies in the future.

One well-known example of a compound extreme event is the 2010 western Russian heatwave (RHW). The RHW was one of the most severe heatwaves on record, breaking temper-

ature records of several centuries (Barriopedro et al., 2011). It was accompanied by extensive wild and peat fires with smoke plumes about 1.6 km high at the peak of the heatwave in early August, and estimated emissions of around 77 Tg carbon due to multiple fire events (Guo et al., 2017). Carbon losses due to reduced vegetation activity were estimated to be in the same order of magnitude as losses due to fires (90 Tg, Bastos et al., 2014). The amount of emitted carbon monoxide was almost comparable to the anthropogenic emissions in this region (Konovalov et al., 2011). Approximately 55 000 cases of death have been attributed to health impacts of the RHW (Barriopedro et al., 2011).

The RHW was associated with an atmospheric blocking situation (Matsueda, 2011), which led to a persistent anticyclonic weather pattern in eastern Europe (Dole et al., 2011; Petoukhov et al., 2013; Schubert et al., 2014; Kornhuber et al., 2016).

However, to fully understand the developments and impacts of heatwaves or droughts, apart from hydrometeorological drivers, associated land surface dynamics and feedbacks need to be considered (Seneviratne et al., 2010). For instance, under persistent anticyclonic and dry conditions, land–atmosphere feedbacks are expected to further amplify the magnitude of heatwaves via enhanced sensible heat fluxes, as shown also for the RHW (Miralles et al., 2014; Hauser et al., 2016). These feedback mechanisms highlight the importance of depleted soil moisture to heatwaves. In 2010 a negative soil moisture anomaly contributed to increased temperatures (Hauser et al., 2016). It is a general observation that the combination of anticyclonic weather regimes and initially dry conditions prior to the event amplifies heatwaves in most cases (Quesada et al., 2012).

The direct impacts of such extreme events on ecosystems are manifold. Summer heat and drought typically reduce (or even inhibit) photosynthesis, hence reducing the carbon uptake potential of ecosystems (Reichstein et al., 2013). However, the magnitude of these impacts varies between ecosystems (Frank et al., 2015), and the resulting net effects are still under debate, particularly for heatwaves (Sippel et al., 2018). However, in-depth investigations of a number of individual events such as the European heatwave 2003 (Ciais et al., 2005), the 2000–2004 and 2012 droughts in North America (Schwalm et al., 2012; Wolf et al., 2016), and the RHW (Bastos et al., 2014) agree on an overall tendency towards negative impacts on the carbon accumulation potential.

The RHW has been thoroughly investigated from a hydrometeorological point of view linking the atmospheric blocking to the large-scale positive anomalies in air temperatures and negative anomalies in water availability (e.g., Barriopedro et al., 2011; Rahmstorf and Coumou, 2011). The event has also been well investigated, with an emphasis on the biospheric impacts describing the negative anomalies in ecosystem productivity and related vegetation indices (e.g., Bastos et al., 2014). However, comparing the reports of areas affected by the RHW reveals some discrepancies. Hydrometeorological anomalies point to much larger areas affected compared to biosphere response patterns. Figure 1 shows the zonal evolution of the RHW in both domains. We find that the spatiotemporal patterns of the temperature anomaly do not match the zonal anomaly in vegetation productivity anomalies. Thus, an integrated assessment including the hydrometeorological and biospheric domains simultaneously may further our understanding of the RHW.

The figure reveals an unusually warm period during spring and one longer heatwave during summertime (Fig. 1a). Temperature anomalies exceed more than 10 K in both spring and summer, but they lead to distinctive anomalies in gross primary productivity (GPP). Positive GPP anomalies occur during the spring event, whereas negative GPP anomalies occur during the summer heatwave. The positive GPP response in spring might be a reaction to warmer, more optimal spring temperatures (Wang et al., 2017) possibly accompanied by enough water availability. However, negative GPP anomalies in summer occur only in areas south of 55° N (Fig. 1c), indicating that the GPP response involves far more processes than high temperatures and drought during the unique RHW. As already indicated by Smith (2011), the connection between biosphere and hydrometeorology is much more complex than just a direct one-to-one mapping. Further complicating this issue is the fact that the summer event cannot be investigated without the previous spring as both seasons are inherently related via memory effects in water availability. Increased GPP in spring due to warm temperatures can negatively influence soil moisture and thus GPP during summer (Buermann et al., 2013; Wolf et al., 2016; Sippel et al., 2017). In particular, Buermann et al. (2013) show for North American boreal forests that earlier springs are followed by reduced productivity in summer because of water constraints.

In summary, comparing these two Hovmöller diagrams shows that (1) the affected latitudinal range of the negative GPP anomaly is much smaller than the positive temperature anomaly and (2) the evolution of the summer impacts should consider potential carry-over effects of positive GPP anomalies during spring, as earlier studies showed that earlier spring onset and increased spring GPP may negatively influence soil moisture and thus GPP during summer (Buermann et al., 2013). The objective of this paper is to revisit the RHW and to investigate the GPP response during the spring event and the summer heatwave in detail by investigating spatiotemporal anomalies in hydrometeorological drivers and ecological variables.

This kind of integrated assessment requires a generic methodological approach. Here, we use a multivariate extreme event detection approach that (1) does not differentiate between a positive and negative extreme event, and (2) can equally be applied on any set of time series, regardless of whether they describe the biospheric or hydrometeorological domain. We expect that we can reveal previously overlooked facets in the RHW and discuss whether our approach may fa-

facilitate a broader perspective and improved interpretation of extreme events and their impacts.

2 Methods and data

2.1 Rationale

One approach to detect extreme events like the RHW could be to identify the peaks over some threshold in the marginal distribution of a variable (or its anomaly) of interest. For instance, one could identify values that deviate by more than 2 standard deviations from the variable's mean values (Hansen et al., 2012; Sippel et al., 2015). However, univariate approaches only allow us to characterize an event by, e.g., extremely high temperature anomalies, lack of precipitation, or very low soil moisture but not their compound anomaly. However, from earlier studies (e.g., Miralles et al., 2014; Hauser et al., 2016) we know that more than one variable is involved in the RHW, and a multivariate extreme event detection (i.e., a compound event, Leonard et al., 2014; Zscheischler and Seneviratne, 2017) is more feasible. Multivariate algorithms to detect extreme events are expected to offer more robust detection capabilities when accounting for dependencies and correlations among the selected variables (e.g., Zimek et al., 2012; Bevacqua et al., 2017; Flach et al., 2017; Mahony and Cannon, 2018). Multivariate extreme event detection considers all observable dimensions of the domain simultaneously. With a multivariate approach one may, for instance, detect very rare combinations of variables even if the individual variables are not extreme. In the following, we detect the anomalies in a multivariate variable space in two sets of variables describing (1) the hydrometeorological conditions and (2) the biospheric response. The workflow involves a data pre-processing to compute anomalies, a step for dimensionality reduction to not be biased by redundancies among variables. Based on the reduced data space, an anomaly score is computed that can then be used as a threshold. For various reasons, however, in practice the threshold needs to be computed across multiple spatial grid cells of comparable phenology.

2.2 Data and pre-processing

Our data set for analyzing the hydrometeorological domain includes those variables which we consider to be of particular importance for processes taking place during extreme events in the biosphere based on prior process knowledge (Larcher, 2003) and empirical analysis (von Buttler et al., 2018). The hydrometeorological data set consists of air temperature, radiation, relative humidity (original resolution 0.71° , all three from ERA-INTERIM, Dee et al., 2011), precipitation (original resolution 1° , Adler et al., 2003), and surface moisture (resolution 0.25° , <http://www.gleam.eu>, last access: 12 October 2018, v3.1a, Miralles et al., 2011; Martens et al., 2017). We consider surface moisture to be a hydrometeorological

variable due to its importance for drought detection, although we notice that surface moisture is influenced by biospheric processes. We use gross primary productivity (GPP), latent heat flux (LE), sensible heat flux (H) (resolution of 0.25° , all three from FLUXCOM-RS, Tramontana et al., 2016), and the fraction of absorbed photosynthetic active radiation (original resolution 1 km, FAPAR, moderate resolution imaging spectroradiometer (MODIS) based FAPAR; Myneni et al., 2002) to describe the land surface dynamics.

The selected variables cover the spatial extent of Europe (latitude $34.5\text{--}71.5^\circ$ N, longitude: $-18\text{--}60.5^\circ$ E) and are re-gridded on a spatial resolution of 0.25° from 2001 to 2011 in an 8-daily temporal resolution. The temporal extent is selected as it is covered by all data sets used in the study. To check for differences in land cover types, we estimate the dominant land cover type of the European Space Agency Climate Change Initiative land cover classification on a spatial resolution of 0.25° (original: 300 m). To check for consistency of our findings among other variables (Sect. 3.2), we additionally use terrestrial ecosystem respiration (TER) and net ecosystem productivity (NEP, both originating from FLUXCOM-RS, Tramontana et al., 2016).

The actual event detection is realized on the anomalies of these data sets. To compute the anomalies, for each variable under consideration, we first estimate the seasonality as a smoothed median seasonal cycle per grid cell. We use the median instead of the mean as it is less susceptible to outliers. We then subtract these seasonal cycles from each variable and year to obtain a multivariate data cube of anomalies (Fig. 2, step 1). Small data gaps are set to zeros to ensure that they are not detected as anomalies. The gap filling is necessary for a multivariate detection approach as there are many more cases in which one variable is missing in the multivariate cube compared to a univariate data stream.

2.3 Feature extraction and anomaly detection

We use a multivariate anomaly detection algorithm proposed by Flach et al. (2017) and apply it separately to two sets of variables for the biosphere and hydrometeorology. The method expects a multivariate set of anomalies and projects them to a reduced space via principal component analysis, retaining a number of principal components that explain more than 95% of the variance (Fig. 2, step 3b). This procedure accounts for linear correlations in the data only by removing redundancies among the variable anomalies.

We compute an anomaly score via kernel density estimation (KDE, Parzen, 1962; Harmeling et al., 2006) in the reduced anomaly space (Fig. 2, step 4). KDE showed very good performance among different other options to detect multivariate anomalies in previous experiments (Flach et al., 2017). One strength of KDE is that it considers nonlinear dependencies among dimensions (Fig. 3). The anomaly scores are transformed into normalized ranks between 1.0 (very anomalous, data point in the margins of the multivariate dis-

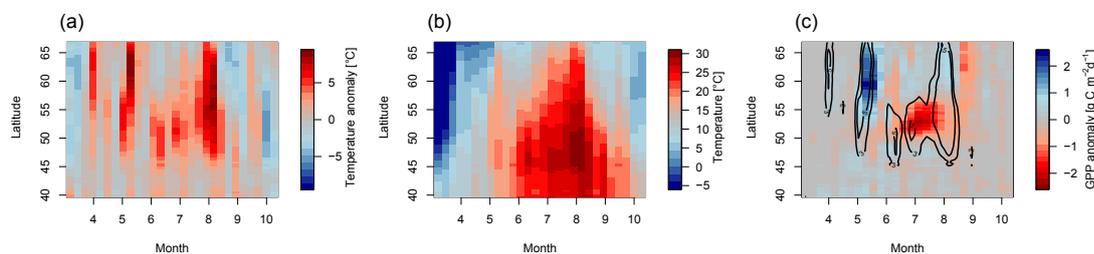


Figure 1. Longitudinal average (30.25 to 60.0° E) of (a) temperature anomalies (reference period: 2001–2011), (b) absolute temperature, and (c) GPP anomalies in 2010 with a contour of temperature anomalies (+3, +5 K).

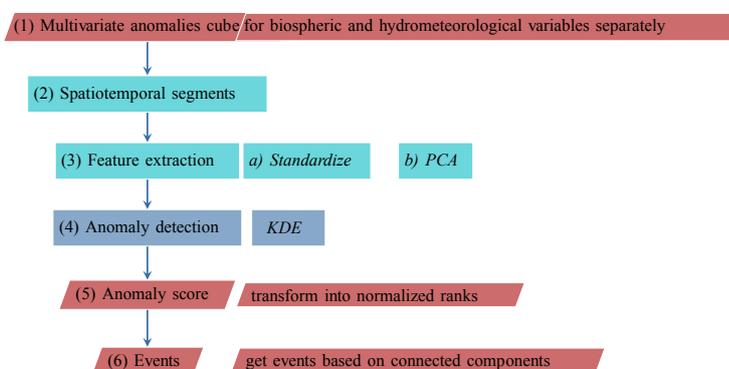


Figure 2. Data processing for detecting multivariate anomalies.

tribution) and 0.0 (completely normal, data point in the dense region of the multivariate distribution; Fig. 2, step 5). In this univariate index of compound extremes, it is legitimate to use a classical threshold that can be intuitively analyzed. However, to avoid an equal spatial distribution of event occurrences we do not apply this multivariate anomaly detection per pixel, but rather by region.

2.4 Spatiotemporal segmentation

The spatiotemporal segmentation aims to identify spatial areas of comparable phenology, climate, and seasonality. To identify these regions, we follow the methodology described by Mahecha et al. (2017) and extend it to the multivariate case. The main idea is that the (now spatial) principal components of the mean seasonal cycles can be used for classifying regions according to their characteristic temporal dynamics.

The procedure for extracting spatial segments of similar grid cells works as follows (for a detailed description, see Mahecha et al., 2017).

1. We estimate the median seasonal cycle in each grid cell and of each variable individually and standardize the median seasonal cycles to zero mean and unit vari-

ance to get the cycles comparable across different units (Fig. 4, step 1).

2. To remove the effect of different phasing (similar but only lagged seasonal cycles), we sort the median seasonal cycles according to a variable showing a strong seasonality, which is temperature in our case. Thus, we memorize how to bring temperature in a sorted increasing or decreasing order (the “permutation” of temperature) and apply the same permutation to the other median seasonal cycles (Fig. 4, step 2). We prepare the data for dimensionality reduction by concatenating the seasonal cycle of all variables to a matrix seasonal cycles \times space. We apply a principal component analysis (PCA) to reduce the dimension of the concatenated median seasonal cycles.
3. We select locations (grid cells) of similar phenology and climate by dividing the orthogonal principal component subspace into equally sized bins (Fig. 4, step 3). We used $N_{PC} = 4$ components in this step, explaining 71 % of variance. The bins are sufficiently small compared to the length of the principal components to ensure a fine binning of very similar phenology and climate.

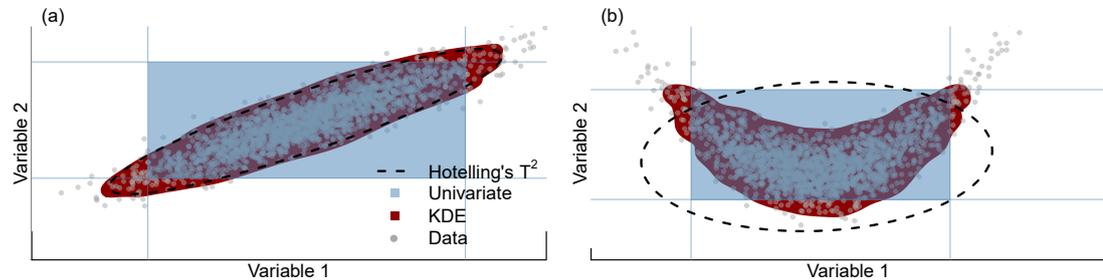


Figure 3. Illustration of the multivariate anomaly detection algorithm with two variables. The data have (a) linear dependencies (multivariate normal) and (b) a nonlinear dependency structure. Univariate extreme event detection (peak-over-threshold in the marginal distribution of a variable) does not follow the shape of the data, whereas algorithms assuming a multivariate normal distribution (Hotelling's T^2 , Lowry and Woodall, 1992) are suitable for case (a); kernel density estimation (KDE) gets the shape of the data in both cases (a) and (b); 5 % extreme anomalies are outside the shaded areas (region of “normality”) for all three algorithms.

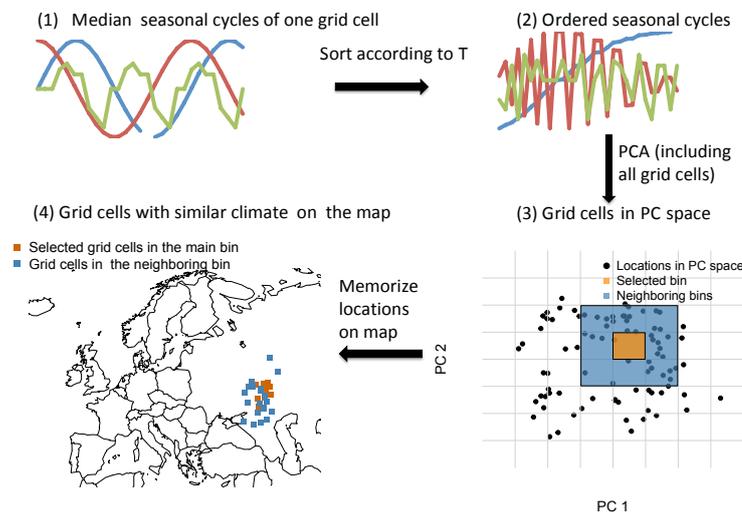


Figure 4. Illustration of the spatial segmentation procedure with two principal components.

4. We compute the multivariate anomaly score in an overlapping moving window for all grid cells that fall into one of the bins (the central bin and the neighboring bins, Fig. 4, step 4).

A final detail to consider is the effect of changing seasonal variance (temporal heteroscedasticity). These patterns lead to detecting extreme events predominantly during the high-variance seasons (i.e., summer times). To avoid seasonal biases in the extreme event detection, we additionally apply the entire anomaly detection scheme to seasonally overlapping moving windows across years.

Within the spatiotemporal segmentation procedure, we ensure that the number of observations is at least 198 (9 time

steps \times 11 years, at least one spatial replicate). To reunify the spatiotemporal segments, we assign the normalized anomaly scores temporally to the time step in the center of the temporal moving window and spatially to the grid cell in the central bin of similar climate and phenology.

2.5 Statistics of extreme events

We assume that 5 % of the data are anomalous in each overlapping spatiotemporal segment and convert the anomaly scores into binary information. However, the main results of compensation effects are not sensitive to this threshold selection (Appendix Table A1, varying the threshold between 1 % and 10 %). To compute statistics based on the

spatiotemporal structure of each extreme event, we follow an approach developed by Lloyd-Hughes (2011) and Zscheischler et al. (2013) and compute the connections between spatiotemporal extremes if they are connected within a $3 \times 3 \times 3$ (long \times lat \times time) cube. Each connected anomaly is considered as a single event (Fig. 2, step 6). In this way, we observe event-based statistics, i.e., affected area (km^2), affected volume ($\text{km}^2 \text{days}^{-1}$), centroids of the area, and histograms of the single variable anomalies stratified according to different ecosystem types (land cover classes). Furthermore, we observe the response of individual variables to the multivariate event by computing the area weighted sum of the variable during the event in which the variable of interest is positive relative to the seasonal cycle (res^+) or negative (res^-), respectively. For many biospheric variables, one expects a mainly negative response to hydrometeorological extreme events like heatwaves or droughts (Larcher, 2003; von Buttlar et al., 2018). Thus, we define compensation of a specific variable to be the absolute fraction of res^+ from res^- . The balance of a variable is the sum of res^+ and res^- . Centroids of res^+ and res^- are computed as the average of the affected longitudes, latitudes, and time period, weighted with the number of affected grid cells at this longitude, latitude, and time period, and its respective anomaly score. They are used to compute the spatial and temporal distance between res^+ and res^- . Affected area, volume, response, and centroids take the spherical geometry of the Earth into account by weighting the affected grid cells with the cosine of the respective latitude.

3 Results

3.1 Extreme events in western Russia in 2010

We identify two multivariate extreme events in the set of hydrometeorological variables in western Russia 2010, based on the spatiotemporal connectivity. The two extreme events are separated by approximately 1 week of normal conditions towards the end of May.

- Hydrometeorological spring event: anomaly of the hydrometeorological variables in western Russia during May ranging from longitude 30.25 to 60.0° E, latitude $\geq 55^\circ$ N. (Fig. 5a, b)
- Hydrometeorological summer event: anomaly of the hydrometeorological variables in western Russia, June to August, ranging from longitude 28.75 to 60.25° E, latitude 48.25 to 66.75° N. This event is usually referred to as Russian heatwave (RHW) 2010 (e.g., Barriopedro et al., 2011; Rahmstorf and Coumou, 2011) (Fig. 5c, d).

Both multivariate hydrometeorological anomalies partly overlap with a multivariate anomaly in the set of biosphere variables (biospheric spring event and biospheric summer

event). Of specific interest is that the area affected by anomalous hydrometeorological summer conditions is remarkably larger than the one detectable in the biospheric variables (biospheric summer event, 2.4×10^6 vs. $1.1 \times 10^6 \text{ km}^2$, Table 1). This fact already indicates that biosphere responses are more nuanced than the hydrometeorological events and do not simply follow the extent of the hydrometeorological anomaly. As indicated, e.g., also by Smith (2011), a hydrometeorological extreme event does not necessarily imply an extreme response.

3.1.1 Hydrometeorological events

As GPP is a key determinant of ecosystem–atmosphere carbon fluxes, we focus on the gross primary productivity (GPP) response to the multivariate hydrometeorological anomaly: we find that the GPP response is entirely positive during the short-lasting hydrometeorological spring event (+17.8 Tg C, Table 1), while it is mainly negative during the summer event (+8.8, −49 Tg C, Table 1). A part of the GPP summer losses (18 %) associated with the RHW in the southern region are instantaneously reduced by over-productive vegetation in the higher latitudes, which are hit by the extreme event. Please note that the carbon balance in summer accounts for the GPP response to the same hydrometeorological extreme event, namely the RHW, which leads to contrasting responses in adjacent regions. If we estimate the annually integrated effect of the anomalies, another 36 % of the carbon losses are compensated during spring in higher latitudes. We did not find extreme events after summer, which implies a fast recovery of vegetation activity after summer. Integration over the spring and summer events thus equals the annual integration. Overall, we find that 54 % of the negative GPP anomalies are compensated either because of the positive spring anomalies or across ecosystems hit by the same event during summer. These compensation effects reduce the negative carbon impact of integrated annual hydrometeorological events from −49.0 to −24 Tg C in total (Table 1). We want to emphasize that the negative impact of the RHW in terms of GPP is just reduced, and still negative in total.

3.1.2 Biospheric events

Moving the focus to the multivariate biosphere events (biospheric spring and biospheric summer event), which overlap with the hydrometeorological events, we find that GPP responses based on the biospheric spring event are almost entirely positive (+33.8 Tg C), and based on the biospheric summer event almost entirely negative (−82.6 Tg C). If we consider the annually integrated effect of the anomalies, spring carbon gains are estimated to offset 41 % of the subsequent carbon losses in summer (56 days earlier) in the higher latitudes (514 km distance of the centroids, Table 1). To further examine these findings, we check for these kinds of compensation effects among different variables and another GPP

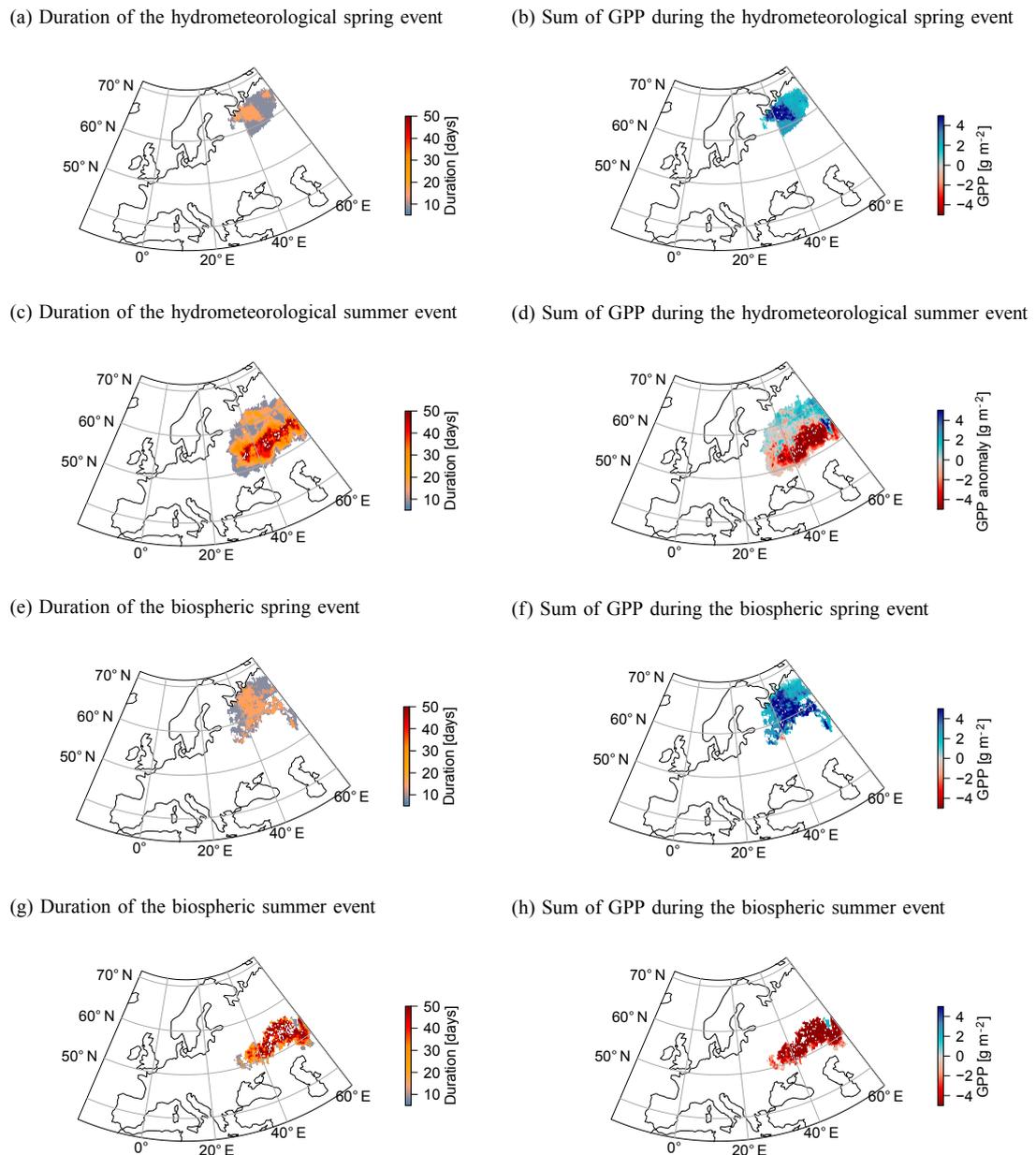


Figure 5. Left column: temporal duration of the (a) hydrometeorological spring event, (c) hydrometeorological summer event, and biospheric events (e, g). Right column: corresponding GPP response, i.e., the sum of deviations from the seasonal cycle during the event for the (b) hydrometeorological spring event, (d) hydrometeorological summer event, and biospheric events (f, h). While the GPP response during the hydrometeorological spring event is entirely positive (more productive than usual, b), GPP response during the hydrometeorological summer event differs between higher latitudes ($> 55^{\circ}$ N, short-lasting, positive) and lower latitudes (long-lasting, negative).

Table 1. Statistics of the extreme events, based on their spatiotemporal connectivity structure: affected area, affected volume, positive and negative GPP response ($\text{res}^{+/-}$) to the event, compensation of the negative response (comp.), as well as average spatial and temporal distance between the parts of the events with positive and negative responses.

event	area (km ²)	volume (km ² days ⁻¹)	GPP comp.	$\text{res}_{\text{GPP}}^{+}$	$\text{res}_{\text{GPP}}^{-}$	spatial (km)	temporal (d)
hydrometeorological							
spring	0.77×10^6	0.81×10^7	–	17.8 Tg	–		
summer	2.44×10^6	5.79×10^7	0.18	8.8 Tg	–49.0 Tg	499	–4
integrated	3.29×10^6	6.60×10^7	0.56	26.6 Tg	–49.0 Tg	452	–34
biospheric							
spring	1.25×10^6	1.48×10^7	117.04	33.8 Tg	–0.3 Tg	756	–16
summer	1.06×10^6	4.22×10^7	0.00	0.4 Tg	–82.4 Tg	962	50
integrated	2.28×10^6	5.70×10^7	0.41	34.2 Tg	–82.7 Tg	514	–56

data set in the following section. Note that the data set of biosphere variables includes GPP itself. Computing the responses based on the extent of the biospheric event is nevertheless useful, as an extreme event in the biosphere variables is not exclusively restricted to extreme conditions in the hydrometeorological conditions (Smith, 2011).

3.2 Compensation in other data sets and variables

The annually integrated compensation effect in GPP is highly consistent among different variables. For instance, NEP (excluding fire) shows such a kind of compensation, but also FAPAR and LE (Table 2). Sensible heat flux, on the other hand, is high during the hydrometeorological summer event (biospheric summer event) as well as the hydrometeorological spring event (biospheric spring event), as expected for strong positive temperature anomalies. However, some of the remote sensing data products might be affected by high fire induced aerosol loadings during the heatwave that affect atmospheric optical thickness (e.g., Guo et al., 2017; Kononov et al., 2011). Exploring an almost entirely climate-driven GPP product (FLUXCOM RS + METEO, Jung et al., 2017), we also find the integrated compensation effect, although much less pronounced (Appendix Fig. B1). Thus, we are confident that the observed compensation effect is not related to the optical thickness during the RHW.

3.3 Influence of vegetation types

In Fig. 6 we present the histograms of GPP anomalies for different land cover classes (forests, grasslands, and crops) based on the hydrometeorological spring event and hydrometeorological summer event (biospheric spring event and biospheric summer event, respectively, Fig. C1) to highlight two aspects: first, during the spring event (hydrometeorological spring or biospheric spring), forests react almost entirely with positive GPP anomalies (Fig. 6a). Forests in this region are energy-limited, so the timing of the extreme event leads

to hydrometeorological conditions (e.g., positive temperature anomalies in spring, more incoming radiation accompanied by enough water availability) which are favorable for vegetation productivity, as absolute spring temperatures are still below the temperature optimum of GPP (Fig. 8a, Wolf et al., 2016; Wang et al., 2017).

Second, during the hydrometeorological summer event, we observe positive to neutral GPP responses in forests, whereas crops and grasslands react strongly negatively (Fig. 6b). The positive vs. negative GPP responses almost entirely reflect the map of dominant vegetation types (forest vs. agricultural ecosystems, Fig. 7). However, different vegetation types exhibit a transition from higher latitudes (predominantly forest ecosystems) to lower latitudes (dominated by agricultural ecosystems). Thus, the different responses of vegetation types might be confounded by the fact that absolute temperatures also follow a latitudinal gradient (Fig. 1b). Absolute temperatures for agricultural ecosystems are higher and far beyond the temperature optimum of GPP (Fig. 8c). Additionally, agricultural ecosystems are drying out in summer (low soil moisture, Fig. 8c). In contrast, forest-dominated ecosystems at higher latitudes experience temperatures just slightly above the temperature optimum of GPP, accompanied by high soil moisture (Fig. 8b). The response of forest ecosystems partly reflects a latitudinal gradient: forest ecosystems in the lower latitudes react positively to the spring temperature anomaly and then tend to react more negatively to the summer heatwave than forest ecosystems in higher latitudes. Forest ecosystems in higher latitudes are still productive in terms of GPP during the peak of the heatwave (Fig. 9). We find negligible anomalies in autumn for both ecosystems, which implies a fast recovery after the heatwave.

To disentangle the variable importance of the different confounding factors, we run a simple linear regression model which tries to explain GPP as a function of the hydrometeorological driver variables (temperature, precipitation, radiation, and surface moisture, including their anomalies and

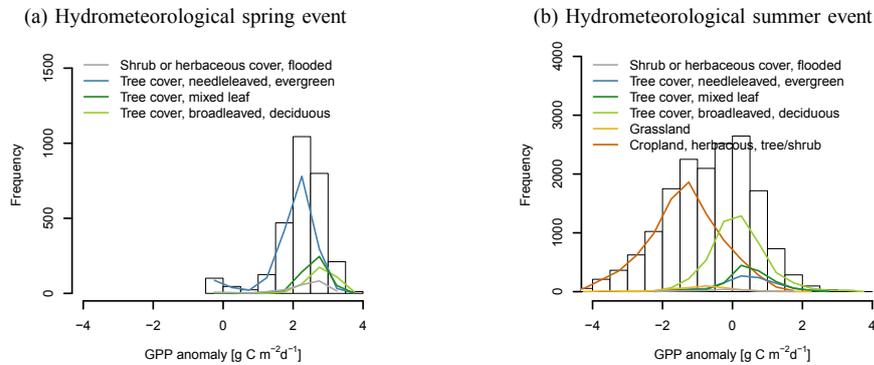


Figure 6. Histogram of GPP anomalies (reference period: 2001–2011) for different land cover classes based on the spatiotemporal extent of (a) the hydrometeorological spring event and (b) the hydrometeorological summer event. Bars denote the sum of all vegetation classes.

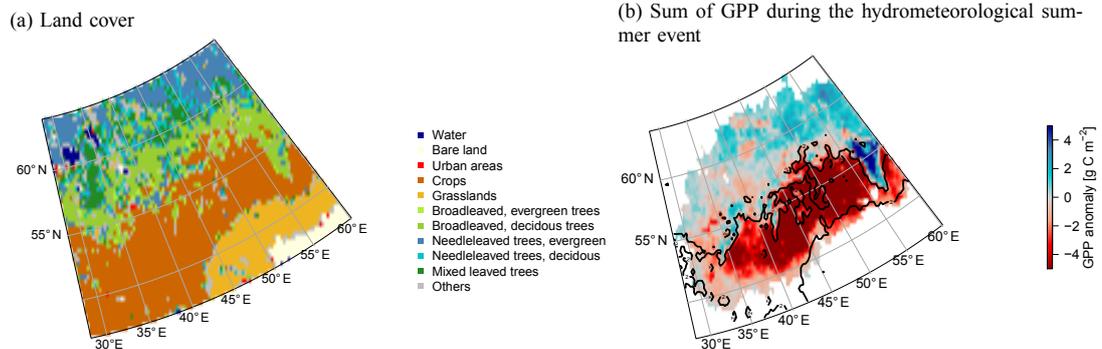


Figure 7. (a) Dominant land cover classes of a spatial extent of the RHW. (b) The boundaries of the different ecosystem types (forest-dominated ecosystems vs. agriculture-dominated ecosystems, denoted by the black contour line) match the observed patterns of the GPP response (reference period for the calculating anomalies: 2001–2011) during the hydrometeorological summer event.

absolute values), as well as vegetation type, duration and latitude (Appendix D). We use an algorithm following Chevan and Sutherland (1991) which extracts the independent contribution of the variable importance related to this particular variable regardless of the model complexity or dependencies among variables. The model reveals from a statistical point of view that vegetation type and the latitudinal gradient are the most important variables explaining GPP during the summer event, followed by the hydrometeorological drivers. Access to deeper water and soil type as well as nonlinear feedbacks are factors which are not represented in the model but might explain the high importance of latitude. Apart from vegetation type being important for the GPP response, underlying water use efficiency (calculated according to Zhou et al. (2014) is consistently higher in forest-dominated ecosystems compared to agriculture-dominated ecosystems (Appendix

Fig. E1a), and higher evaporative fraction in forest ecosystems during the peak of the heatwave (Appendix Fig. E1b).

4 Discussion

In this paper we show that the hydrometeorological extreme events affecting western Russia in spring and summer 2010 do not directly map to the observed vegetation responses. Positive to neutral GPP responses prevail in higher latitudes during summer, whereas strong negative impacts on GPP can be found in lower latitudes. We interpret this effect by different water management strategies of forest vs. agricultural ecosystems (Teuling et al., 2010; van Heerwaarden and Teuling, 2014) that meet a general latitudinal temperature gradient. Apart from a more efficient water usage of forest-dominated ecosystems, access to deeper soil water might be another reason for ecosystem-specific responses (Fan et al.,

Table 2. Negative responses to the RHW are partly compensated based on the integrated biospheric or hydrometeorological events in 2010. The finding is consistent over different variables and data sets.

variable	hydrometeorological events			biospheric events		
	res ⁺	res ⁻	comp. (%)	res ⁺	res ⁻	comp. (%)
NEP	17.53 Tg	-34.03 Tg	51.5	23.45 Tg	-48.49 Tg	48.4
LE	19.90 Tg	-53.97 Tg	36.9	16.34 Tg	-102.81 Tg	15.9
FAPAR	1.89	-4.03	47.0	2.52	-6.61	38.1
TER	18.97 Tg	-11.06 Tg	171.4	13.71 Tg	-23.43 Tg	58.5

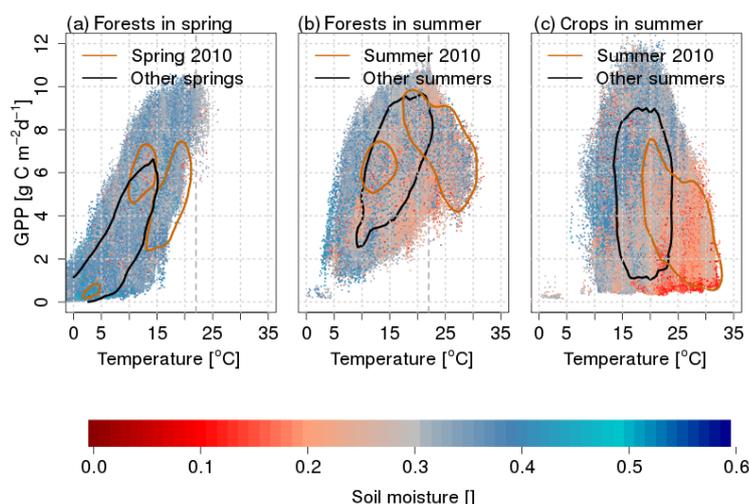


Figure 8. Temperature optimality for GPP in (a) forests during spring, (b) forests during summer, and (c) crops during summer. Contour lines enclose 75 % of the data points.

2017; Yang et al., 2016). Note that the latitudinal temperature gradient alone might explain differences in the response within ecosystems in summer and between spring and summer, but does not sufficiently explain differentiated GPP responses in summer among different ecosystems (predominantly forest vs. agricultural ecosystems).

Another important aspect is that the combination of the anomalous spring and the unique heatwave in summer might be inherently connected via land surface feedbacks. Buermann et al. (2013) showed that warmer springs going hand in hand with earlier vegetation activity negatively affect soil moisture in summer, and thereby vegetation activity. It is a general observation that warm and dry springs enhance summer temperatures during droughts, which suggests the presence of soil-moisture temperature feedbacks across seasons (Haslinger and Blöschl, 2017). In the case of the Russian heatwave 2010, soil moisture was one of the main drivers (Hauser et al., 2016), hand in hand with persistent atmospheric pressure patterns (Miralles et al., 2014). Thus, we suspect that the spring event is connected to the summer heat-

wave in 2010, if not setting the preconditions for a heatwave of this unique magnitude.

The integration of the carbon balance over spring and summer might be justified by assumed connections between spring and summer as outlined before. However, we would like to note that common annual integration and assessment of compensatory effects on the carbon balance over events during the growing season equal the integration over spring and summer for this particular case, as we did not find any events after summertime. The absence of events after the summer heatwave implies a fast recovery of the ecosystems.

Compensations of parts of the negative impacts on the carbon balance during hydrometeorological extreme events have been reported in earlier studies. On the one hand, Wolf et al. (2016) report that a warm spring season preceding the 2012 US summer drought reduced the impact on the carbon cycle. Yet on the other hand, the increased spring productivity amplified the reduction in summer productivity by spring–summer carry-over effects via soil moisture depletion: higher spring productivity leads to higher water con-

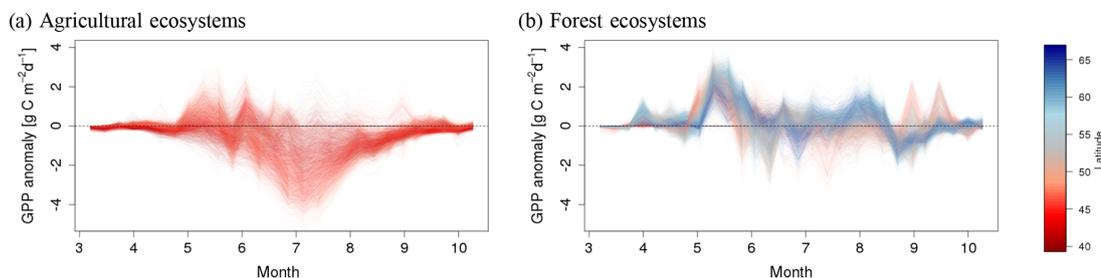


Figure 9. Temporal evolution of the GPP anomaly (reference period: 2001–2011) for (a) agricultural ecosystems and (b) forest ecosystems, colored according to the latitude.

sumption in spring. The high water additionally consumed during spring reduces the water availability in summer and thereby affects productivity during the following summer. However, it remains unclear whether this observation was a singular case or whether it could become a characteristic pattern to be regularly expected in a warmer world. In this study, we provide some evidence for presumed comparable effects. In contrast to the discussion in Wolf et al. (2016), enhanced productivity does not exclusively occur temporally, i.e., spring partly compensates for summer losses, but rather spatially adjacent forest ecosystems are reducing the negative impact of agricultural ecosystems on the carbon balance. Spatially adjacent ecosystems partly compensating carbon losses due to drought or heatwaves have been observed earlier, e.g., in mountainous ecosystems that respond differently than lowlands during the European heatwave 2003 (Reichstein et al., 2007).

Following up on compensatory effects, Sippel et al. (2017) use ensemble model simulations to disentangle the contribution of spring compensation vs. spring–summer carry-over effects on a larger scale. They show that in general warm springs compensate for parts of summer productivity losses in Europe, whereas spring–summer carry-over effects are constantly counteracting by enhancing summer losses. Also, Mankin et al. (2017, 2018) note that increased spring productivity with spring–summer carry-over effects can be observed in Earth system models. We can confirm the general finding that spring partly compensates for summer productivity losses in observations for our case study on the RHW. Without using model simulations it is difficult to quantify spring–summer carry-over effects via soil moisture depletion. In the case of the RHW only very few areas are anomalously productive in terms of GPP in spring and unproductive in summer as well. Thus, we suspect that exclusively temporal spring–summer carry-over effects play a rather small role for the RHW. However, we also emphasize that longer-term effects, such as effects in subsequent years through species changes (Wagg et al., 2017), have not been considered in the

present study and likely remain hard to quantify beyond dedicated experiments.

The RHW is among the best studied extreme events in the Northern Hemisphere. However, the enhanced productivity of northern forests which diminishes the negative carbon impact of the RHW as reported in this study has only received marginal attention so far. For instance, Wright et al. (2014) mention positive NDVI anomalies in spring 2010, but then focus largely on productivity losses in the Eurasian wheat belt. Similarly, Bastos et al. (2014) focus on a spatial extent of the biosphere impacts that only partly includes forest ecosystems at higher latitudes. Our estimation of carbon losses due to decreased vegetation activity (82 TgC) is comparable to the one of Bastos et al. (2014) (90 TgC). Similar to the results of our study, Yoshida et al. (2015) report reductions in photosynthetic activity in agriculture-dominated ecosystems during the RHW, but only small to no reductions in forest ecosystems during summertime. However, their interpretations focus on the summer heatwave. Nevertheless, re-evaluating impact maps (published, e.g., in Wright et al., 2014; Yoshida et al., 2015; Zscheischler et al., 2015) in the light of our findings suggests that their evidence supports the presence of contrasting responses, differing among ecosystems during the RHW. When it comes to extreme events, the general tendency in many existing studies is naturally to focus on negative impacts as they are of particular interest for society (Bastos et al., 2014; Wright et al., 2014; Yoshida et al., 2015; Zscheischler et al., 2015).

5 Conclusions

We re-analyzed biospheric and hydrometeorological conditions in western Russia in 2010 with a generic spatiotemporal multivariate anomaly detection algorithm. We find that the hydrometeorological conditions and the biospheric responses exhibit two anomalous extreme events, one in late spring (May) and one over the entire summer (June, July, August), covering large areas of western Russia. For the summer event, we find that the spatially homogeneous anomaly

pattern (characterized by high solar radiation and temperature, low relative humidity, and precipitation) translates into a bimodal and contrasting biosphere response. Forest ecosystems in higher latitudes show a positive anomaly in gross primary productivity, while agricultural systems decrease their productivity dramatically.

If we consider the annually integrated effect of the anomalous hydrometeorological conditions in 2010, we find that forest ecosystems reduce the negative impact of the productivity losses experienced in agricultural ecosystems by 54 % (36 % during spring, 18 % during summer). Please note that the annually integrated impact of the 2010 events on the carbon balance stays strongly negative. Our findings do not alleviate the consequences of extreme events for food security in agricultural ecosystems.

From a methodological point of view, this study emphasizes the importance of considering the multivariate nature of anomalies. From this study, we learn that it is insightful to consider the possibility of both negative as well as positive impacts and to assess their annually integrated statistics. Although the integrated impact of gross primary production on the hydrometeorological conditions in 2010 is strongly negative, it is important to notice the partial compensatory effects due to differently affected ecosystem types, as well as timing of the extreme events.

Data availability. The data are available and can be processed at <https://www.earthsystemdatalab.net/index.php/interact/data-lab/>, last access: 15 October 2018.

Appendix A: Sensitivity of the threshold selection

Table A1. Compensation effects of the integrated hydrometeorological events (spring and summer) are not sensitive to varying the threshold for extreme event detection between 93 % and 99 % (7 % and 1 % of extreme data in each spatiotemporal segment). A slight tendency towards more pronounced compensation effects can be seen for the 90 % threshold. Such a kind of enhancing the positive response is expected for lower thresholds, as the hydrometeorological conditions are not perceived as “extreme” anymore.

Threshold	Compensation [%]				
	90 %	93 %	95 %	97 %	99 %
GPP	65	53	54	58	55
NEP	60	52	52	51	46
LE	49	36	37	38	32
FAPAR	70	46	47	50	50
TER	150	147	171	191	197

Appendix B: Comparison with METEO + RS

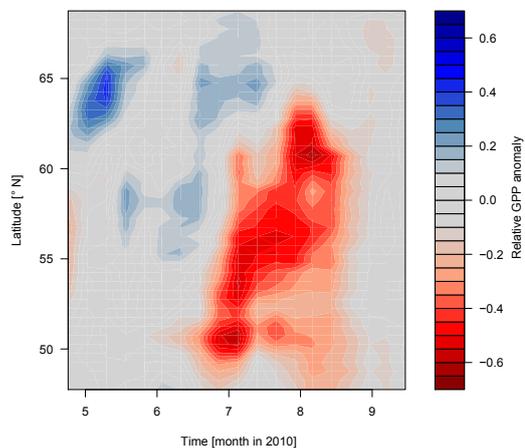


Figure B1. The longitudinal (30.25–60.25° E) average of the GPP anomalies during the RHW 2010, based on the Climate Research Unit observation-based climate variables (CRUNCEPv6, New et al., 2000) driven GPP product originating from FLUXCOM RS+METEO. Jung et al. (2017) show similar but weaker compensation effects; 28 % of the negative GPP response to the RHW is compensated based on the shown latitude–longitude subset.

Appendix C: Biosphere response

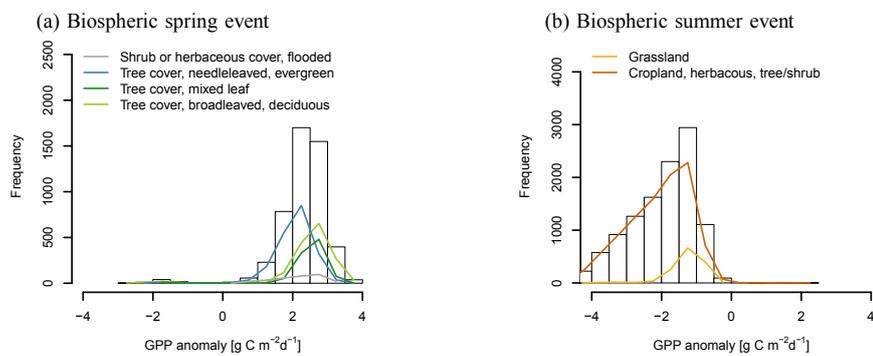


Figure C1. Histogram of GPP anomalies (reference period: 2001–2011) for different land cover classes constrained by (a) biospheric spring and (b) biospheric summer event.

Appendix D: Factors explaining the GPP response

As several factors might contribute to the GPP response to the hydrometeorological anomalies in spring and summer 2010, we assume that a linear model can partly explain the variance in GPP and improve our understanding of the extreme events in spring and summer via the variable importance of the model. Thus, we model GPP of all pixels during spring and separately during summer as a function of the factors temperature (T), precipitation (P), global radiation (Rg), soil moisture (SM), and their corresponding anomalies. We include land cover type, duration, and latitude as possible drivers of the full model (spring $R^2 = 0.86$, summer $R^2 = 0.35$). We use a variable importance partitioning algorithm according to Chevan and Sutherland (1991) to get the variable importance of the full model while accounting for redundancies (e.g., dependencies) among the factors and model complexity. The partitioning algorithms compute all possible combinations of submodels (excluding one or several factors). By combining the differences of R^2 measures

of the submodels in an intelligent way (for more details, see Chevan and Sutherland, 1991), it is possible to partition the total importance of each variable into an independent contribution and a joint contribution. Results show that the hydrometeorological spring event is mainly a response to very favorable hydrometeorological conditions (higher radiation due to the lack of precipitation, high absolute spring temperatures beyond the optimum of GPP), which is indicated by the high independent contributions of the variables. As only forest ecosystems are affected, vegetation type plays a minor role (Fig. D1a). The lower explanatory power of the model for the summer event indicates that there are potentially non-linear feedback loops not captured by the model or factors playing a role, which we did not include in the model. One of the latter candidates is the access to deeper water, also indicated by the high variable importance of latitude. Apart from latitude vegetation type is the most important factor driving the GPP response during the summer event (Fig. D1b).

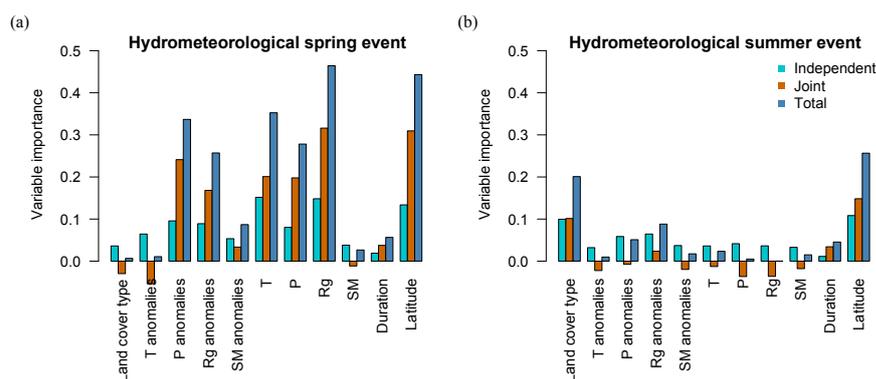


Figure D1. Independent, joint, and total contribution of the factors explaining (a) GPP response during the hydrometeorological spring event and (b) during the hydrometeorological summer event. Used abbreviations are T (temperature), P (precipitation), Rg (radiation), and SM (soil moisture).

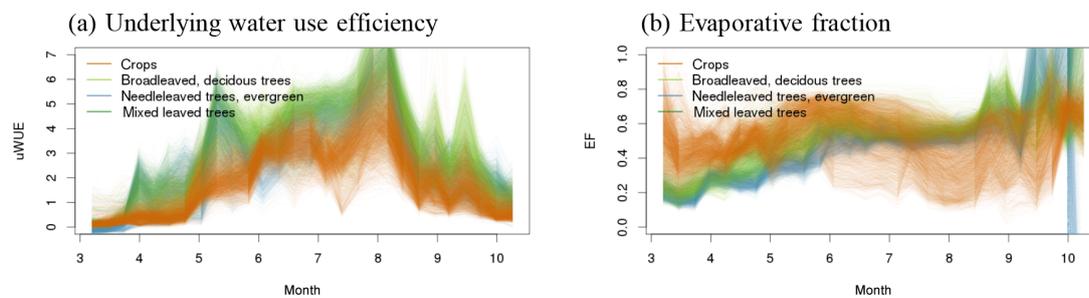
Appendix E: Water use efficiency and evaporative fraction of different land cover types

Figure E1. (a) Underlying water use efficiency (uWUE) and (b) evaporative fraction (EF) of the area affected by the RHW in 2010. uWUE is calculated according to Zhou et al. (2014) including vapor pressure deficit. In contrast to WUE, uWUE attempts to correct for differences in temperature and vapor pressure deficit to a certain degree.

Author contributions. MF and MDM designed the study in collaboration with SS, FG, ABa, ABr, and MR. MF conducted the analysis and wrote the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research received funding by the European Space Agency (project “Earth System Data Lab”) and the European Union’s Horizon 2020 research and innovation program (project “BACI”, grant agreement no. 64176). The authors are grateful to the FLUXCOM initiative (<http://www.fluxcom.org>, last access: 12 October 2018) for providing the data. MF acknowledges support by the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS). Furthermore, the authors would like to thank Sebastian Bathiany for crucial discussions on the topic, Jürgen Knauer for his expertise on water use efficiency, Julia Kiefer for her kind language check, as well as Victor Brovkin and Sophia Walther for helping to improve the manuscript. Two anonymous reviewers provided valuable suggestions for improvement.

The article processing charges for this open-access publication were covered by the Max Planck Society.

Edited by: Paul Stoy

Reviewed by: two anonymous referees

References

- Adler, R. F., Huffman, G. F., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvins, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), *J. Hydrometeorol.*, 4, 1147–1167, 2003.
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and Garcia-Herrera, R.: The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe, *Science*, 332, 220–224, 2011.
- 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, <https://doi.org/10.5194/bg-11-3421-2014>, 2014.
- Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), *Hydrol. Earth Syst. Sci.*, 21, 2701–2723, <https://doi.org/10.5194/hess-21-2701-2017>, 2017.
- Buermann, W., Bikash, P. R., Jung, M., Burn, D. H., and Reichstein, M.: Earlier springs decrease peak summer productivity in North American boreal forests, *Environ. Res. Lett.*, 8, 024027, <https://doi.org/10.1088/1748-9326/8/2/024027>, 2013.
- Chevan, A. and Sutherland, M.: Hierarchical Partitioning, *The American Statistician*, 45, 90–96, 1991.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevalier, 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, 2005.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., Quan, X.-W., Xu, T., and Murray, D.: Was there a basis for anticipating the 2010 Russian heat wave?, *Geophys. Res. Lett.*, 38, L06702, <https://doi.org/10.1029/2010GL046582>, 2011.
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic regulation of plant rooting depth, *P. Natl. Acad. Sci. USA*, 82, 201712381, <https://doi.org/10.1073/pnas.1712381114>, 2017.
- Flach, M., Gans, F., Brenning, A., Denzler, J., Reichstein, M., Rodner, E., Bathiany, S., Bodesheim, P., Guaniche, Y., Sippel, S., and Mahecha, M. D.: Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques, *Earth Syst. Dynam.*, 8, 677–696, <https://doi.org/10.5194/esd-8-677-2017>, 2017.
- Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., 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., Zavalá, M. A., and Zscheischler, J.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, *Global Change Biol.*, 21, 2861–2880, 2015.
- Guo, M., Li, J., Xu, J., Wang, X., He, H., and Wu, L.: CO₂ emissions from the 2010 Russian wildfires using GOSAT data, *Environ. Pollut.*, 226, 60–68, 2017.
- Hansen, J., Sato, M., and Ruedy, R.: Perception of climate change, *P. Natl. Acad. Sci. USA*, 109, E2415–E2423, 2012.
- Harmeling, S., Dornhege, G., Tax, D., Meinecke, F., and Müller, K.-R.: From outliers to prototypes: Ordering data, *Neurocomputing*, 69, 1608–1618, 2006.
- Haslinger, K. and Blöschl, G.: Space-Time Patterns of Meteorological Drought Events in the European Greater Alpine Region Over the Past 210 Years, *Water Resour. Res.*, 53, 9807–9823, 2017.
- Hauser, M., Orth, R., and Seneviratne, S. I.: Role of soil moisture versus recent climate change for the 2010 heat wave in Russia, *Geophys. Res. Lett.*, 43, 2819–2826, 2016.
- Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneeth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Pa-

- pale, D., Poulter, B., Ráduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO₂ sink changes to temperature, *Nature*, 541, 516–520, 2017.
- Konovalov, I. B., Beekmann, M., Kuznetsova, I. N., Yurova, A., and Zvyagintsev, A. M.: Atmospheric impacts of the 2010 Russian wildfires: integrating modelling and measurements of an extreme air pollution episode in the Moscow region, *Atmos. Chem. Phys.*, 11, 10031–10056, <https://doi.org/10.5194/acp-11-10031-2011>, 2011.
- Kornhuber, K., Petoukhov, V., Petri, S., Rahmstorf, S., and Coumou, D.: Evidence for wave resonance as a key mechanism for generating high-amplitude quasi-stationary waves in boreal summer, *Clim. Dynam.*, 49, 1961–1979, 2016.
- Larcher, W.: *Physiological plant ecology: ecophysiology and stress physiology of functional groups*, Springer Science & Business Media, Berlin, 2003.
- Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M.: A compound event framework for understanding extreme impacts, *WIREs Clim. Change*, 5, 113–128, 2014.
- Lesk, C., Rowhani, P., and Ramankutty, N.: Influence of extreme weather disasters on global crop production, *Nature*, 529, 84–87, 2016.
- Lloyd-Hughes, B.: A spatio-temporal structure-based approach to drought characterisation, *Int. J. Climatol.*, 32, 406–418, 2011.
- Lowry, C. A. and Woodall, W. H.: A Multivariate Exponentially Weighted Moving Average Control Chart, *Technometrics*, 34, 46–53, 1992.
- Mahecha, M. D., Gans, F., Sippel, S., Donges, J. F., Kaminski, T., Metzger, S., Migliavacca, M., Papale, D., Rammig, A., and Zscheischler, J.: Detecting impacts of extreme events with ecological in situ monitoring networks, *Biogeosciences*, 14, 4255–4277, <https://doi.org/10.5194/bg-14-4255-2017>, 2017.
- Mahony, C. R. and Cannon, A. J.: Wetter summers can intensify departures from natural variability in a warming climate, *Nature Comm.*, 9, 783, <https://doi.org/10.1038/s41467-018-03132-z>, 2018.
- Mankin, J. S., Smerdon, J. E., Cook, B. I., Williams, A. P., and Seager, R.: The Curious Case of Projected Twenty-First-Century Drying but Greening in the American West, *J. Climate*, 30, 8689–8710, 2017.
- Mankin, J. S., Seager, R., Smerdon, J. E., Cook, B. I., Williams, A. P., and Horton, R. M.: Blue Water Trade-Offs With Vegetation in a CO₂-Enriched Climate, *Geophys. Res. Lett.*, 45, 3115–3125, 2018.
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geosci. Model Dev.*, 10, 1903–1925, <https://doi.org/10.5194/gmd-10-1903-2017>, 2017.
- Matsueda, M.: Predictability of Euro-Russian blocking in summer of 2010, *Geophys. Res. Lett.*, 38, L06801, <https://doi.org/10.1029/2010GL046557>, 2011.
- Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, *Hydrol. Earth Syst. Sci.*, 15, 453–469, <https://doi.org/10.5194/hess-15-453-2011>, 2011.
- Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C., and Vilà-Guerau de Arellano, J.: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation, *Nat. Geosci.*, 7, 345–349, 2014.
- Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M. A., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote Sens. Environ.*, 83, 214–231, 2002.
- New, M., Hulme, M., and Jones, P.: Representing Twentieth-Century Space–Time Climate Variability. Part II: Development of 1901–96 Monthly Grids of Terrestrial Surface Climate, *J. Climate*, 13, 2217–2238, 2000.
- Parzen, E.: On Estimation of a Probability Density Function and Mode, *Ann. Math. Stat.*, 33, 1065–1076, 1962.
- Petoukhov, V., Rahmstorf, S., Petri, S., and Schellnhuber, H.-J.: Quasiresonant amplification of planetary waves and recent Northern Hemisphere weather extremes, *P. Natl. Acad. Sci. USA*, 110, 5336–5341, 2013.
- Quesada, B., Vautard, R., Yiou, P., Hirschi, M., and Seneviratne, S. I.: Asymmetric European summer heat predictability from wet and dry southern winters and springs, *Nat. Clim. Change*, 2, 736–741, 2012.
- Rahmstorf, S. and Coumou, D.: Increase of extreme events in a warming world, *P. Natl. Acad. Sci. USA*, 108, 17905–17909, 2011.
- Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S., Viovy, N., Cramer, W., Granier, A., Ogee, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., Grünwald, T., Heimann, M., Heinesch, B., Knohl, A., Kutsch, W., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.-M., Pilegaard, K., Pumpanen, J., Rambal, S., Schapf, S., Seufert, G., Soussana, J. F., Sanz, M. J., Vesala, T., and Zhao, M.: Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis, *Global Change Biol.*, 13, 634–651, 2007.
- 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, 2013.
- Scheffran, J., Brzoska, M., Kominek, J., Link, P. M., and Schilling, J.: Climate Change and Violent Conflict, *Science*, 336, 869–871, 2012.
- Schubert, S. D., Wang, H., Koster, R. D., Suarez, M. J., and Groisman, P. Y.: Northern Eurasian Heat Waves and Droughts, *J. Climate*, 27, 3169–3207, 2014.
- Schwalm, C. R., Williams, C. A., Schaefer, K., Baldocchi, D., Black, T. A., Goldstein, A. H., Law, B. E., Oechel, W. C., Kyaw Tha Paw U, and Scott, R. L.: Reduction in carbon uptake during turn of the century drought in western North America, *Nat. Geosci.*, 5, 551–556, 2012.
- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, *Earth Sci. Rev.*, 99, 125–161, 2010.

- Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment, in: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (IPCC SREX Report), edited by: Field, C., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., Mastrandrea, M., Mach, K., Plattner, G.-K., Allen, S., Tignor, M., and Midgley, 109–230, Cambridge University Press, Cambridge, 2012.
- Sippel, S., Zscheischler, J., Heimann, M., Otto, F. E. L., Peters, J., and Mahecha, M. D.: Quantifying changes in climate variability and extremes: Pitfalls and their overcoming, *Geophys. Res. Lett.*, 42, 9990–9998, 2015.
- 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, *Environ. Res. Lett.*, 12, 075006, <https://doi.org/10.1088/1748-9326/aa7398>, 2017.
- Sippel, S., Reichstein, M., Ma, X., Mahecha, M. D., Lange, H., Flach, M., and Frank, D.: Drought, Heat, and the Carbon Cycle: a Review, *Curr. Clim. Change Rep.*, 4, 266–286, <https://doi.org/10.1007/s40641-018-0103-4>, 2018.
- Smith, M. D.: An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research, *J. Ecol.*, 99, 656–663, 2011.
- Teuling, A. J., Seneviratne, S. I., Stöckli, R., Reichstein, M., Moors, E. J., Ciais, P., Luysaert, S., van den Hurk, B., Ammann, C., Bernhofer, C., Dellwik, E., Gianelle, D., Gielen, B., Grünwald, T., Klumpp, K., Montagnani, L., Moureaux, C., Sottocornola, M., and Wohlfahrt, G.: Contrasting response of European forest and grassland energy exchange to heatwaves, *Nat. Geosci.*, 3, 722–727, 2010.
- 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, <https://doi.org/10.5194/bg-13-4291-2016>.
- van Heerwaarden, C. C. and Teuling, A. J.: Disentangling the response of forest and grassland energy exchange to heatwaves under idealized land–atmosphere coupling, *Biogeosciences*, 11, 6159–6171, <https://doi.org/10.5194/bg-11-6159-2014>, 2014.
- von Buttler, J., Zscheischler, J., Rammig, A., Sippel, S., Reichstein, M., Knohl, A., Jung, M., Menzer, O., Arain, M. A., Buchmann, N., Cescatti, A., Gianelle, D., Kiely, G., Law, B. E., Magliulo, V., Margolis, H., McCaughey, H., Merbold, L., Migliavacca, M., Montagnani, L., Oechel, W., Pavelka, M., Peichl, M., Rambal, S., Raschi, A., Scott, R. L., Vaccari, F. P., van Gorsel, E., Varlagin, A., Wohlfahrt, G., and Mahecha, M. D.: Impacts of droughts and extreme-temperature events on gross primary production and ecosystem respiration: a systematic assessment across ecosystems and climate zones, *Biogeosciences*, 15, 1293–1318, <https://doi.org/10.5194/bg-15-1293-2018>, 2018.
- Wagg, C., O'Brien, M. J., Vogel, A., Scherer-Lorenzen, M., Eisenhauer, N., Schmid, B., and Weigelt, A.: Plant diversity maintains long-term ecosystem productivity under frequent drought by increasing short-term variation, *Ecology*, 98, 2952–2961, 2017.
- Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R. P., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurre, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Liu, L., Müller, C., Kumar, S. N., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Rezaei, E. E., Ripoche, D., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., and Asseng, S.: The uncertainty of crop yield projections is reduced by improved temperature response functions, *Nature Plants*, 3, 17102, <https://doi.org/10.1029/2018JG004489>, 2017.
- Wolf, S., Keenan, T. F., Fisher, J. B., Baldocchi, D. D., Desai, A. R., Richardson, A. D., Scott, R. L., Law, B. E., Litvak, M. E., Brunzell, N. A., Peters, W., and van der Laan-Luijckx, I. T.: Warm spring reduced carbon cycle impact of the 2012 US summer drought, *P. Natl. Acad. Sci. USA*, 113, 5880–5885, 2016.
- Wright, C. K., de Beurs, K. M., and Henebry, G. M.: Land surface anomalies preceding the 2010 Russian heat wave and a link to the North Atlantic oscillation, *Environ. Res. Lett.*, 9, 124015, <https://doi.org/10.1088/1748-9326/9/12/124015>, 2014.
- Yang, Y., Donohue, R. J., and McVicar, T. R.: Global estimation of effective plant rooting depth: Implications for hydrological modeling, *Water Resour. Res.*, 52, 8260–8276, 2016.
- Yoshida, Y., Joiner, J., Tucker, C., Berry, J., Lee, J. E., Walker, G., Reichle, R., Koster, R., Lyapustin, A., and Wang, Y.: The 2010 Russian drought impact on satellite measurements of solar-induced chlorophyll fluorescence: Insights from modeling and comparisons with parameters derived from satellite reflectances, *Remote Sens. Environ.*, 166, 163–177, 2015.
- Zhou, S., Yu, B., Huang, Y., and Wang, G.: The effect of vapor pressure deficit on water use efficiency at the subdaily time scale, *Geophys. Res. Lett.*, 41, 5005–5013, 2014.
- Zimek, A., Schubert, E., and Kriegel, H.-P.: A survey on unsupervised outlier detection in high-dimensional numerical data, *Stat. Anal. Data Min.*, 5, 363–387, 2012.
- Zscheischler, J. and Seneviratne, S. I.: Dependence of drivers affects risks associated with compound events, *Science Advances*, 3, e1700263, <https://doi.org/10.1126/sciadv.1700263>, 2017.
- Zscheischler, J., Mahecha, M. D., Harmeling, S., and Reichstein, M.: Detection and attribution of large spatiotemporal extreme events in Earth observation data, *Ecol. Inform.*, 15, 66–73, 2013.
- Zscheischler, J., Orth, R., and Seneviratne, S. I.: A submonthly database for detecting changes in vegetation-atmosphere coupling, *Geophys. Res. Lett.*, 42, 9816–9824, 2015.

Chapter 5

Vegetation modulates the impact of climate extremes on gross primary production

This chapter was submitted to Biogeosciences on 14th. March 2020:

Flach, M., Brenning, A., Gans, F., Reichstein, M., Sippel, S. & Mahecha, M. D. (2020): Vegetation modulates the impact of climate extremes on gross primary production, Biogeosciences Discuss., doi: 10.5194/bg-2020-80, in review.



Vegetation modulates the impact of climate extremes on gross primary production

Milan Flach^{1,2}, Alexander Brenning², Fabian Gans¹, Markus Reichstein^{1,3}, Sebastian Sippel^{4,5}, and Miguel D. Mahecha^{1,3,6}

¹Max Planck Institute for Biogeochemistry, Department Biogeochemical Integration, P. O. Box 10 01 64, D-07701 Jena, Germany

²Friedrich Schiller University Jena, Department of Geography, Jena, Germany

³German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany

⁴Norwegian Institute of Bioeconomy Research, Ås, Norway

⁵ETH Zürich, Institute for Atmospheric and Climate Science, Switzerland

⁶University of Leipzig, Remote Sensing Center for Earth System Research, Germany

Correspondence: Milan Flach (milan.flach@bgc-jena.mpg.de)

Abstract. Drought and heat events affect the uptake and sequestration of carbon in terrestrial ecosystems. Factors such as the duration, timing and intensity of extreme events influence the magnitude of impacts on ecosystem processes such as gross primary production (GPP), i.e. the ecosystem uptake of CO₂. Preceding soil moisture depletion may exacerbate these impacts. However, some vegetation types may be more resilient to climate extremes than others. This effect is insufficiently understood at the global scale and is the focus of this study. Using a global upscaled product of GPP that scales up in-situ land CO₂ flux observations with global satellite remote sensing, we study the impact of climate extremes at the global scale. We find that GPP in grasslands and agricultural areas is generally reduced during heat and drought events. However, we also find that forests, if considered globally, appear not in general to be particularly sensitive to droughts and heat events that occurred during the analyzed period or even show increased GPP values during these events. On the one hand, this is in many cases plausible, e.g. when no negative preconditioning has occurred. On the other hand, however, this may also reflect a lack of sensitivity in current remote sensing derived GPP products to the effects of droughts and heatwaves. The overall picture calls for a differentiated consideration of different land cover types in the assessments of risks of climate extremes for ecosystem functioning.

1 Introduction

We expect that climate change leads to increases in frequencies, durations, intensities, and spatial extents of droughts and heatwaves in the next decades (Meehl et al., 2000; Olesen and Bindi, 2002; Seneviratne et al., 2012; Coumou and Robinson, 2013; Cook et al., 2015; Zscheischler and Seneviratne, 2017). Ecosystems will respond to the events ahead in multiple ways. In particular the processes controlling the terrestrial carbon balance, i.e. photosynthesis and respiratory processes as well as fires and e.g. pest-induced mortality are expected to be affected (Peuelas et al., 2004; Ciais et al., 2005; Vetter et al., 2008; Reichstein et al., 2013; Bastos et al., 2014; Yoshida et al., 2015; Wolf et al., 2016; Brando et al., 2019) (for a recent review see Sippel et al. (2018)). Given that these responses represent feedbacks to the coupled climate–ecosystem dynamics, it is important to



understand which factors generally influence the magnitudes of such impacts at the global scale (Frank et al., 2015). Previous studies have shown that event duration can be as important as intensity in controlling the reduction of gross primary production (GPP), which represents the total ecosystem carbon uptake (Granier et al., 2008; von Buttlar et al., 2018; Orth and Destouni, 2018). In particular, compound extreme events, e.g., the combination of drought and heat stress can increase the impact on

25 GPP as compared to singular stressors (Ciais et al., 2005; AghaKouchak et al., 2014; Zscheischler et al., 2018; von Buttlar et al., 2018). Several case studies point to the crucial role of timing in influencing the magnitude of impacts on ecosystem functioning. Warm and early springs may partly compensate for severe carbon impacts of summer droughts (Wolf et al., 2016). In contrast, soil moisture depletion in spring can even enhance carbon losses during summer (Buermann et al., 2013; Sippel et al., 2017a; Buermann et al., 2018).

30 Probably the least understood aspect is the question how strongly land cover types modulate drought and heat impacts on the fundamental processes controlling ecosystems carbon dynamics, such as gross primary production, ecosystem respiration, and net ecosystem exchange. Evidence from eddy covariance stations (von Buttlar et al., 2018) and case studies using spatiotemporal remote sensing derived data (Wolf et al., 2016; Flach et al., 2018) suggest that certain ecosystems are less vulnerable to heat and drought events than others. However, the question to what degree land cover types shape the impacts of droughts and

35 heatwaves globally remains unclear. Here we aim to specifically investigate the importance of land cover type in controlling the impacts of climate extremes relative to other factors.

When discussing impacts of climate extremes, the crucial question is their definition. If values over some global thresholds are used to detect extremes e.g. in some meteorological variable and investigate anomalies in ecological processes, one might find very different impact patterns as compared to events that are extreme relative to their expected value. Another approach

40 is to consider the joint probability of multiple variables contributing to an event. Here we rely on a multivariate extreme event detection algorithm that can detect extremes in multi-dimensional data sources (Flach et al., 2017, 2018) and restrict our analysis to those events that can be also considered a relative drought and heat event. We estimate anomalies regionally i.e. defining extreme events relative to the typical conditions of the regional growing season. We apply this method jointly to air temperature, surface moisture, and incoming shortwave radiation as fundamental variables to detect relative extreme events.

45 Each event describes a spatiotemporal context that can be described by its spatial extent and duration (Zscheischler et al., 2013; Mahecha et al., 2017). The impacts are then assessed in these areas as anomalies in gross primary production (GPP). Our study addresses the impacts in the time range between 2003 and 2018 globally in different land cover classes and builds on nonlinear predictive models to understand the importance of the driving factors (for details see Methods, Section 2).

2 Methods

50 For detecting hydrometeorological extreme events across ecosystems we need (i) a set of variables describing hydrometeorological extreme events and their impacts on productivity (Section 2.1), (ii) a detection algorithm (Section 2.2), and (iii) an approach to evaluate the hydrometeorological extremes with regard to responses in different ecosystems (Section 2.3).



2.1 Data

To detect hydrometeorological extreme events we use 2-m air temperature, incoming shortwave radiation (both from ERA5, original resolution 0.25° , Copernicus Climate Change Service (C3S) (2017)), and surface moisture (v3.2b, original resolution 0.25° from the GLEAM model-data integration framework, (Miralles et al., 2011; Martens et al., 2017)). We consider surface moisture as a hydrometeorological variable due to its importance for drought detection although it is influenced by vegetation. The impacts of the identified extremes are quantified as anomalies in gross primary productivity (GPP, original resolution $\frac{1}{12}^\circ$ from FLUXCOM-RS, Tramontana et al. (2016)). Anomalies in GPP are computed as deviations from the mean seasonal cycle excluding the extreme year itself. The selected hydrometeorological variables have global coverage and a common spatial resolution of 0.25° , and are used at an eight-daily temporal resolution covering the 2003–2018 period. Land cover classes at $\frac{1}{12}^\circ$ resolution were obtained from MODIS (collection 5, Friedl et al. (2010)) We group the available land cover classes in forest ecosystems (land cover classes containing "forest"), agricultural ecosystems (containing "crop"), and, all remaining other land cover types.

2.2 Preprocessing and anomaly detection

We compute deviations from a smoothed median seasonal cycle in the hydrometeorological variables, which we denote as anomalies. For detecting extreme events, we apply a multivariate anomaly detection procedure described in detail in (Flach et al., 2018). It (i) accounts for seasonal changes in the variance of the anomalies using a moving window technique, and (ii) uses climatic similarities to obtain more robust thresholds for extreme event detection via spatial replicates as proposed by Mahecha et al. (2017) (for more details see the B).

The extreme event detection algorithm itself is applied to the set of hydrometeorological anomaly time series and returns anomaly scores computed by kernel density estimation. Kernel density estimation showed good performance among other possible methods and accounts for nonlinearities in the data (Flach et al., 2017). The resulting anomaly scores can be interpreted as a univariate index of deviation from the general multivariate pattern. We consider the highest 5% of the anomaly scores to be extreme events (95th percentile), which is within the typical range of percentiles defining extreme events (McPhillips et al., 2018).

2.3 Framework for extracting event-based statistics

We use the extracted binary information (extreme / non-extreme) to compute statistics based on the spatio-temporal structure of the extreme events similar to (Lloyd-Hughes, 2011; Zscheischler et al., 2013; Mahecha et al., 2017; Chen et al., 2019). Extreme voxels are considered to belong to the same extreme event if they are connected within a $3 \times 3 \times 3$ (long x lat x time) cube. Note that this definition includes connections over edges. We compute event-based statistics from the 1000 largest extreme events globally as introduced also for the Russian heatwave (Flach et al., 2018). Specifically, we calculate affected volume, centroids, mean and integral of GPP separately for positive and negative anomalies, as well as the distance between the centroids of the positive and the negative anomalies of GPP during the event. We consider an event to be predominantly a

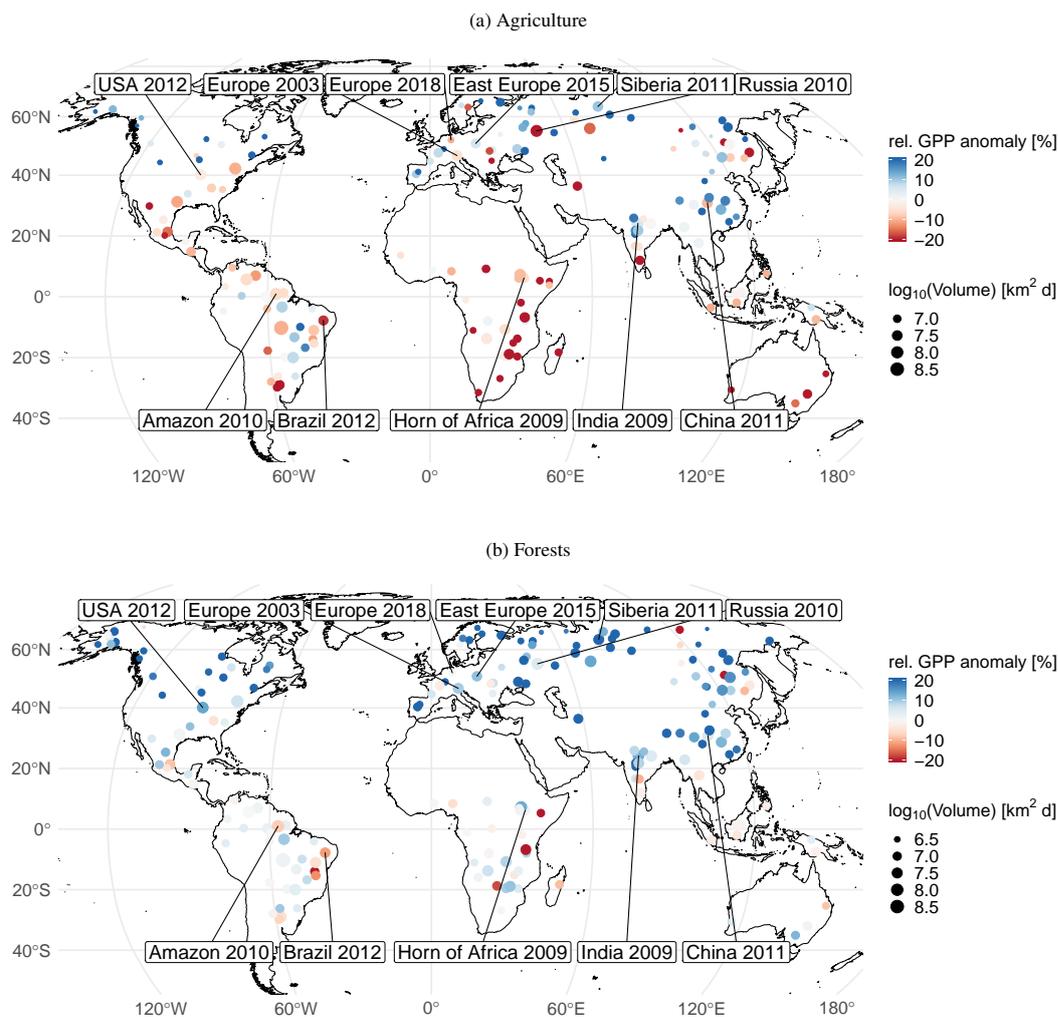


Figure 1. Relative drought and heat events coloured with the relative anomaly in gross primary production for (a) agricultural and (b) forest ecosystems. Point sizes are proportional to the affected volume of the space-time event. The largest and some well known events are labelled.

85 relative drought (relative heatwave) if more than 50% of the surface moisture (temperature) values during the extreme event are beneath (exceed) the 5th (95th) percentile of the variable. We select drought ($n = 98$) and heat ($n = 44$) events and combined drought–heat events ($n = 71$), which are taking place during the growing season (total $n = 213$), i.e. the centroid of the event is



within the half year encompassing the seasonal GPP maximum. Our statistics account for the spherical geometry of the Earth by weighting with the cosine of latitude.

90 Furthermore, we evaluate if the positive and negative anomalies in GPP during the event predominantly have a spatial or temporal component. Therefore, we split the event in parts with enhanced and parts with reduced productivity. Between those two parts, we compute the spatio-temporal distance between the centroids of each part. We consider positive and negative GPP anomalies to occur predominantly spatially if the temporal distance of the centroids is almost simultaneous, i.e. less than one time step in the data (eight days). GPP anomalies are considered to be predominantly temporally changing if the spatial distance of the centroids is less than 110 km (approximately one degree at the equator). Both, spatial and temporal components

95 can be found for centroids which are more than 110 km and more than eight days away.

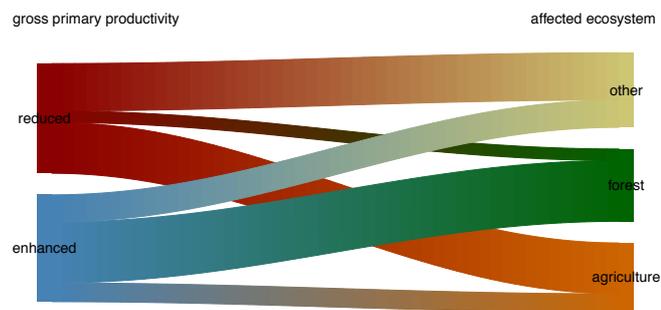


Figure 2. Proportion of GPP anomalies with reduced or enhanced productivity and their distribution in the different ecosystems (growing season events from 2003-2018). Bar sizes are proportional to the affected volume of the identified events. Forests tend to be associated with enhanced productivity rates, while agricultural ecosystems tend to be associated with reduced productivity.

2.4 Statistical model of GPP during extreme events

As we detect heatwaves and droughts relative to the mean seasonal patterns, positive or negative GPP anomalies during the droughts and heatwaves may additionally be influenced by differences in the conditions in the hydrometeorological variables

100 during the extreme event, differences in background climate in which the vegetation is growing, or duration and affected area of the event. We use gradient boosting machines (Friedman, 2001) to predict average GPP anomalies during the event as a function of mean surface moisture, mean temperature, mean radiation during the event, duration, affected area, land cover class, and mean climate during the growing season, i.e. mean temperature and surface moisture during all growing seasons between 2003 and 2018. We tune model parameters (shrinkage parameter, depth of the trees, bag fraction, minimal number of observations

105 per node) by following a workflow described in Elith et al. (2008) using a hyper grid search from 100 different random initialisations of splitting the data into training (75%) and testing (remaining 25%). We compute uncertainty of the variable importance measure described in (Friedman, 2001) from each of the 100 best models of the hyper grid search. Additionally



we use an approach based on Local Interpretable Model-agnostic Explanations (LIME), which tries to predict each single observation in a black box model based on locally weighted regression (Ribeiro et al., 2016). Here, this approach helps to understand (1) the effect of specific land cover classes, and (2) the direction of the effect.

3 Results

Our analysis based on a 5% threshold in the multivariate anomaly scores leads to a detection of 213 events (98 relative droughts, 44 relative heatwaves, 71 compound drought–heatwaves) between 2003 and 2018.

If we only discriminate forest and agricultural ecosystems, we find substantial differences in the direction of the GPP anomalies during extreme droughts and heatwaves in the growing season. In agricultural and other non-forest land-cover types, GPP was reduced during the identified events (agricultural land-cover types: 64% (56–72%) reduction, Figure 1 (a); other ecosystems 60% (53–67%), Appendix Figure A1). In forested areas, instead, a majority of 71% (63–78%, 95% confidence interval) of events shows enhanced productivity (Figure 1 (b)). The dichotomy described in the instantaneous response patterns confirms the overall statistics. Events with their centroid in France 2003, Russia 2010, and Germany 2018 all show bidirectional GPP anomalies that coincide with land-cover type transitions between predominantly forested land cover and others (a detailed illustration of the different events is provided in the supplementary materials)Figure 2 summarizes these findings across all events by relating the global integral areas of positive and negative anomalies in GPP during extreme events to the dominant land cover type.

The events analyzed here are based on relative radiation, heat and water availability anomalies (see Methods). To better understand the role of absolute climate conditions we show the reported GPP anomalies in the terms of absolute temperatures and surface moisture levels in Figure 3(a). The figure shows that reduced rates of GPP tend to coincide with very low surface moisture and high temperature (eight-daily averages). Delineating different ecosystems within this space shows that they are arranged along decreasing surface moisture values. Most extreme events in forests tend to occur under slightly higher surface moisture conditions compared to agricultural and other ecosystems (Figure 3(b)). Forests are hit less frequently critical dry conditions for which we predominantly observe reduced productivity. In contrast, we observe reduced productivity during the events for agricultural ecosystems, which experience frequently critical hot and dry conditions (Figure 3(b)).

While Figure 3(a) shows that temperature and soil moisture have some effect on the direction of the impacts, they are insufficient to explain the observed patterns in detail. To unravel the importance of land cover type and other factors we predict average GPP anomalies using gradient boosting machines ($R^2 = 0.43$, Friedman (2001) Section 2.4) and explore their relative variable importance. Growing season temperature, event duration, and land cover type are, in decreasing order, are the most important variables in the statistical model (Figure 4(a)).

Apart from identifying important variables that explain the GPP anomalies during drought and heat anomalies, we disentangle the direction of each factor's effect in the model, and, in particular for specific land cover classes. Whereas growing season temperature and duration show a negative model coefficient, i.e. a longer duration and a warmer climate are associated with a stronger impact, as expected, productivity in different land cover types is influenced in contrasting ways: Land cover

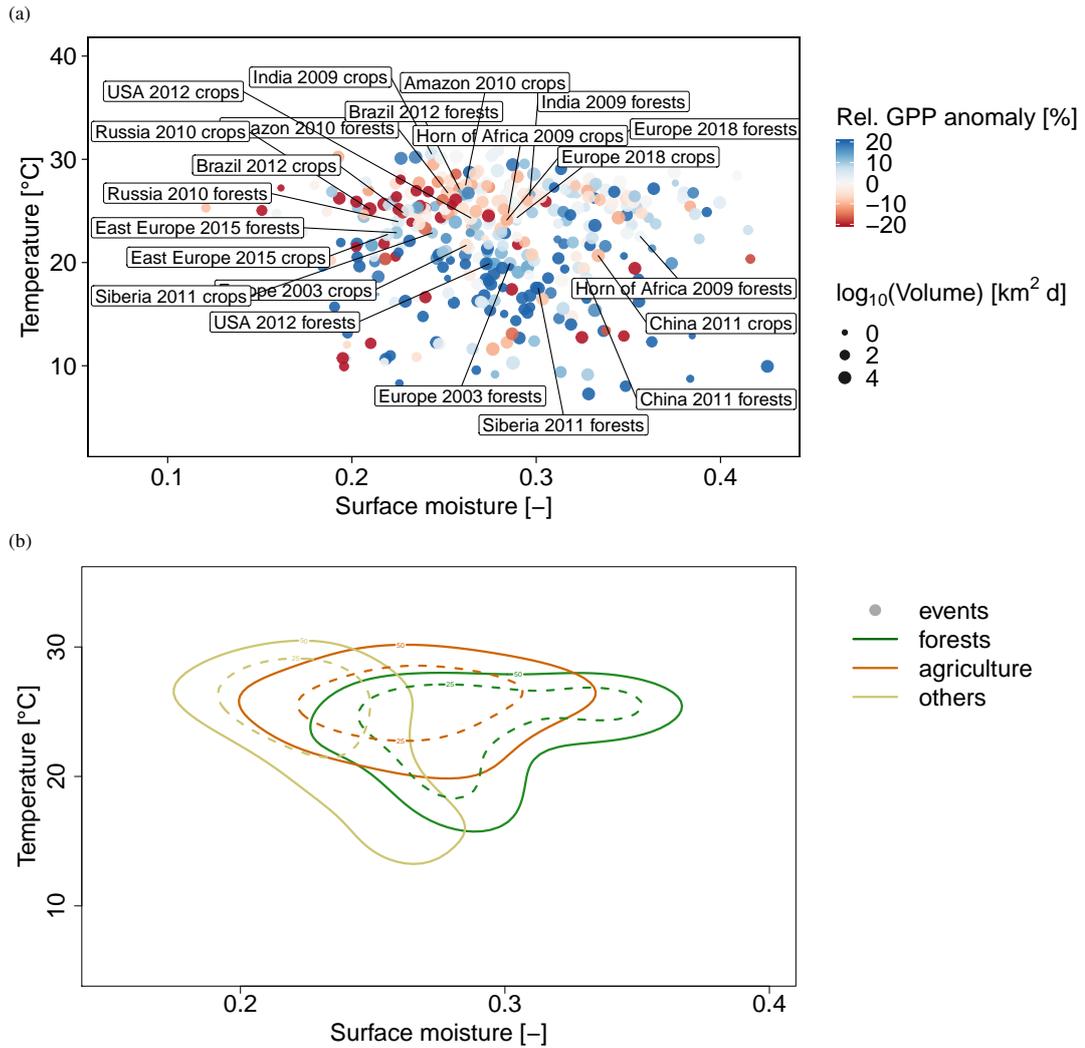


Figure 3. (a) Mean temperature and surface moisture during the relative drought and heat events for forests and agricultural ecosystems. Size and color of the points denote the affected space-time volume and the direction of the impact on productivity. (b) Average conditions in temperature and surface moisture for all ecosystems. Colored lines enclose 25% and 50% of the events within forest, agricultural and other ecosystems.



types including forests and woody savannas show increased average GPP during the extreme events. In contrast, agricultural ecosystems (land cover types including crops), grasslands, savannas, and other land cover types reduce average GPP anomalies (Figure 4(b)). Warmer growing season climates and higher temperatures during the event are associated with negative GPP anomalies. In contrast, greater availability radiation and higher surface moisture have a positive influence on the impact.

145 We showed that land cover type is one of the major factors influencing the GPP anomaly during the event. A single hydrometeorological extreme event can affect two or more adjacent land cover types simultaneously with potentially contrasting impacts (spatial contrasting anomalies), but enhanced productivity can also be observed earlier than reduced productivity (or vice versa, temporally contrasting anomalies). To explicitly quantify the role of spatial vs. temporal effects on the GPP anomalies during extreme events we split each event in parts with enhanced and reduced GPP anomalies and compute the centroidal
150 distance in space and time. In fact, positive and negative GPP anomalies mostly co-occur simultaneously in adjacent spatial regions (116 events of 213 events in total within ± 8 days, Figure 5). Especially for large scale events (large volume), a considerable distance of the anomalies can be observed in space and time. However, taking only the temporal distance into account, we have more events with enhanced productivity before the reduced productivity (temporal distance < -8 days, $n = 44$) than after (> 8 days, $n = 33$).

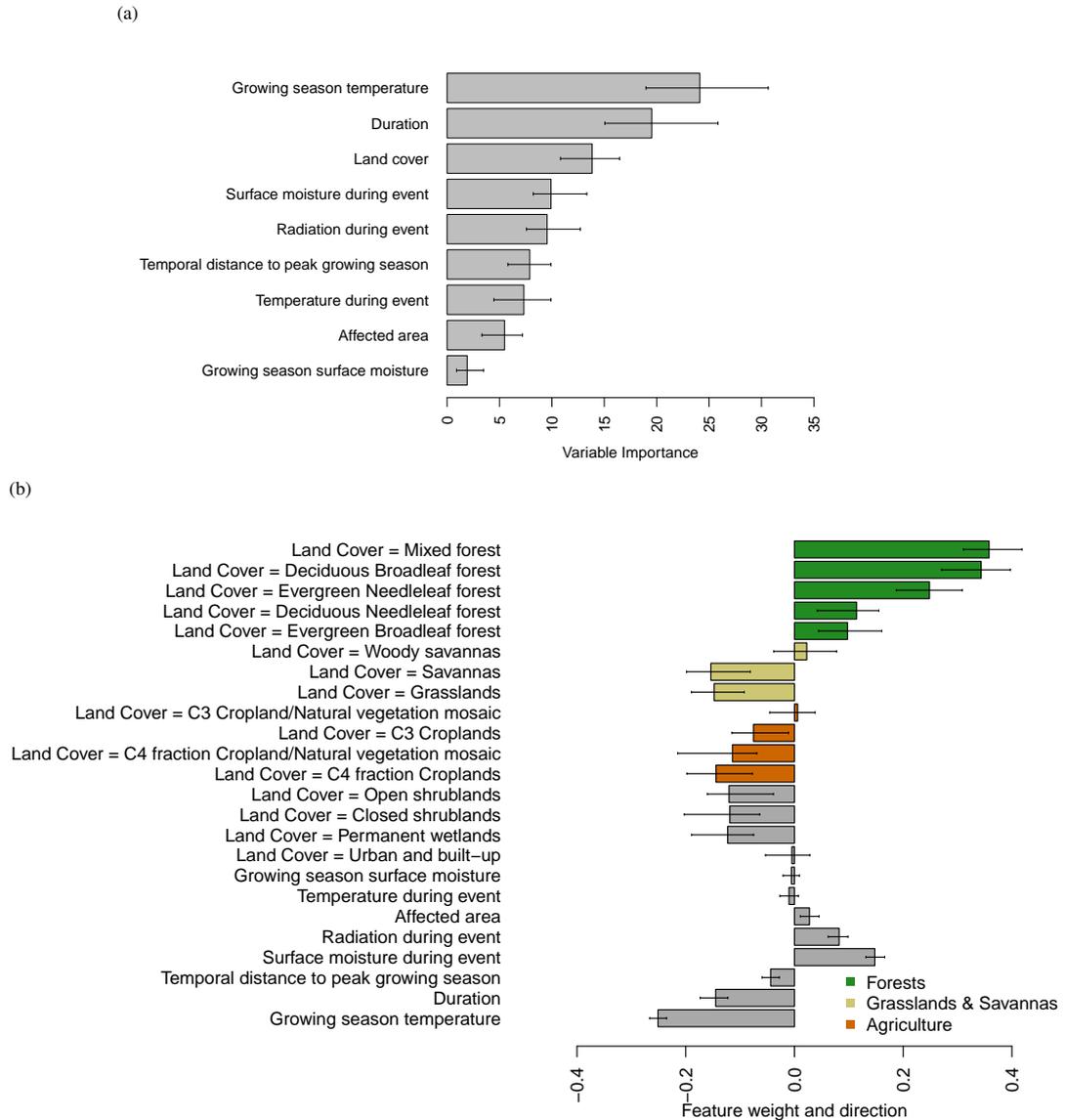


Figure 4. (a) Variable importance of the ten best gradient boosting machines predicting average GPP anomalies during the events, and (b) direction and feature weight of the variables explaining GPP anomalies of the individual events based on linear regression via local interpretable model-agnostic explanations (LIME).

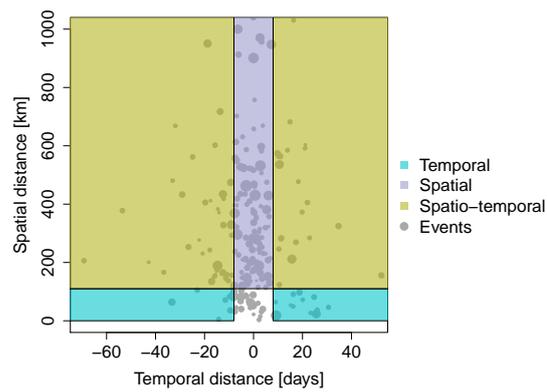


Figure 5. Each extreme event is split into parts with enhanced and reduced GPP anomalies. The centroidal distance between both parts in space and time shows whether contrasting GPP anomalies are predominantly taking place temporally, spatially or spatio-temporally. Point sizes are proportional to the event's affected volume.



155 4 Discussion

Contrasting responses of ecosystems to climate extremes, e.g. in the US in 2012 (Wolf et al., 2016) or in Russia in 2010 (Flach et al., 2018), are not singular cases but are shown to be frequent phenomena in response to hydrometeorological extreme events at the global scale. Within the same extreme event, reduced and enhanced productivity can be observed simultaneously in adjacent spatial regions. This finding complements previous studies on temporal (Wolf et al., 2016; Sippel et al., 2017a; 160 Buermann et al., 2018) or spatial contrasting responses (Jolly et al., 2005; Zaitchik et al., 2006; Lewińska et al., 2016).

This study provides evidence that the impacts of extreme drought or heat anomalies on GPP during growing seasons is, firstly a function of event duration and long-term climate, but secondly, also depends on the affected land cover type. In particular the tendency towards positive vs. negative responses seems to be controlled by tree cover (similar to the results of Ivits et al. (2014); Walther et al. (2019)), i.e. forests seem to show higher resilience to drought and heat anomalies on the 165 short term, which is reflected in a tendency towards positive GPP anomalies during the events. However, our results are based on events that are extreme relative to the regional normal conditions. In the supplementary materials we illustrate a range of events in more detail. For instance, a relative drought or heatwave in a typically wet ecosystem can boost productivity as well as a heatwave in ecosystems that are typically cold (see cases reported e.g. for China 20011, India 2009, and the Siberian heatwave 2011). Both water stress and temperature affect ecophysiological processes in a nonlinear manner. Heat events below 170 optimal temperatures enhance photosynthesis (Wang et al., 2017), or photosynthesis may be enhanced by the radiation surplus during dry periods (Walther et al., 2019) especially at higher latitudes (Bachmair et al., 2018) and as long as ecophysiological limits are not violated. Yet, the prevalence of certain land cover types is partly controlled by climatic gradients, and therefore land cover cannot really be considered independently of the mean climatological conditions that likewise play a role (Figure 3(a)). Climate conditions also lead to adaptation of physiological processes. For instance, forests in dry ecosystems may be 175 characterized by a more conservative water use strategy (Teuling et al., 2010; van Heerwaarden and Teuling, 2014; Ramos et al., 2015) and adapted to drought compared to analogous land cover types whose biogeographic history experienced colder and more moderate conditions (Doughty et al., 2015). Moreover, forests have access to deeper soil water compared to other ecosystems (Yang et al., 2016; Fan et al., 2017). The degree of isohydricity may further differentiate the response of forests, as it differs between tree species (Roman et al., 2015; Ruehr et al., 2015; Yi et al., 2017).

180 Our study only reports on GPP responses during the climatic anomaly without considering the legacy of the events. Responses may emerge with some time lag between weeks to months (Schwalm et al., 2012; Ruehr et al., 2015), or even at longer time scales (years) (Saatchi et al., 2013; Anderegg et al., 2015). Hence, finding enhanced productivity of forests during some heat event does not exclude increased mortality in the long-term. Forest ecosystems are known to potentially respond much delayed to environmental stress, which can trigger strong secondary impacts like insect outbreaks (Hicke et al., 2006; Rouault et al., 2006; Allen et al., 2010), or fires (Brando et al., 2014). In contrast, agricultural systems are known to be very directly 185 vulnerable to droughts (De Keersmaecker et al., 2016; Bachmair et al., 2018). We choose the growing season as time period of interest, which is notably different than summer for some regions, e.g. in the Mediterranean where more positive responses to warm anomalies in the cold season may be expected (Sippel et al., 2017b), and also impacts of droughts may be less than



during the dry season (Huang et al., 2018). Note that due to complex interactions between GPP and ecosystem respiration no
190 direct translation of the results into net ecosystem exchange is expected (Richardson et al., 2007).

Another aspect to discuss is data quality. Gross primary productivity from FLUXCOM-RS may inherit errors from the
underlying remote sensing products; these have, in particular, been discussed for tropical forests (Asner et al., 2004; Asner and
Alencar, 2010; Wu et al., 2018). Recently, Stocker et al. (2019) showed at the global scale that remote sensing retrieved GPP
underestimates drought impacts due to soil moisture effects on light use efficiency. Comparing our estimates of GPP impacts to
195 published data from eddy covariance stations for two case studies (US 2012, (Wolf et al., 2016), and Europe 2003 (Ciais et al.,
2005; Reichstein et al., 2007)) indicates that we do indeed underestimate GPP impact. Thus, we suspect that in addition to the
GPP estimates used by Stocker et al. (2019), also FLUXCOM-RS GPP underestimates the impacts of climate extreme events
specifically for forest ecosystems. As FLUXCOM-RS exhibits a good agreement for forests globally with GPP estimates based
on solar-induced fluorescence (Walther et al., 2019), the lack of sensitivity to drought and heat impacts in forest ecosystems
200 may be a more general issue in remote sensing data.

5 Conclusions

We conclude that a more differentiated consideration of the role of land cover reveals firstly major differences between forest
and agricultural ecosystems. These differences may originate from a different (micro-)climate or different water management
strategies including the access to deeper soil water or point to more strongly lagged impacts in forest ecosystems. However,
205 the lack of sensitivity of forest ecosystems to droughts and heatwaves is stronger than we would expect it to be. Thus, we think
that our results also point towards deficiencies in FLUXCOM-RS derived GPP which are potentially a more general issue in
remote sensing derived indices of vegetation activity. These deficiencies call for the development of new global GPP products
with a higher sensitivity to droughts and heatwaves, which can unravel the role of forest ecosystems in a more frequently hot
and dry future climate.

210 *Data availability.* We use data originating from the FLUXCOM initiative (<http://www.fluxcom.org>), the GLEAM model data integration
framework (<https://www.gleam.eu/>), and ERA5 (<https://cds.climate.copernicus.eu/cdsapp#!home>). The harmonized data set is available
within the project Earth System Data Lab (ESDL) and can be accessed here: [https://www.earthsystemdatalab.net/index.php/interact/data-](https://www.earthsystemdatalab.net/index.php/interact/data-lab/)
lab/.



Appendix A: Other ecosystems

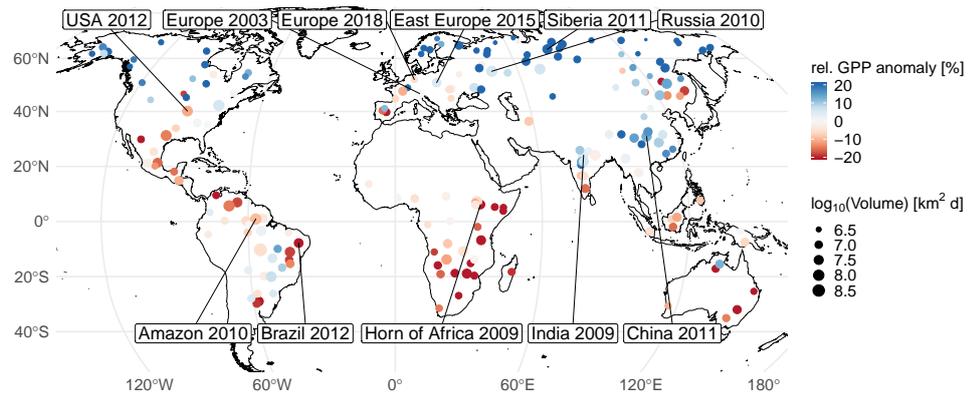


Figure A1. Relative GPP anomalies of other ecosystems (except forests and agriculture) during droughts and heatwaves. Point sizes are proportional to the (space–time) volume of the extreme event.



215 Appendix B: Technical details on the spatial segmentation

We follow the procedure described and developed by Mahecha et al. (2017), which was extended to the multivariate case by Flach et al. (2018). In summary, the used approach defines climatically and phenologically similar regions by using the leading principal components (here: three) of the seasonal cycles of the hydrometeorological variables (temperature, surface moisture, radiation) in addition to the vegetation proxy (gross primary productivity). Similar cycles appear in the same region
220 of the obtained principal component space (Figure B1). Thus, a simple classification can be obtained by dividing the principal component space into equally sized cubes. Here we use 25 breaks for each of the first three principal components, which leads to 814 classes globally of similar climate and phenology. For each pixel, we sample four random spatial replicates from each region to efficiently run the following anomaly detection workflow globally (previously the procedure was used for Europe only).

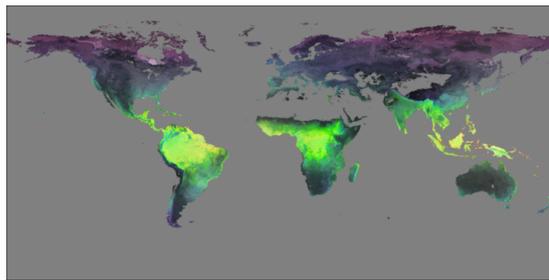


Figure B1. Map of the first three leading principal components (PCs) colored according to the colorspace hue (PC1), saturation (PC2), and, lightness (PC3).

<https://doi.org/10.5194/bg-2020-80>
Preprint. Discussion started: 27 March 2020
© Author(s) 2020. CC BY 4.0 License.



225 *Author contributions.* MF and MDM designed the study in collaboration with AB, FG, SS, MR. MF conducted the analysis and wrote the manuscript with contributions from all co-authors.

Competing interests. The authors declare they have no conflict of interests.

Acknowledgements. This research was supported by the European Space Agency (project "Earth System Data Lab") and the European Union's Horizon 2020 research and innovation programme (project "BACI", grant agreement no 64176). The authors are grateful to the
230 FLUXCOM initiative (<http://www.fluxcom.org>) for providing the data. MF acknowledges support by the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS).



References

- AghaKouchak, A., Cheng, L., Mazdiyasi, O., and Farahmand, A.: Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought, *Geophys. Res. Lett.*, 41, 8847–8852, 2014.
- 235 Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Venetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. H. T., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J.-H., Allard, G., Running, S. W., Semerci, A., and Cobb, N.: A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, *Forest Ecology and Management*, 259, 660–684, 2010.
- Anderegg, W. R. L., Schwalm, C. R., Biondi, F., Camarero, J. J., Koch, G., Litvak, M., Ogle, K., Shaw, J. D., Shevliakova, E., Williams, A. P., Wolf, A., Ziaco, E., and Pacala, S.: Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models, *Science*, 349, 524–528, 2015.
- Asner, G. P. and Alencar, A.: Drought impacts on the Amazon forest: the remote sensing perspective, *New Phytologist*, 187, 569–578, 2010.
- Asner, G. P., Nepstad, D., Cardinot, G., and Ray, D.: Drought stress and carbon uptake in an Amazon forest measured with spaceborne imaging spectroscopy, *PNAS*, 101, 6039–6044, 2004.
- 245 Bachmair, S., Tanguy, M., Hannaford, J., and Stahl, K.: How well do meteorological indicators represent agricultural and forest drought across Europe?, *Environmental Research Letters*, 13, 034042, 2018.
- 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.
- Brando, P. M., Balch, J. K., Nepstad, D. C., Morton, D. C., Putz, F. E., Coe, M. T., Silverio, D., Macedo, M. N., Davidson, E. A., Nobrega, C. C., Alencar, A., and Soares-Filho, B. S.: Abrupt increases in Amazonian tree mortality due to drought-fire interactions, *Proceedings of the National Academy of Sciences*, 111, 6347–6352, 2014.
- Brando, P. M., Paolucci, L., Ummenhofer, C. C., Ordway, E. M., Hartmann, H., Cattau, M. E., Rattis, L., Medjibe, V., Coe, M. T., and Balch, J.: Droughts, Wildfires, and Forest Carbon Cycling: A Pantropical Synthesis, *Annual Review of Earth and Planetary Sciences*, 47, 555–581, 2019.
- 255 Buermann, W., Bikash, P. R., Jung, M., Burn, D. H., and Reichstein, M.: Earlier springs decrease peak summer productivity in North American boreal forests, *Environ. Res. Lett.*, 8, 024027, 2013.
- Buermann, W., Forkel, M., O’Sullivan, M., Sitch, S., Friedlingstein, P., Haverd, V., Jain, A. K., Kato, E., Kautz, M., Lienert, S., Lombardozi, D., Nabel, J. E. M. S., Tian, H., Wiltshire, A. J., Zhu, D., Smith, W. K., and Richardson, A. D.: Widespread seasonal compensation effects of spring warming on northern plant productivity, *Nature*, 562, 110–111, 2018.
- 260 Chen, W., Zhu, D., Huang, C., Ciais, P., Yao, Y., Friedlingstein, P., Sitch, S., Haverd, V., Jain, A. K., Kato, E., Kautz, M., Lienert, S., Lombardozi, D., Poulter, B., Tian, H., Vuichard, N., Walker, A. P., and Zeng, N.: Negative extreme events in gross primary productivity and their drivers in China during the past three decades, *Agricultural and Forest Meteorology*, 275, 47–58, 2019.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, 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, 2005.



- Cook, B. I., Ault, T. R., and Smerdon, J. E.: Unprecedented 21st century drought risk in the American Southwest and Central Plains, *Science Advances*, 1, e1400082, 2015.
- 270 Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS), pp. accessed in April 2019, <https://cds.climate.copernicus.eu/cdsapp#!/home>, 2017.
- Coumou, D. and Robinson, A.: Historic and future increase in the global land area affected by monthly heat extremes, *Environ. Res. Lett.*, 8, 034018, 2013.
- De Keersmaecker, W., van Rooijen, N., Lhermitte, S., Tits, L., Schaminée, J., Coppin, P., Honnay, O., and Somers, B.: Species-rich semi-natural grasslands have a higher resistance but a lower resilience than intensively managed agricultural grasslands in response to climate anomalies, *Journal of Applied Ecology*, 53, 430–439, 2016.
- 275 Doughty, C. E., Metcalfe, D. B., Girardin, C. A. J., Amézquita, F. F., Cabrera, D. G., Huasco, W. H., Silva-Espejo, J. E., Araujo-Murakami, A., da Costa, M. C., Rocha, W., Feldpausch, T. R., Mendoza, A. L. M., da Costa, A. C. L., Meir, P., Phillips, O. L., and Malhi, Y.: Drought impact on forest carbon dynamics and fluxes in Amazonia, *Nature*, 519, 78–82, 2015.
- 280 Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, *Journal of Animal Ecology*, 77, 802–813, 2008.
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic regulation of plant rooting depth, *Proceedings of the National Academy of Sciences*, 82, 201712381, 2017.
- Flach, M., Gans, F., Brenning, A., Denzler, J., Reichstein, M., Rodner, E., Bathiany, S., Bodesheim, P., Guanache, Y., Sippel, S., and Mahecha, M. D.: Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques, *Earth System Dynamics*, 8, 677–696, 2017.
- 285 Flach, M., Sippel, S., Gans, F., Bastos, A., Brenning, A., Reichstein, M., and Mahecha, M. D.: Contrasting biosphere responses to hydrometeorological extremes: revisiting the 2010 western Russian heatwave, *Biogeosciences*, 15, 6067–6085, 2018.
- Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., 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., Zavalá, 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.
- 290 Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: Remote Sensing of Environment, *Remote Sensing of Environment*, 114, 168–182, 2010.
- Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics*, 29, 1189–1232, 2001.
- 295 Granier, A., Bréda, N., Longdoz, B., Gross, P., and Ngao, J.: Ten years of fluxes and stand growth in a young beech forest at Hesse, North-eastern France, *Annals of Forest Science*, 65, doi:10.1051/forest:2008052, 2008.
- Hicke, J. A., Logan, J. A., Powell, J., and Ojima, D. S.: Changing temperatures influence suitability for modeled mountain pine beetle (*Dendroctonus ponderosae*) outbreaks in the western United States, *Journal of Geophysical Research: Biogeosciences*, 111, doi:10.1029/2005JG000101, 2006.
- 300 Huang, M., Wang, X., Keenan, T. F., and Piao, S.: Drought timing influences the legacy of tree growth recovery, *Global Change Biology*, 24, 3546–3559, 2018.
- Ivits, E., Horion, S., Fensholt, R., and Cherlet, M.: Drought footprint on European ecosystems between 1999 and 2010 assessed by remotely sensed vegetation phenology and productivity, *Global Change Biology*, 20, 581–593, 2014.
- Jolly, W. M., Dobbertin, M., Zimmermann, N. E., and Reichstein, M.: Divergent vegetation growth responses to the 2003 heat wave in the Swiss Alps, *Geophysical Research Letters*, 32, doi: 10.1029/2005GL023252, 2005.
- 305



- Lewińska, K., Ivits, E., Schardt, M., and Zebisch, M.: Alpine Forest Drought Monitoring in South Tyrol: PCA Based Synergy between scPDSI Data and MODIS Derived NDVI and NDII7 Time Series, *Remote Sensing*, 8, doi:10.3390/rs808063, 2016.
- Lloyd-Hughes, B.: A spatio-temporal structure-based approach to drought characterisation, *International Journal of Climatology*, 32, 406–418, 2011.
- 310 Mahecha, M. D., Gans, F., Sippel, S., Donges, J. F., Kaminski, T., Metzger, S., Migliavacca, M., Papale, D., Rammig, A., and Zscheischler, J.: Detecting impacts of extreme events with ecological in situ monitoring networks, *Biogeosciences*, 14, 4255–4277, 2017.
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geoscientific Model Development*, 10, 1903–1925, 2017.
- 315 McPhillips, L. E., Chang, H., Chester, M. V., Depietri, Y., Friedman, E., Grimm, N. B., Kominoski, J. S., McPhearson, T., Méndez-Lázaro, P., Rosi, E. J., and Shafiei Shiva, J.: Defining Extreme Events: A Cross-Disciplinary Review, *Earth's Future*, 6, 441–455, 2018.
- Meehl, G. A., Zwiers, F. W., Evans, J. L., Knutson, T., Mearns, L. O., and Whetton, P.: Trends in Extreme Weather and Climate Events: Issues Related to Modeling Extremes in Projections of Future Climate Change, *BAMS*, 81, 427–436, 2000.
- Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, *Hydrology and Earth System Sciences*, 15, 453–469, 2011.
- 320 Olesen, J. E. and Bindi, M.: Consequences of climate change for European agricultural productivity, land use and policy, *European Journal of Agronomy*, 16, 239–262, 2002.
- Orth, R. and Destouni, G.: Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe, *Nature Communications*, 9, doi:10.1038/s41467-018-06013-7, 2018.
- 325 Peuelas, J., Gordon, C., Llorens, L., Nielsen, T., Tietema, A., Beier, C., Bruna, P., Emmett, B., Estiarte, M., and Gorissen, A.: Noninvasive Field Experiments Show Different Plant Responses to Warming and Drought Among Sites, Seasons, and Species in a North-South European Gradient, *Ecosystems*, 7, 598–612, 2004.
- Ramos, A., Pereira, M. J., Soares, A., do Rosário, L., Matos, P., Nunes, A., Branquinho, C., and Pinho, P.: Agricultural and Forest Meteorology, *Agricultural and Forest Meteorology*, 202, 44–50, 2015.
- 330 Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S., Viovy, N., Cramer, W., Granier, A., Ogee, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., Grünwald, T., Heimann, M., Heinesch, B., Knohl, A., Kutsch, W., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.-M., Pilegaard, K., Pumpanen, J., Rambal, S., Schaphoff, S., Seufert, G., Soussana, J. F., Sanz, M. J., Vesala, T., and Zhao, M.: Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis, *Global Change Biology*, 13, 634–651, 2007.
- 335 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, 2013.
- Ribeiro, M. T., Singh, S., and Guestrin, C.: "Why Should I Trust You?", in: *KDD '16. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*, pp. 1135–1144, ACM Press, San Francisco, CA, USA, 2016.
- 340 Richardson, A. D., Hollinger, D. Y., Aber, J. D., Ollinger, S. V., and Braswell, B. H.: Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, *Global Change Biology*, 13, 788–803, 2007.
- Roman, D. T., Novick, K. A., Brzostek, E. R., Dragoni, D., Rahman, F., and Phillips, R. P.: The role of isohydric and anisohydric species in determining ecosystem-scale response to severe drought, *Oecologia*, 179, 641–654, 2015.



- Rouault, G., Candau, J.-N., Lieutier, F., Nageleisen, L.-M., Martin, J.-C., and Warzée, N.: Effects of drought and heat on forest insect
345 populations in relation to the 2003 drought in Western Europe, *Annals of Forest Science*, 63, 613–624, 2006.
- Ruehr, N. K., Gast, A., Weber, C., Daub, B., and Arneith, A.: Water availability as dominant control of heat stress responses in two contrasting
tree species, *Tree Physiology*, 36, 164–178, 2015.
- Saatchi, S., Asefi-Najafabady, S., Malhi, Y., Aragao, L. E. O. C., Anderson, L. O., Myneni, R. B., and Nemani, R.: Persistent effects of a
severe drought on Amazonian forest canopy, *PNAS*, 110, 565–570, 2013.
- 350 Schwalm, C. R., Williams, C. A., Schaefer, K., Baldocchi, D., Black, T. A., Goldstein, A. H., Law, B. E., Oechel, W. C., U, K. T. P., and
Scott, R. L.: Reduction in carbon uptake during turn of the century drought in western North America, *Nature Geoscience*, 5, 551–556,
2012.
- Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein,
M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment, in:
355 *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (IPCC SREX Report)*, edited by Field, C.,
Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., Mastrandrea, M., Mach, K., Plattner, G.-K., Allen, S., Tignor, M., and Midgley, pp.
109–230, Cambridge University Press, 2012.
- 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, *Environ. Res. Lett.*, 12, 075 006,
360 2017a.
- Sippel, S., El-Madany, T. S., Mahecha, M. D., Migliavacca, M., Carrara, A., Flach, M., Kaminski, T., Otto, F. E. L., Thonicke, K., Vossbeck,
M., and Reichstein, M.: Warm winter, wet spring, and extreme response in ecosystem functioning on the Iberian Peninsula, *BAMS*,
98, S80–S85, 2017b.
- Sippel, S., Reichstein, M., Ma, X., Mahecha, M. D., Lange, H., Flach, M., and Frank, D.: Drought, Heat, and the Carbon Cycle: a Review,
365 *Current Climate Change Reports*, 4, 266–286, 2018.
- Stocker, B. D., Zscheischler, J., Keenan, T. F., Prentice, I. C., Seneviratne, S. I., and Peñuelas, J.: Drought impacts on terrestrial primary
production underestimated by satellite monitoring, *Nature Geoscience*, 12, 264–270, 2019.
- Teuling, A. J., Seneviratne, S. I., Stöckli, R., Reichstein, M., Moors, E. J., Ciais, P., Luysaert, S., van den Hurk, B., Ammann, C., Bernhofer,
C., Dellwik, E., Gianelle, D., Gielen, B., Grünwald, T., Klumpp, K., Montagnani, L., Moureaux, C., Sottocornola, M., and Wohlfahrt, G.:
370 Contrasting response of European forest and grassland energy exchange to heatwaves, *Nature Geoscience*, 3, 722–727, 2010.
- 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, 2016.
- van Heerwaarden, C. C. and Teuling, A. J.: Disentangling the response of forest and grassland energy exchange to heatwaves under idealized
375 land–atmosphere coupling, *Biogeosciences*, 11, 6159–6171, 2014.
- Vetter, M., Churkina, M., Jung, M., Reichstein, M., Zaehle, S., Bondeau, A., Chen, Y., Ciais, P., Feser, F., Freibauer, A., Geyer, R., Jones,
C., Papale, D., Tenhunen, J., Tomelleri, E., Trusilova, K., Viovy, N., and Heimann, M.: Analyzing the causes and spatial pattern of the
European 2003 carbon flux anomaly using seven models, *Biogeosciences*, 5, 561–583, 2008.
- von Buttlar, J., Zscheischler, J., Rammig, A., Sippel, S., Reichstein, M., Knohl, A., Jung, M., Menzer, O., Arain, M. A., Buchmann, N.,
380 Cescatti, A., Gianelle, D., Kiely, G., Law, B. E., Magliulo, V., Margolis, H., McCaughey, H., Merbold, L., Migliavacca, M., Montagnani,
L., Oechel, W., Pavelka, M., Peichl, M., Rambal, S., Raschi, A., Scott, R. L., Vaccari, F. P., van Gorsel, E., Varlagin, A., Wohlfahrt, G.,



- and Mahecha, M. D.: Impacts of droughts and extreme-temperature events on gross primary production and ecosystem respiration: a systematic assessment across ecosystems and climate zones, *Biogeosciences*, 15, 1293–1318, 2018.
- Walther, S., Duveiller, G., Jung, M., Guanter, L., Cescatti, A., and Camps-Valls, G.: Satellite Observations of the Contrasting Response of
385 Trees and Grasses to Variations in Water Availability, *Geophys. Res. Lett.*, 46, 1429–1440, 2019.
- Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R. P., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurrealde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Liu, L., Müller, C., Kumar, S. N., Nendel, C., O’Leary, G., Olesen, J. E., Palosuo, T., Priesack, E.,
390 Rezaei, E. E., Ripoche, D., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., and Asseng, S.: The uncertainty of crop yield projections is reduced by improved temperature response functions, *Nature Plants*, 3, doi:10.1038/nplants.2017.102, 2017.
- Wolf, S., Keenan, T. F., Fisher, J. B., Baldocchi, D. D., Desai, A. R., Richardson, A. D., Scott, R. L., Law, B. E., Litvak, M. E., Brunsell, N. A., Peters, W., and van der Laan-Luijckx, I. T.: Warm spring reduced carbon cycle impact of the 2012 US summer drought, *Proceedings of the National Academy of Sciences*, 113, 5880–5885, 2016.
395
- Wu, J., Kobayashi, H., Stark, S. C., Meng, R., Guan, K., Tran, N. N., Gao, S., Yang, W., Restrepo-Coupe, N., Miura, T., Oliveira, R. C., Rogers, A., Dye, D. G., Nelson, B. W., Serbin, S. P., Huete, A. R., and Saleska, S. R.: Biological processes dominate seasonality of remotely sensed canopy greenness in an Amazon evergreen forest, *New Phytologist*, 217, 1507–1520, 2018.
- Yang, Y., Donohue, R. J., and McVicar, T. R.: Global estimation of effective plant rooting depth: Implications for hydrological modeling,
400 *Water Resources Research*, 52, 8260–8276, 2016.
- Yi, K., Dragoni, D., Phillips, R. P., Roman, D. T., and Novick, K. A.: Dynamics of stem water uptake among isohydric and anisohydric species experiencing a severe drought, *Tree Physiology*, 506, 153, 2017.
- Yoshida, Y., Joiner, J., Tucker, C., Berry, J., Lee, J. E., Walker, G., Reichle, R., Koster, R., Lyapustin, A., and Wang, Y.: The 2010 Russian drought impact on satellite measurements of solar-induced chlorophyll fluorescence: Insights from modeling and comparisons with
405 parameters derived from satellite reflectances, *Remote Sensing of Environment*, 166, 163–177, 2015.
- Zaitchik, B. F., Macalady, A. K., Bonneau, L. R., and Smith, R. B.: Europe’s 2003 heat wave: a satellite view of impacts and land–atmosphere feedbacks, *International Journal of Climatology*, 26, 743–769, 2006.
- Zscheischler, J. and Seneviratne, S. I.: Dependence of drivers affects risks associated with compound events, *Science Advances*, 3, doi:10.1126/sciadv.1700263, 2017.
- 410 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, 2013.
- Zscheischler, J., Westra, S., Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, *Nature Climate Change*, 8, 469–477, 2018.

Chapter 6

Discussion

6.1 Alternative pathways

One aspect to discuss is that it would have been possible to use different algorithms or ensembles of algorithms in chapter 3, and following up on this in chapter 4 and 5. A methodological comparison of different algorithms as performed in chapter 3 can never be complete by definition due to an ever growing set of potential algorithms and simultaneous developments of methods in the field (e.g. Maximal divergent intervals Barz et al., 2017; Rodner et al., 2016). Methods can fail due to slightly different assumptions (e.g. training data without anomalies Pimentel et al., 2014, which may be fixable by applying a method repeatedly). However, in the method comparison of chapter 3, I tried to focus on methods which (1) can detect anomalies in an unsupervised way and (2) which are non-parametric, or whose model parameters can be fixed with some rule of thumb to facilitate automated detection in Earth observations at the least. Nevertheless, another set of algorithms for usage in chapter 4 or 5 would have been possible.

In particular, I selected one of the three best algorithms (kernel density estimation) to detect extreme events in chapter 4 although other choices would have been possible. Another choice would have been the recurrence approach. Although the performance in terms of area under the receiver operator curve is similar, the recurrence approach is more sensitive to the chosen parameter (ε) than kernel density estimation (Supplementary material of chapter 3). Similar, the third of the three best algorithms, a k-nearest neighbours approach is more sensitive with respect to choosing the number of nearest neighbors k. Using ensembles of the three best algorithms would be another approach, which was not chosen due to computational constraints, although I note that efficient and fast implementations of the algorithms, e.g. on GPU machines are ready to be applied soon Michailidis, 2013; Rawald et al., 2017.

6.2 Sensitivity of the results

Another aspect to discuss is the sensitivity of the results. Two choices are particularly relevant for chapter 4 and 5: the threshold to detect extreme events, and the sensitivity of the results with respect to applying the extreme events detection scheme regionally (instead of locally for each pixel, or globally for one region).

First, it is shown in chapter 4 that the results are not particularly sensitive to the chosen threshold (5% extreme events per region). The sensitivity of the results with respect to the chosen threshold was not evaluated in detail in chapter 5, as I expect a similar sensitivity globally as for the Russian Heatwave 2010. Furthermore, the chosen threshold is one of the standard choices in this field McPhillips et al., 2018. Thus, similar to chapter 4, positive GPP anomalies during extreme events in forest ecosystems are expected to be found as well as for lower thresholds (10% extreme events per region) as well as for more strict thresholds. However, when using very strict thresholds, like 1%, the results for the Russian Heatwave started to get a stronger tendency towards more negative GPP anomalies. Thus, it is expected that more stricter thresholds are associated also globally with a tendency towards more negative GPP anomalies.

Second, detecting extreme events in different regions adds another aspect to the sensitivity of the results. Detecting extreme events locally (per time series in a pixel) Ivits et al., 2014 or with one threshold globally Zscheichler et al., 2014 leads to differences in the results. However, the sensitivity of our results to slight modifications in the parameters used to identify similar regions has not been discussed so far. Testing the sensitivity of the extreme event detection scheme in a latitudinal transect over Africa and Europe shows that a rather low sensitivity to slight modifications in the regionalisation parameters can be expected globally (Supplementary materials, Chapter S1.3).

One aspect in the context of the regionalisation scheme is particularly relevant for chapter 5 and has not been discussed in detail in this chapter. In this particular chapter, I use a subsampling procedure for each different region in order to detect the events. The main pixel is extended by four pixels from the respective region, which seems to be very little at first glance. Four samples are, however, chosen to keep computational time low. Yet, given that the subsampling procedure is repeated for each pixel, and the number of pixels is high and can be regarded as a spatial replicate of the subsampling procedure, the repeated subsampling procedure may be sufficient to cover the variability in each region. Using spatial replicates is common, e.g. for predicting carbon, water and energy fluxes globally (Jung et al., 2017; Tramontana et al., 2016).

Furthermore, the results of any analysis as such can be highly sensitive to the data under scrutiny. Here FLUXCOM-RS data (Tramontana et al., 2016) is used in

chapter 5, and additionally the fraction of absorbed photosynthetic active radiation in chapter 4. It can be questioned whether data originating e.g. from dynamic global vegetation models (Sitch et al., 2008) show similar patterns. It is possible that patterns of more positive responses of forest ecosystems are a feature of machine learning, i.e. learning the model from biased remote sensing data (Stocker et al., 2019) in contrast to process based models. Recently (Stocker et al., 2019) shows that impacts of droughts on primary productivity in general are underestimated when retrievals of gross primary productivity are based on remote sensing. Apart from the points already mentioned in the discussion of chapter 5, it is important to note that one general result here is the differentiated response of specific vegetation types. Even if the impacts of gross primary productivity are underestimated in general in the data used here (the FLUXCOM-RS data used here is not analyzed by the paper of (Stocker et al., 2019)), the differentiated response of specific vegetation types is still remaining one main result.

6.3 Timing of droughts and heatwaves

In chapter 5, solely the impacts of climate extreme events during the growing season are assessed, although timing of droughts and heatwaves may be crucial Sippel et al., 2016. Growing season includes effects of timing between spring, summer, and autumn. In particular, vegetation is highly sensitive in spring and summer (Darenova et al., 2017; Denton et al., 2016), with potentially contrasting impacts (Buermann et al., 2013, 2018; Wolf et al., 2016). However, the definition of growing season used here (half year around maximum gross primary productivity) does not include extreme events in winter (for temperate regions), or dry seasons for some ecosystems. Furthermore, other definitions (like using a threshold of gross primary productivity per month) may be useful, e.g. for Africa, where more than one rainy season can lead to several distinct growing seasons.

The importance of extreme events in the non-growing season time cannot be neglected. For example, winter warming can lead to the complete loss of berries in the following summer (Bokhorst et al., 2008). In general, high damage on evergreen plants occurs after midwinter warming (which leads to shallow snow depth) which is followed up by extreme frosts (Bjerke et al., 2014). Snow cover is also crucial for the survival of young trees (Drescher & Thomas, 2012), which can impact the carbon cycle apart from biodiversity (Section 6.5). Conversely, in mediterranean ecosystems winter warming leads to enhanced productivity rather than damage of vegetation (Sippel et al., 2017). However, generally, severe impacts of droughts have been found for semi-arid ecosystems (Ma et al., 2015). There, non-growing season droughts in the dry season may have stronger impacts on vegetation than droughts in the growing

season (Huang et al., 2018). Hence, the effect of timing is difficult to generalize and strongly dependent on the ecosystem under scrutiny as well as the event type.

6.4 Lagged effects in forests

In media tree mortality in Germany recently gains a lot attention (DPA, 2019; Köppe, 2019; Weiß, 2019). One shortcoming in the public perception of tree mortality as well as in chapter 4 and 5 is the lack of considering legacy effects: tree mortality in 2019 may not be instantaneously driven by the dry spring or the hot summer 2019, but the previous year may play an important role (Arnone et al., 2008). The exceptionally hot and dry summer 2018 may trigger lagged tree mortality in 2019. Insect populations, in general, are affected when temperature is reaching conditions above some optimum. However, bark beetles are known to undergo a process similar to summer dormancy during extreme heat years (Rouault et al., 2006). Lacking water during extreme droughts (but not during moderate droughts, (Raffa, 2001)) can reduce resin flow and therefore weaken the trees with respect to attacking insects. In the year after the drought, e.g. for 2004 (after the 2003 drought and heat) exploding bark beetle populations and damage on trees is reported. Possible reasons are already weakened host-trees and strong insect populations, which may benefit from warm spring or winters following the summer drought and heat (Hicke et al., 2006; Rouault et al., 2006). Thus, the sequence of two unusual years in large parts of Germany (2018 and 2019) in terms of heat and dryness further complicates the attribution of tree mortality lamented by foresters to one of the specific events. This sequence of events is one reason why the focus of chapter 5 is on instantaneous effects of drought and heat on vegetation.

Another aspect which complicates the assessment of lagged effects is the question of the time scale on which lagged effects are taking place. Reported time scales range from weeks to several years, or even more (Anderegg et al., 2015; Ruehr et al., 2015; Saatchi et al., 2013; Schwalm et al., 2017). For instance, (Anderegg et al., 2015) reports one to four years of recovery time after extreme droughts. Another study reports less than one year for most regions globally (Schwalm et al., 2017), but talks about higher recovery times in the Northern latitudes and the tropics (in agreement with (Saatchi et al., 2013)). Thus, to assess lagged effects of climate extremes, longer time series than used in chapter 4 and 5 would be desirable.

6.5 A more general view on impacts of droughts and heatwaves

The focus of chapter 4 and 5 is on the impact of droughts and heatwaves on ecosystems and specifically on primary productivity as any changes in productivity can affect the global carbon cycle (Le Quéré et al., 2018). However, impacts of extreme events are manifold (see also chapter 1.1) and not just focussed on carbon. More generally, ecosystem services can be affected strongly by droughts and heatwaves. Related to reductions in primary productivity are reductions in agricultural systems providing food for the society (Wegren, 2010). The development of mitigation strategies of impacts on agriculture is of particular importance (Lopez-Nicolas et al., 2017). Furthermore, droughts also strongly affect the water supply and quality which is an important ecosystem service for society (Benotti et al., 2010; Shen et al., 2007; Vliet & Zwolsman, 2008). Another ecosystem service is tourism. It can be affected by heatwaves (Buckley & Foushee, 2011), and droughts (Scott & Lemieux, 2010) e.g. through water levels of rivers used for recreation.

Biodiversity is affected by droughts and heatwaves in terrestrial ecosystems (Jentsch et al., 2011), and is also discussed for marine ecosystems recently (Smale et al., 2019). Loss of biodiversity due to extreme events is likely and can lead to shifts in species compositions (Jentsch & Beierkuhnlein, 2008). For example, changes in species composition due to droughts favour C4 plants (White et al., 2000). C4 plants cope better with droughts, which can also be seen in a more neutral modelling coefficient of C4 crops compared to C3 crops (see results of chapter 5 (Figure 4b)). Additionally, biodiversity is associated with less impacts from climate extreme events, e.g. more diverse grassland communities are less affected by droughts than less diverse grassland communities (Kreyling et al., 2008). Hence, biodiversity is affected by climate extremes and may also serve as a counteracting force for buffering the impacts.

Lacking water directly impacts microbial activity in the soil, i.e. ecosystem respiration is decreased (Jentsch et al., 2011) (Davidson & Janssens, 2006). However, species communities in the soil may be affected as well, e.g. by selectively affecting specific species (bacteria) less than others (fungi) (Yuste et al., 2010), but can also be less pronounced than expected (Eisenhauer et al., 2011). Furthermore, a drought followed by re-wetting and drying patterns can enhance carbon emissions from soils and may also disrupt soil aggregates or reduce nutrient retention (Borken & Matzner, 2009).

6.6 Climate extremes beyond droughts and heat-waves

Apart from drought and heat events, which I focussed on in chapter 4 and 5, other climate extreme events are also impacting the carbon cycle. Specifically, extreme frosts, heavy precipitation and heavy storms are known to play a role in this context (Frank et al., 2015).

Extreme frosts can lead to ice breakage or frost damage. In particular for the latter one, timing is of crucial importance: frost damage happens particularly often in early spring after short warming periods. Heavy precipitation can lead to flooding and burial of carbon (see (Frank et al., 2015), and references therein). Storms can lead to tree mortality through wind throw. Two famous examples are the storm Lothar in Europe 1999 (Chambers et al., 2007), and the hurricane Katrina in the US 2005 (Lindroth et al., 2009), which both offset large parts of the carbon sink in the respective regions. All these kinds of extreme events, such as storms, frost damage or flooding have not been considered in the thesis.

However, I want to emphasize here that the methods developed in chapter 3 are generic enough to detect any type of extreme event, and not just drought and heat events. Drought and heat events were in the focus of this thesis due to their expected high impacts on primary productivity (e.g.)(Zscheischler et al., 2014b). An outline of how a generic detection of any type of extreme events could look like is given in chapter 6.8.

6.7 Differences between the multivariate and univariate perspective

Part of the overarching aim of this thesis was also to *facilitate a broader multivariate perspective which complements previous approaches to detect extreme events*. Hence, one question to ask concerns differences between the multivariate and univariate perspective. What do we really gain from a multivariate extreme event detection compared to a common univariate extreme event detection?

The disadvantages of a multivariate detection scheme may hinder their application in analysis of extreme events. First, it is more complicated to understand the methods behind a multivariate perspective. I experienced it to be much more complicated to communicate a multivariate detection in the climate science community. Second, the multivariate detection scheme as used in this thesis, detects any type of extreme event in any direction of the variables, independent of whether a variable has values near the upper or lower bounds of the marginal distribution. Assigning one or several variables and a direction of the extreme event is a second step after the detection

step. Thus, in any application of a multivariate extreme event detection scheme, these two disadvantages have to be kept in mind.

However, a multivariate perspective on extreme events has several advantages. First, the detection in any direction of the variables can also be an advantage, as it makes the multivariate detection scheme per se more objective than a standard univariate detection scheme. For example, an increased productivity anomaly prior to the Russian heatwaves 2010, has not been studied in detail before (chapter 4), presumably due to a focus on higher impacts, i.e. on negative anomalies in productivity. However, note that multivariate directional approaches exist, which are used to detect extreme events as well. These approaches are e.g. based on copulas or principal component analysis (Torres et al., 2017).

Second, the multivariate approach is very generic and can detect various types of anomalies in many different variables (Chapter 3). Thus, it is easily possible to streamline different views on the very same extreme event with a multivariate perspective. Different views can originate from looking at the very same extreme event with different variables, often utilizing variables which are commonly used in the discipline of the researcher. A multivariate perspective e.g. on the Russian heatwave facilitates a broader view on the area which was hit by the multivariate extreme event. Focussing on single variables only, one can potentially miss parts which are affected by the inherently multivariate event (e.g. a focus on temperature solely misses 52% of the affected multivariate volume, a focus on surface moisture misses 31% of the affected volume (see Supplementary materials S1.1).

Third, sensitivity with respect to the chosen threshold for detecting extreme events is simplified. In the case of multivariate extreme event detection, one multivariate threshold to detect extreme events needs to be adjusted for sensitivity analysis. In the case of univariate extreme event detection, one threshold needs to be adjusted for each variable, and potentially combined with all possible combinations from other variables. The number of possible combinations hinders sensitivity analysis when combining several univariate extreme event detection schemes.

Fourth, multivariate extreme event detection is more precise than univariate extreme event detection as it uses information from multivariate covariate patterns (Santos-Fernandez, 2013). The developed multivariate extreme event detection scheme follows the shape of the multivariate distribution, i.e. linear multivariate methods like the mahalanobis distance to the mean of the data (Hotelling's T^2) accounts for correlation among the variables. Thus, multivariate algorithms are stricter than univariate thresholds, when looking at the direction aligned with the main correlation patterns. This means that events are usually closer to the upper or lower bound of the marginal distributions in this direction. Anomalies in this direction, which are identified in a univariate setting but not in a multivariate are usually considered to be false

positive detections in the univariate setting (ignoring the covariance structure) (Garrett, 1989). On the contrary, in the orthogonal direction, which is misaligned with the main correlation pattern, the multivariate algorithm detects events which would not be detected by a univariate detection scheme. Recovering these non-detected true anomalies in a univariate setting may be a much harder task than figuring out the false positive detections (Garrett, 1989). Kernel density estimation is used in the thesis to furthermore account for potential non-linearities in the general multivariate pattern.

An alternative to multivariate extreme event detection schemes is to apply a univariate peak-over threshold scheme repeatedly. However, it has to be applied with care as dependency structures are not accounted for, and a rigorous application may lead to an overestimation of extreme events (Garrett, 1989). Applying univariate thresholds for each variable (blue box) overestimates the total number of extreme events with an increasing number of variables. For example, a 5% threshold for each of three uniformly distributed variables would lead to flag $(1 - 0.05)^3 = 0.95^3 = 85.7\%$ as normal, and respectively 14.3% of the data as extreme, whereas a multivariate algorithm exactly flags the chosen 5% to be extreme. This setting would lead to a difference in the number of data points flagged as extreme between the univariate and the multivariate setting, i.e. 2.86 more events flagged as extreme in the repeatedly applied univariate setting. This factor is lowered by covariance structures in the data, but only disappears for perfect correlation between the variables.

Last, a multivariate perspective has the potential to detect novelties. In chapter 4, it was shown that the objectiveness and possibility to apply a multivariate detection scheme to any set of variables facilitates a different perspective and yields interesting new results. The result of contrasting vegetation responses to well studied climate extremes like the Russian heatwave 2010 is one example. However, similar results would have been possible by applying univariate extreme event detection schemes to several variables repeatedly, and considering any direction of the limits of a marginal distribution and then merging the results. An up-until now "unknown" type of an extreme event may be an event, which can be detected by a multivariate algorithm but not by a univariate detection scheme. For linear dependency structures, these kind of events are located mostly orthogonal to the main covariance structure of the variables. Therefore, these events appear to be historically much more rare, smaller and of lower impacts than expected (see Supplementary S1.2), and were therefore not analysed in detail in this thesis. However, with climate change, an intensification of those events may be expected in regions where the climate moves in the direction orthogonal to the main covariance structure (Mahony & Cannon, 2018).

6.8 Outlook

The developed multivariate approach (chapter 3) and the studies of ecosystem-specific responses for a case study (chapter 4) and globally, (chapter 5) open up a variety of possible follow-up research questions. More case studies could be used to study differentiated responses of vegetation types, as well as trying to disentangle impacts in 2019 associated with legacy effects of 2018 and from instantaneous impacts in 2019. Another option would be to put more focus on the "unknown" events, i.e. events which can be detected by a multivariate approach, but not by a univariate approach. The latter option was considered but not explored in detail in this thesis, due to the rare number of this type of extreme events. Furthermore, it would be possible to focus more on the near real-time detection and prediction of multivariate extreme events, which would require monitoring and analysis of constantly incoming data streams. An automated operational detection scheme would be necessary to exercise this, which can be set up on the basis of this thesis. However, in the following I want to emphasize two further research directions: One is to further tap into the potential of the multivariate algorithms and to develop a multivariate typology of extreme events, which can be applied to historical data and might be of interest in the context of climate change (Section 6.8.1). Another direction is to use the tested methods of chapter 3 and to adjust them to detect novel climates and the timing when a region or ecosystem departs from what one would consider normal variability in this region (Section 6.8.2).

6.8.1 Further potentials of the multivariate perspective

One potential of the developed multivariate perspective on extreme events is to fully built on the potential of the multivariate algorithm to detect extreme events in any direction of the variables. With the developed multivariate approach, it is possible to distinguish many more nuanced differences among climate extreme events, e.g. one can not only detect droughts, heatwaves and compounding droughts and heatwaves (like in chapter 5, but also droughts and heatwaves with extremely high, medium or extremely low radiation, probably differentiated even further with any other (climate) variable. This kind of refined differentiation of extreme events would allow for developing an entire typology of extreme events. Typologies have been developed e.g. for different types of flooding (Merz & Blöschl, 2003), crop vulnerability to droughts (Simelton et al., 2009), or climate change adaptation strategies (Biagini et al., 2014). In the attribution community, i.e. attributing climate extreme events to climate change, (Otto, 2017) calls for establishing a typology of the potential attribution of climate extremes to human induced climate change. However, a typology of multivariate climate extremes would already be novel as such, to the best of my

knowledge and facilitates further research questions.

First, a typology of extreme events could be used to identify potential high impact events. Although it is expected that compound events are associated with higher impacts than their univariate equivalent (e.g. only drought, not heat) (Zscheischler et al., 2018), this is usually only shown for specific types of extreme events, e.g. for impacts of compounding droughts and heatwaves on vegetation productivity (Buttler et al., 2018). This generalisation of impacts of compound events can be further differentiated e.g. it can be assumed that compounding droughts and heatwaves with concurrent high wind speed increase wildfire risk (Abatzoglou et al., 2018; Di Virgilio et al., 2019; Pinol et al., 1998; Sharples, 2009). Furthermore, compounding droughts and heatwaves with concurrent high radiation are associated with different impacts on vegetation productivity than compounding droughts and heatwaves with concurrent low radiation (for a more detailed example of a typology of extreme events see Supplementary materials S1.2). Also other impacts, like impacts on humans, or specifically impacts targeted for agricultural system, can be of interest in order to develop (regionally adapted) early warning systems for different types of extreme events, which are expected to lead to highest impacts as learnt from history in the respective region.

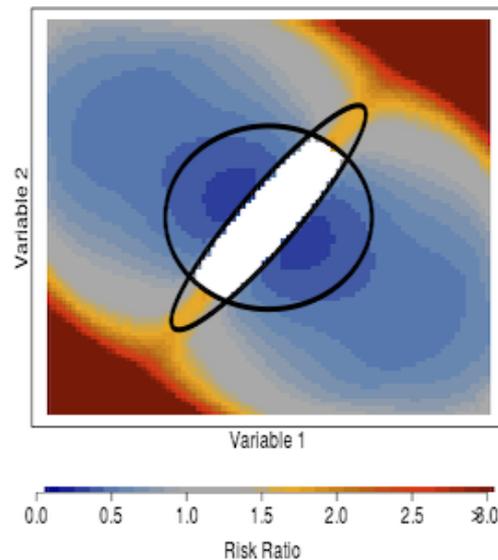


Figure 6.1 – Change in the risk ratio of two multivariate normally distributed variables, compared to (univariate) risk assessment ignoring the covariance structure.

Second, a multivariate perspective on and typology of extreme events facilitates the refinement of our understanding of changing patterns in the extremes with climate change. For example, the dependency structure between temperature and precipitation and future changes therein lead to a higher frequency of compounding droughts and heatwave than one would expect from univariate risk assessment in historical

data (Sarhadi et al., 2018) and in model projections of future climate (Zscheischler & Seneviratne, 2017) (Figure 6.1). Similarly, including more variables may reveal more nuanced details in changes of the future dependency structures, and thus point towards much higher increase in numbers of specific types of extremes than one would expect from the change in the marginal distribution. Vice versa, other types of extreme events may get less frequent not only due to trends in a warming climate, but also due to regionally changing dependency structures in the earth system models.

6.8.2 Novel climates

Another potential future research direction is the usage of the multivariate perspective to compute a multivariate measure of the time of emergence / novel climates.

Climate models can be used to compute the time when the climate departs from normal variability in the respective region for a specific climate variable. This is usually referred to as time of emergence from natural climate variability or novel climates, and is an important indicator for regional risk assessment from climate models (Hawkins & Sutton, 2012). It is usually computed for single variables e.g. for precipitation (Giorgi & Bi, 2009; Sui et al., 2014), temperature (Frame et al., 2017; Hawkins & Sutton, 2012; Mahlstein et al., 2011; Sui et al., 2014), sea levels (Lyu et al., 2014), ocean biogeochemistry (Keller et al., 2014), and ice cover (Screen & Deser, 2019). In principle the idea is applicable to any (climate) variable (Hawkins & Sutton, 2012), and has even been applied to climate extremes (King et al., 2015).

However, the idea of the time of emergence from natural variability is not restricted to marginal distributions. It may be an inherently multivariate question of several climate variables. A first multivariate approach to detect novel climates in a bivariate distribution of temperature and precipitation was developed by (Mahony et al., 2017) and applied to the global scale (Mahony & Cannon, 2018). The technique is based on the Mahalanobis distance to the mean of the data, which is also used in chapter 3. One main result in this context of novel climates is that the departure from normal variability is taking place earlier compared to a univariate perspective, in particular when the associated climate change signal is not aligned with the main direction of the correlation between temperature and precipitation **see conceptual figure** (Mahony & Cannon, 2018).

One further research potential in the context of novel climates is to extend previously developed approaches with a set of variable describing the climate of a specific region in its entirety. First applications of the time of emergence to three variables (precipitation, daily maximum and minimum temperatures) are currently developed in order to analyse migratory bird populations (La Sorte et al., 2019). However, including variables like radiation or wind speeds would allow for a much more detailed description of the current regional climate in a specific region. It can be expected

that similarly to the results of Mahony & Cannon, 2018 the departure time from normal variability is even faster than for two variables, when the direction of the departure is not aligned with the multivariate covariance structure. One particularly challenging aspect may be the visualisation of the departure from normal variability in more than three dimensions.

Furthermore, including variables important to vegetation development, like water availability, may allow to characterize novel climates with respect to normal conditions for ecosystems. Another possibility would be to compute a time of emergence for indices of vegetation productivity, like the fraction of absorbed photosynthetic active radiation, solar induced fluorescence, or primary productivity.

Another potential in the context of multivariate novel climates is to built on the techniques which were tested and applied in this thesis. So far, techniques based on the Mahalanobis distance are used for first multivariate applications on novel climates (La Sorte et al., 2019; Mahony & Cannon, 2018; Mahony et al., 2017). Although the simple interpretation is clearly an advantage of the Mahalanobis distance, it assumes a multivariate normal distribution of the data. Hence, the Mahalanobis distance is limited to detect deviations from linear dependency structures. Linear dependencies may not be the case for climate variables. Thus, using nonlinear techniques to compute the time of emergence may facilitate a refined view on novel climates. One possibility of a nonlinear technique is kernel density estimation, which was used in chapter 4 and 5 to detect extreme events. However, many more methods are possible, as computing the time of emergence is a well defined novelty detection problem in the context of computer science (see chapter 2.2.2). The reference period of normal variability can be considered as training data of any (supervised) novelty detection algorithm. One promising method is the kernel null foley sammon transform (Bodesheim et al., 2013), which was not suitable for the (unsupervised) detection of extreme events (chapter 3). However, it might be suitable for detecting novel climates, because it can accurately map the differences to the training data (even in multiple classes) (Bodesheim et al., 2013) which may be an advantage for detecting novel climates.

Chapter 7

Concluding summary

The overarching aim of the thesis was to *improve the detection and understanding of climate extremes and their impacts by facilitating a broader multivariate perspective which complements previous approaches to detect extreme events.*

First, in chapter 3, the objective was to evaluate which *combination of multivariate anomaly detection algorithms and feature extraction is best suitable for detecting anomalous events.* In this paper, I identified several suitable methods for extreme event detection. The methods are ready to be applied to earth observations. They can be applied to any set of variables and the chosen multivariate algorithms are objective, i.e. detect extreme events in any direction of the variable's multivariate distribution. In particular, one detection scheme was identified to be suitable for high dimensional, correlated, and non-linear data. The selected detection scheme consists of subtracting the seasonality, reducing the dimensionality and accounting for linear correlations in the data via principal component analysis, and detecting multivariate extreme events in the reduced feature space via non-linear kernel density estimators. Detection schemes in earth system science are commonly univariate (ignore correlations between the data). Multivariate approaches so far are based on the Mahalanobis distance to the mean of the distribution (restricted to linear settings) or based on copulas (not applicable in high dimensional settings and direction dependent). Thus, compared to other approaches, the developed approach has the advantage to be non-linear, applicable in high-dimensional settings and to account for dependencies in the data.

Second, the objective of chapter 4 was to apply the newly developed detection scheme to one well known event, and to *evaluate whether a broader multivariate perspective facilitates our understanding of extreme events and their impacts by revealing previously overlooked facets.* In this context, it was possible to reveal enhanced spring productivity prior to the Russian heatwave 2010 due to the objectiveness of the algorithm. Furthermore, it was possible to detect enhanced productivity in higher latitudes during the Russian heatwave 2010, and to complement different perspec-

tives on the Russian heatwave which were based on the focus on specific variables. Ecosystem type was identified to be the most important factor which is associated with a contrasting response between enhanced productivity of forest ecosystems and tremendously reduced productivity of agricultural ecosystems. Confounding factors are latitude (e.g. soil type, absolute temperatures), and duration of the heatwave. Regions associated with enhanced productivity can also be found in maps of already published articles of this famous extreme event. However, due to a different focus on mostly negative impacts, the enhanced productivity was previously overlooked. In this paper, I suspect that overseeing enhanced productivity might also be related to the choice of a common direction when detecting extreme events and their impacts. Contrastingly, the multivariate algorithms facilitate a broader perspective because they are directionless by definition.

Third, one directly subsequent question is to evaluate whether this kind of differentiated and potentially contrasting productivity response to the Russian heatwave is just a single case or is a more frequent phenomena in current climate around the world. This leads to the objective in chapter 5, i.e. to *evaluate the importance of different vegetation types on shaping the impact of climate extremes relative to other factors*. This paper shows that apart from background climate, duration of the event is the most important factor in determining impacts of growing season droughts and heatwaves globally. However, the direction of the impact (enhanced vs. reduced productivity) is associated with different vegetation types (forests vs. grasslands and crops). The study shows forests are often associated with normal to slightly enhanced productivity during droughts and heatwaves globally. It shows that the contrasting productivity as seen during the Russian heatwave is a rather frequent phenomenon, instead of just a singular case. Furthermore, the differentiated productivity according to different vegetation types is a spatially contrasting effect, which complements previous studies on temporally contrasting compensation effect (spring vs. summer extreme events).

Overall, the overarching aim of the thesis was to *improve the detection and understanding of climate extremes and their impacts by facilitating a broader multivariate perspective which complements previous approaches to detect extreme events*. It was possible to facilitate a broader perspective and to detect multivariate droughts and heatwaves. Furthermore, the thesis improved the understanding of the impacts of droughts and heatwaves on different vegetation types.

However, limitations are e.g. present in the selection of the variables (e.g. lacking the influence of soil). Furthermore, the study was limited to remote sensing retrieved gross primary productivity, which transfers potential biases of this technique onto the results of the differentiated vegetation response to climate extremes. Although the results of differentiated vegetation responses and in particular a less sensitive

response of forests to droughts and heatwaves may be plausible in many cases, I also suspect that the sensitivity of current remote sensing derived productivity estimates to drought and heat effects is limited and should be an issue for future research. The focus of analyzing the impacts of climate extremes, was the instantaneous responses during the growing season. Lagged responses of different vegetation types still need to be scrutinized further. An example of this kind of lagged impacts would be the bark beetle outbreaks which can currently (2019) be observed in Germany's forests, presumably due to last year's drought (2018).

The multivariate perspective allows to easily extend the focus on droughts and heatwaves in this thesis to other types of extreme events, such as frost damage, heavy precipitation or wind throw. Furthermore, the multivariate perspective allows to develop an entire multivariate typology of extreme events, in order to identify current high-impact types of extreme events, potential changes in future climate, or to rely on multivariate methods to detect novel climates in the future and the departure time from normal variability.

Acknowledgments

First, I would like to thank my supervisor Miguel Mahecha, whose door was always open. I am grateful for innumerable scientific discussions, endless inspiring ideas, but also guidance and focus when necessary. In this context I also acknowledge the possibility and freedom of collaborations in the research group "Empirical Inference", and the many opportunities to participate in workshops and conferences.

Second, special thanks goes to my supervisor Alexander Brenning, who always took his time when needed. I am especially grateful for his advice on statistical issues, his support on precise formulations in the manuscripts and his critical view from a step-back-view.

Third, I am grateful to Markus Reichstein for the opportunity to join his dynamically growing, inspiring and vibrant department, his scientific insights and discussions, and for his calm way in any situation.

My thanks goes also to all of the co-authors of the manuscripts. Specifically to Joachim Denzler, Erik Rodner, and Paul Bodesheim for their methodological expertise and support especially in the first stage of this thesis. My special thanks goes also to Sebastian Sippel for scientific and non-scientific discussions, for many ideas always embedded in a broader context, and his positive but also critical feedback on any scientific topic, including the introduction of this thesis. Furthermore I would like to thank Fabian Gans for being always there, for his support in coding, his endless help with debugging, as well as his critical views and many technical and non-technical ideas which substantially influenced this thesis. I am also grateful to Sebastian Bathiany for inspiring discussions, a meteorological perspective on the topic and for hosting me and my family in Wageningen.

Furthermore, I am grateful to several people providing the basics. Michael Hauhs and Holger Lange for teaching me time series analysis and introducing me into a scientific way of thinking. In this context, I would also like to thank Christopher Krich, Sophia Walther, as well as Clara and Andreas Neudecker for a place to sleep in Jena when needed and the possibility to talk about anything. I am grateful to Leona Flach for English proofreading and my parents for their support in any situation.

Last, I would like to thank my wife Ilka for her continuous support in any situation and my children Tamo and Emil for their charming way after work.

Bibliography

- Abatzoglou, J. T., J. K. Balch, B. A. Bradley, & C. A. Kolden (2018). Human-related ignitions concurrent with high winds promote large wildfires across the USA. *International Journal of Wildland Fire* 27 (6), doi: 10.1071/WF17149.
- AghaKouchak, A, D Feldman, M Hoerling, T Huxman, & J Lund (2015). Recognize anthropogenic drought. *Nature* 524, 409–411.
- AghaKouchak, A., L. Cheng, O. Mazdiyasi, & A. Farahmand (2014). Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought. *Geophys. Res. Lett* 41, 8847–8852.
- Allen, C. D., D. D. Breshears, & N. G. McDowell (2015). On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere* 6, 129.
- Anderegg, W. R. L. et al. (2015). Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. *Science* 349 (6247), 524–528.
- Arnone III, J. A. et al. (2008). Prolonged suppression of ecosystem carbon dioxide uptake after an anomalously warm year. *Nature* 455 (7211), 383–386.
- Åström, D. O., B. Forsberg, & J. Rocklöv (2011). Heat wave impact on morbidity and mortality in the elderly population: A review of recent studies. *Maturitas* 69 (2), 99–105.
- Barriopedro, D., F. E. M, J. Luterbacher, R. M. Trigo, & R Garcia-Herrera (2011). The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe. *Science* 332 (6027), 220–224.
- Barz, B., Y. Guaniche, E. Rodner, & J. Denzler (2017). Maximally Divergent Intervals for Extreme Weather Event Detection, doi: 10.1109/OCEANSE.2017.8084569.
- Bastos, A, C. M. Gouveia, R. M. Trigo, & S. W. Running (2014). Analysing the spatio-temporal impacts of the 2003 and 2010 extreme heatwaves on plant productivity in Europe. *Biogeosciences* 11 (13), 3421–3435.
- Bauer, T. E. S. & K. W. Bauer (2015). A Comparison of Multivariate Outlier Detection Methods For Finding Hyperspectral Anomalies, doi: 10.1109/AERO.2007.353062.
- Beenen, J., C. von Eichhorn, S. Höll, U. Nimz, M. Schories, & M. Sprenger (2018). *Trockenheit in Deutschland. Auf Grund. Deutschland erlebt gerade die größte Dürre seit Beginn der Messungen - weil es schon seit Beginn des Jahres zu wenig regnet. Eine Reise zu Gewässern, denen das Wasser ausgeht.* In: *Süddeutsche Zeitung. Panorama*. <https://projekte.sueddeutsche.de/artikel/panorama/duerre-in-deutschland-e407144/> accessed on September, 30th., 2019.
- Beer, C. et al. (2010). Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science* 329 (5993), 843–838.
- Begzsuren, S, J. E. Ellis, D. S. Ojima, M. B. Coughenour, & T Chuluun (2004). Livestock responses to droughts and severe winter weather in the Gobi Three Beauty National Park, Mongolia. *Journal of Arid Environments* 59 (4), 785–796.
- Benotti, M. J., B. D. Stanford, & S. A. Snyder (2010). Impact of Drought on Wastewater Contaminants in an Urban Water Supply. *Journal of Environment Quality* 39 (4), 1196–1200.

- Berry, H. L., K. Bowen, & T. Kjellstrom (2009). Climate change and mental health: a causal pathways framework. *International Journal of Public Health* 55 (2), 123–132.
- Bersimis, S, S Psarakis, & J Panaretos (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International* 23 (5), 517–543.
- Bevacqua, E., D. Maraun, I. Hobæk Haff, M. Widmann, & M. Vrac (2017). Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences* 21 (6), 2701–2723.
- Biagini, B., R. Bierbaum, M. Stults, S. Dobardzic, & S. M. McNeeley (2014). Global Environmental Change. *Global Environmental Change* 25, 97–108.
- Bjerke, J. W., S. R. Karlsen, K. A. Høgda, E. Malnes, J. U. Jepsen, S. Lovibond, D. Vikhamar-Schuler, & H. Tømmervik (2014). Record-low primary productivity and high plant damage in the Nordic Arctic Region in 2012 caused by multiple weather events and pest outbreaks, 1–15.
- Black, E, M Blackburn, G Harrison, B Hoskins, & J Methven (2006). Factors contributing to the summer 2003 European heatwave. *Weather* 59, 217–223.
- Black, T. A., W. J. Chen, A. G. Barr, M. A. Arain, Z Chen, Z Nestic, E. H. Hogg, H. H. Neumann, & Yang, P. C. (2000). forest in years with a warm spring. *Geophysical Research Letters* 27, 1271–1274.
- Bodesheim, P., A. Freytag, E. Rodner, M. Kemmler, & J. Denzler (2013). Kernel Null Space Methods for Novelty Detection. *CVPR*, 3374–3381.
- Bokhorst, S, J. W. Bjerke, F. W. Bowles, J Melillo, T. V. Callaghan, & G. K. Phoenix (2008). Impacts of extreme winter warming in the sub-Arctic: growing season responses of dwarf shrub heathland. *Global Change Biology* 14, 2603–2612.
- Borken, W & E Matzner (2009). Reappraisal of drying and wetting effects on C and N mineralization and fluxes in soils. *Global Change Biology* 15 (4), 808–824.
- Brando, P. M. et al. (2014). Abrupt increases in Amazonian tree mortality due to drought-fire interactions. *Proceedings of the National Academy of Sciences* 111 (17), 6347–6352.
- Buckley, L. B. & M. S. Foushee (2011). Footprints of climate change in US national park visitation. *International Journal of Biometeorology* 56 (6), 1173–1177.
- Buermann, W., P. R. Bikash, M. Jung, D. H. Burn, & M. Reichstein (2013). Earlier springs decrease peak summer productivity in North American boreal forests. *Environ. Res. Lett.* 8 (2), 024027.
- Buermann, W. et al. (2018). Widespread seasonal compensation effects of spring warming on northern plant productivity. *Nature* 562, 110–111.
- Buras, A., A. Rammig, & C. S. Zang (2019). Quantifying impacts of the drought 2018 on European ecosystems in comparison to 2003. *Biogeosciences Discussions*, doi: 10.5194/bg-2019-286.
- Buttler, J. von et al. (2018). Impacts of droughts and extreme-temperature events on gross primary production and ecosystem respiration: a systematic assessment across ecosystems and climate zones. *Biogeosciences* 15 (5), 1293–1318.
- Calow, R. C., A. M. MacDonald, A. L. Nicol, & N. S. Robins (2010). Ground Water Security and Drought in Africa: Linking Availability, Access, and Demand. *Ground Water* 48 (2), 246–256.
- Cao, M & F. I. Woodward (1998). Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature* 393, 249–252.
- Chambers, J. Q., J. I. Fisher, H Zeng, E. L. Chapman, D. B. Baker, & G. C. Hurtt (2007). Hurricane Katrina’s Carbon Footprint on U.S. Gulf Coast Forests. *Science* 318, 1107.
- Chandola, V., A. Banerjee, & V. Kumar (2009). Anomaly detection. *ACM Computing Surveys* 41 (3), 1–58.
- Chaves, M. M., J Flexas, & C Pinheiro (2009). Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany* 103, 551–560.

- Chen, J., J. F. Franklin, & T. A. Spies (1993). Contrasting microclimates among clearcut, edge, and interior of old-growth Douglas-fir forest. *Agricultural and Forest Meteorology* 63, 219–237.
- Ciais, P. et al. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437 (7058), 529–533.
- Coumou, D. & S. Rahmstorf (2012). A decade of weather extremes. *Nature Climate Change* 2 (7), 491–496.
- Coumou, D. & A. Robinson (2013). Historic and future increase in the global land area affected by monthly heat extremes. *Environ. Res. Lett.* 8 (3), doi:10.1088/1748-9326/8/3/034018.
- Dai, A., D. Luo, Sönk, & J. Liu (2019). Arctic amplification is caused by sea-ice loss under increasing CO. *Nature Communications* 121, doi: 10.1038-s41467-018-07954-9.
- Darenova, E., P. Holub, L. Krupkova, & M. Pavelka (2017). Effect of repeated spring drought and summer heavy rain on managed grassland biomass production and CO₂ efflux. *Journal of Plant Ecology* 10 (3), 476–485.
- Davidson, E. A. & I. A. Janssens (2006). Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* 440 (7081), 165–173.
- De Bono, A, P Peduzzi, S Kulser, & G Giuliani (2004). Impacts of Summer 2003 Heat Wave in Europe. *United Nations Environmental Programme*, <https://archive\bibrangedashouverte.unige.ch/unige:32255>.
- De Michele, C (2003). A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *J. Geophys. Res* 108 (D2), doi:10.1029/2002JD002534.
- Denton, E. M., J. D. Dietrich, M. D. Smith, & A. K. Knapp (2016). Drought timing differentially affects above-and belowground productivity in a mesic grassland. *Plant Ecology* 218 (3), 317–328.
- Di Virgilio, G., J. P. Evans, S. A. P. Blake, M. Armstrong, A. J. Dowdy, J. Sharples, & R. McRae (2019). Climate Change Increases the Potential for Extreme Wildfires. *Geophys. Res. Lett* 136 (1), 513.
- Donges, J. F., C. F. Schleussner, J. F. Siegmund, & R. V. Donner (2016). Event coincidence analysis for quantifying statistical interrelationships between event time series. *The European Physical Journal Special Topics* 225 (3), 471–487.
- Donges, J. F., R. V. Donner, M. H. Trauth, N. Marwan, H.-J. Schellnhuber, & J. Kurths (2011). Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. *PNAS* 108, 20422–20427.
- Doughty, C. E. et al. (2015). Drought impact on forest carbon dynamics and fluxes in Amazonia. *Nature* 519 (7541), 78–82.
- DPA (2019). *Unser Wald ist massiv geschädigt*. In: *Spiegel Online*. <https://www.spiegel.de/wissenschaft/natur/julia-kloeckner-beraet-mit-bundeslaendern-ueber-zustand-des-waldes-a-1279984.html> accessed on August, 19th., 2019.
- Drescher, M. & S. C. Thomas (2012). Snow cover manipulations alter survival of early life stages of cold-temperate tree species. *Oikos* 122 (4), 541–554.
- DWD (2018). *Erste Bilanz des Deutschen Wetterdienstes zum Jahr 2018 in Deutschland*. In: *Pressemitteilung des DWD*. https://www.dwd.de/DE/presse/pressemitteilungen/DE/2018/20181220_jahr2018_rekord_news.html;jsessionid=2D01A3C0A203924A32B278F1A84ED0DE.live21062 accessed on September, 30th., 2019.
- Dwivedi, S., K. Sahrawat, H. Upadhyaya, & R. Ortiz (2013). “Food, Nutrition and Agrobiodiversity Under Global Climate Change”. *Adv. Agron.*, doi: 10.1016/B978-0-12-407686-0.00001-4.
- Easterling, D. R., G. A. Meehl, C. Parmesan, T. R. K. Changnon, & L. O. Mearns (2000). Climate Extremes: Observations, Modeling, and Impacts. *Science* 289 (5487), 2068–2074.

- Eisenhauer, N., S. Cesarz, R. Koller, K. Worm, & P. B. Reich (2011). Global change belowground: impacts of elevated CO₂, nitrogen, and summer drought on soil food webs and biodiversity. *Global Change Biology* 18 (2), 435–447.
- Epstein, P (2001). Climate change and emerging infectious diseases. *Microbes and Infection* 3, 747–754.
- Erb, K.-H., V. Gaube, F. Krausmann, C. Plutzer, A. Bondeau, & H. Haberl (2007). A comprehensive global 5 min resolution land-use data set for the year 2000 consistent with national census data. *Journal of Land Use Science* 2 (3), 191–224.
- Fan, Y., G. Miguez-Macho, E. G. Jobbágy, R. B. Jackson, & C. Otero-Casal (2017). Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences* 82, 201712381.
- Faranda, D. & S. Vaienti (2013). A recurrence-based technique for detecting genuine extremes in instrumental temperature records. *Geophys. Res. Lett* 40 (21), 5782–5786.
- Feng, M., J. O. Sexton, C. Huang, A. Anand, S. Channan, X.-P. Song, D.-X. Song, D.-H. Kim, P. Noojipady, & J. R. Townshend (2016). Remote Sensing of Environment. *Remote Sensing of Environment* 184 (C), 73–85.
- Filleul, L. et al. (2006). The Relation Between Temperature, Ozone, and Mortality in Nine French Cities During the Heat Wave of 2003. *Environmental Health Perspectives* 114 (9), 1344–1347.
- Fischer, E. M. & R. Knutti (2013). Robust projections of combined humidity and temperature extremes. *Nature Climate Change* 3 (2), 126–130.
- Fischer, E. M. & R. Knutti (2014). Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett* 41, 541–554.
- Fischer, E. M., S. I. Seneviratne, P. L. Vidale, D. Lüthi, & C. Schär (2007). Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave. *Journal of Climate* 20 (20), 5081–5099.
- Frame, D., M. Joshi, E. Hawkins, L. J. Harrington, & M. de Roiste (2017). Population-based emergence of unfamiliar climates. *Nature Climate Change* 7 (6), 407–411.
- Frank, D. et al. (2015). Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. *Global Change Biology* 21 (8), 2861–2880.
- Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, & R. Knutti (2014). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *Journal of Climate* 27 (2), 511–526.
- Garcia-Herrera, R, R. M. Trigo, J Diaz, & J Luterbacher (2010). A Review of the European Summer Heat Wave of 2003. *Critical Reviews in Environmental Science and Technology* 40, 267–306.
- Garrett, R. G. (1989). The chi-square plot: a tool for multivariate outlier recognition. *Journal of Geochemical Exploration* 32, 319–341.
- Genest, C, A. C. Favre, J Béliveau, & C Jacques (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research* 43 (W09401), doi:10.1029/2006WR005275.
- Ghil, M et al. (2011). Extreme events: dynamics, statistics and prediction. *Nonlinear Processes in Geophysics* 18 (3), 295–350.
- Giorgi, F. & X. Bi (2009). Time of emergence (TOE) of GHG-forced precipitation change hot-spots. *Geophys. Res. Lett* 36 (L06709), doi:10.1029/2009GL037593.
- Gleick, P. H. (2014). Water, Drought, Climate Change, and Conflict in Syria. *Weather, Climate, and Society* 6 (3), 331–340.
- Gould, E. A. & S Higgs (2009). Impact of climate change and other factors on emerging arbovirus diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103 (2), 109–121.

- Granier, A., N. Bréda, B. Longdoz, P. Gross, & J. Ngao (2008). Ten years of fluxes and stand growth in a young beech forest at Hesse, North-eastern France. *Annals of Forest Science* 64 (704), doi:10.1051-forest:2008052.
- Griffin, D & K. J. Anchukaitis (2015). How unusual is the 2010-2014 California drought? *Geophys. Res. Lett* 41, 9017–9023.
- Guanche, Y., E. Rodner, M. Flach, S. Sippel, M. D. Mahecha, & J. Denzler (2016). “Detecting Multivariate Biosphere Extremes”. *Proceedings of the 6th International Workshop on Climate Informatics: CI2016*. Ed. by A Banerjee, W Ding, & V Dy. Boulder: National Center for Atmospheric Research: NCAR Technical Note NCAR/TN-529+PROC, 9–12.
- Guo, M., J. Li, J. Xu, X. Wang, H. He, & L. Wu (2017). CO₂emissions from the 2010 Russian wildfires using GOSAT data. *Environmental Pollution* 226, 60–68.
- Hanrahan, B (2018). “Dwindling Rhine paralyzes shipping transport”. *Handelsblatt*.
- Hansen, A., P. Bi, M. Nitschke, P. Ryan, D. Pisaniello, & G. Tucker (2008). The Effect of Heat Waves on Mental Health in a Temperate Australian City. *Environmental Health Perspectives* 116 (10), 1369–1375.
- Hansen, J., M. Sato, & R. Ruedy (2012). Perception of climate change. *PNAS* 109, E2415–E2423.
- Harmeling, S., G. Dornhege, D. Tax, F. Meinecke, & K.-R. Müller (2006). From outliers to prototypes: Ordering data. *Neurocomputing* 69 (13-15), 1608–1618.
- Hawkins, E & R Sutton (2012). Time of emergence of climate signals. *Geophys. Res. Lett* 39 (L01702), doi:10.1029/2011GL050087.
- Heerwaarden, C. C. van & A. J. Teuling (2014). Disentangling the response of forest and grassland energy exchange to heatwaves under idealized land–atmosphere coupling. *Biogeosciences* 11 (21), 6159–6171.
- Hendrix, C. S. & I. Salehyan (2012). Climate change, rainfall, and social conflict in Africa. *Journal of Peace Research* 49 (1), 35–50.
- Hicke, J. A., J. A. Logna, J Powell, & D. S. Ojima (2006). Changing temperatures influence suitability for modeled mountain pine beetle (*Dendroctonus ponderosae*) outbreaks in the western United States. *Journal of Geophysical Research* 111, doi:10.1029–2005JG000101.
- Hirschi, M., S. I. Seneviratne, V. Alexandrov, F. Boberg, C. Boroneant, O. B. Christensen, H. Formayer, B. Orlowsky, & P. Stepanek (2010). Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience* 4 (1), 17–21.
- Huang, M., X. Wang, T. F. Keenan, & S. Piao (2018). Drought timing influences the legacy of tree growth recovery. *Global Change Biology* 24 (8), 3546–3559.
- Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, & P. M. Cox (2013). No increase in global temperature variability despite changing regional patterns. *Nature* 500 (7462), 327–330.
- Ivits, E., S. Horion, R. Fensholt, & M. Cherlet (2014). Drought footprint on European ecosystems between 1999 and 2010 assessed by remotely sensed vegetation phenology and productivity. *Global Change Biology* 20 (2), 581–593.
- Jentsch, A. & C. Beierkuhnlein (2008). Research frontiers in climate change: Effects of extreme meteorological events on ecosystems. *C. R. Geoscience* 340, 612–628.
- Jentsch, A. et al. (2011). Climate extremes initiate ecosystem-regulating functions while maintaining productivity. *Journal of Ecology* 99 (3), 689–702.
- Jung, M. et al. (2017). Compensatory water effects link yearly global land CO₂ sink changes to temperature. *Nature* 541 (7638), 516–520.
- Kantz, H. (2005). *Extremes*. New York Heidelberg Dordrecht London: Springer.
- Katz, R. W. & B. G. Brown (1992). Extreme events in a changing climate: variability is more important than averages. *Climatic Change* 21, 289–302.

- Keenan, R. J., G. A. Reams, F. Achard, J. V. de Freitas, A. Grainger, & E. Lindquist (2015). Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015t. *Forest Ecology and Management* 352, 9–20.
- Keenan, T. F., S Sabate, & C Gracia (2010). The importance of mesophyll conductance in regulating forest ecosystem productivity during drought periods. *Global Change Biology* 16, 1019–1034.
- Keller, K. M., F Joos, & C. C. Raible (2014). Time of emergence of trends in ocean biogeochemistry. *Biogeosciences* 11 (13), 3647–3659.
- King, A. D., M. G. Donat, E. M. Fischer, E. Hawkins, L. V. Alexander, D. J. Karoly, A. J. Dittus, S. C. Lewis, & S. E. Perkins (2015). The timing of anthropogenic emergence in simulated climate extremes. *Environ. Res. Lett.* 10 (9), 094015.
- Knorr, E. M., R. T. Ng, & V Tucakov (2000). Distance-based outliers: algorithms and applications. *The VLDB Journal* 8, 237–253.
- Köppe, J. (2019). Förster warnen vor nächstem Waldsterben. In: *Spiegel Online*. <https://www.spiegel.de/wissenschaft/natur/waldsterben-in-deutschland-foerster-warnen-vor-klimakrise-a-1279002.html> accessed on August, 19th., 2019.
- Kreyling, J., M. Wenigmann, C. Beierkuhnlein, & A. Jentsch (2008). Effects of Extreme Weather Events on Plant Productivity and Tissue Die-Back are Modified by Community Composition. *Ecosystems* 11 (5), 752–763.
- Kurz, W. A., C. C. Dymond, G Stinson, G. J. Rampley, E. T. Neilson, A. L. Carroll, T Ebata, & L Safranyik (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature* 452 (7190), 987–990.
- La Sorte, F. A., D. Fink, & A. Johnston (2019). Time of emergence of novel climates for North American migratory bird populations. *Ecography* 42 (6), 1079–1091.
- Larcher, W. (2003). *Physiological plant ecology : ecophysiology and stress physiology of functional groups*. Berlin, New York: Springer.
- Le Quéré, C. et al. (2018). Global Carbon Budget 2018. *Earth System Science Data* 10 (4), 2141–2194.
- Leonard, M., S. Westra, A. Phatak, M. Lambert, B. van den Hurk, K. McInnes, J. Risbey, S. Schuster, D. Jakob, & M. Stafford-Smith (2014). A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change* 5, 113–128.
- Leys, C., C. Ley, O. Klein, P. Bernard, & L. Licata (2013). Journal of Experimental Social Psychology. *Journal of Experimental Social Psychology* 49 (4), 764–766.
- Li, W, N MacBean, P Ciais, P Defourny, C Lamarche, S Bontemps, R. A. Houghton, & S. Peng (2018). Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015). *Earth System Science Data* 10, 219–234.
- Lindroth, A., F. Lagergren, A. Grelle, L. Klemedtsson, O. Langvall, P. Weslien, & J. Tuulik (2009). Storms can cause Europe-wide reduction in forest carbon sink. *Global Change Biology* 15 (2), 346–355.
- Lobell, D. B., A. Sibley, & J. I. Ortiz-Monasterio (2012). Extreme heat effects on wheat senescence in India. *Nature Climate Change* 2 (3), 186–189.
- Lopez-Nicolas, A, M Pulido-Velazquez, & H Macian-Sorribes (2017). Journal of Hydrology. *Journal of Hydrology* 550, 580–589.
- Lowry, C. A. & D. C. Montgomery (1995). A review of multivariate control charts. *IIE Transactions* 27, 800–810.
- Lowry, C. A. & W. H. Woodall (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* 34, 46–53.

- Lyu, K., X. Zhang, J. A. Church, A. B. A. Slangen, & J. Hu (2014). Time of emergence for regional sea-level change. *Nature Climate Change* 4, 1006–1010.
- Ma, X, A. Huete, S Moran, G. Ponce Capos, & D. Eamus (2015). Abrupt shifts in phenology and vegetation productivity under climate extremes. *J. Geophys. Res* 120, 2036–2052.
- Mach, K. J. et al. (2019). Climate as a risk factor for armed conflict. *Nature*, 193–197.
- Mahecha, M. D., F. Gans, S. Sippel, J. F. Donges, T. Kaminski, S. Metzger, M. Migliavacca, D. Papale, A. Rammig, & J. Zscheischler (2017). Detecting impacts of extreme events with ecological in situ monitoring networks. *Biogeosciences* 14 (18), 4255–4277.
- Mahlstein, I, R Knutti, S Solomon, & R. W. Portmann (2011). Early onset of significant local warming in low latitude countries. *Environ. Res. Lett.* 6 (3), 034009.
- Mahony, C. R. & A. J. Cannon (2018). Wetter summers can intensify departures from natural variability in a warming climate. *Nature Communications* 9, 783.
- Mahony, C. R., A. J. Cannon, T. Wang, & S. N. Aitken (2017). A closer look at novel climates: new methods and insights at continental to landscape scales. *Global Change Biology* 16, 476.
- McLeman, R & B Smit (2006). Migration as an Adaptation to Climate Change. *Climatic Change* 76 (1-2), 31–53.
- McPhillips, L. E. et al. (2018). Defining Extreme Events: A Cross-Disciplinary Review. *Earth's Future* 6 (3), 441–455.
- Meehl, G. A. & C. Tebaldi (2004). More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science* 305 (5686), 994–997.
- Merz, R & G Blöschl (2003). A process typology of regional floods. *Water Resources Research* 39 (12), 1065.
- Meze-Hausken, E (2000). Migration caused by climate change: How vulnerable are people in dryland areas? *Mitigation and Adaptation Strategies for Global Change* 5, 379–406.
- Michailidis, P. D. (2013). Accelerating Kernel Density Estimation on the GPU Using the CUDA Framework. *Applied Mathematical Sciences* 7, 1447–1476.
- Miller, N. L., K. Hayhoe, J. Jin, & M. Auffhammer (2008). Climate, Extreme Heat, and Electricity Demand in California. *Journal of Applied Meteorology and Climatology* 47 (6), 1834–1844.
- Minobe, S, K.-Y. A, N Komori, S.-P. Xie, & R. J. Small (2008). Influence of the Gulf Stream on the troposphere. *Nature* 452, 206–211.
- Miralles, D. G., P Gentine, S. I. Seneviratne, & A. J. Teuling (2019). Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Ann. N.Y. Acad. Sci.* 1436, 19–35.
- Miri, A, A Ghanbari, & A Moghaddamnia (2007). Dust Storms Impacts on Air Pollution and Public Health under Hot and Dry Climate. *International Journal of Energy and Environment* 1 (2), 101–105.
- Monfreda, C., N. Ramankutty, & J. A. Foley (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles* 22 (GB1022), doi:10.1029/2007GB002947.
- Mueller, B. & S. I. Seneviratne (2012). Hot days induced by precipitation deficits at the global scale. *PNAS* 109, 12398–12403.
- Mueller, V, C Gray, & K Kosec (2014). Heat stress increases long-term human migration in rural Pakistan. *Nature Climate Change* 4 (3), 182–185.
- Myneni, R. B., J Dong, C. J. Tucker, R. K. Kaufmann, P Kauppi, L Zhou, V Alexevey, & M. K. Hughes (2001). A large carbon sink in the woody biomass of Northern forests. *PNAS* 98, 14784–14789.

- Nawrotzki, R. J., L. M. Hunter, D. M. Runfola, & F. Riosmena (2015). Climate change as a migration driver from rural and urban Mexico. *Environmental Research Letters* 10 (11), 114023.
- Nemani, R. R., C. D. Keeling, H Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, & S. W. Running (2003). Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999. *Science* 300, 11560–15645.
- Nicholls, N. & L. Alexander (2007). Has the climate become more variable or extreme? Progress 1992-2006. *Progress in Physical Geography: Earth and Environment* 31 (1), 77–87.
- Obermeier, W. A. et al. (2016). Reduced CO₂ fertilization effect in temperate C₃ grasslands under more extreme weather conditions. *Nature Climate Change* 7 (2), 137–141.
- Otto, F. E. L. (2017). Attribution of Weather and Climate Events. *Annual Review of Environment and Resources* 42 (1), 627–646.
- Overland, J. E. & M. Wang (2010). Large-scale atmospheric circulation changes are associated with the recent loss of Arctic sea ice. *Tellus A: Dynamic Meteorology and Oceanography* 62 (1), 1–9.
- Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics* 33, 1065–1076.
- Pendergrass, A. G., R. Knutti, F. Lehner, C. Deser, & B. M. Sanderson (2017). Precipitation variability increases in a warmer climate. *Scientific Reports*, 17966.
- Pereira, L. S., T Oweis, & A Zairi (2002). Irrigation management under water scarcity. *Agricultural Water Management* 57, 175–206.
- Pimentel, M. A. F., D. A. Clifton, L. Clifton, & L. Tarassenko (2014). A review of novelty detection. *Signal Processing* 99 (C), 215–249.
- Pinol, J, J Terradas, & F. LLoret (1998). Climate Warming, Wildfire Hazard, and Wildfire Occurrence in Coastal Eastern Spain. *Climatic Change* 38, 345–357.
- Portmann, F. T., S. Siebert, & P Döll (2010). MIRCA2000 – Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles* 24, GB1011.
- Prospero, J. M. & P. J. Lamb (2003). African Droughts and Dust Transport to the Caribbean: Climate Change Implications. *Science* 302, 1024–1028.
- Raffa, K. F. (2001). Mixed messages across multiple trophic levels: the ecology of bark beetle chemical communication systems. *Chemoecology* 11, 49–65.
- Ramaswamy, S., R. Rastogi, & K. Shim (2000). Efficient Algorithms for Mining Outliers from Large Data Sets. *SIGMOD record* 29 (2), 427–438.
- Rammig, A, M Wiedermann, J. F. Donges, F Babst, W von Bloh, D Frank, K Thonicke, & M. D. Mahecha (2015). Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity. *Biogeosciences* 12 (2), 373–385.
- Rawald, T., M. Sips, & N. Marwan (2017). Computers & Geosciences. *Computers and Geosciences* 104, 101–108.
- Reeves, J, C Foelz, P Grace, & P Best (2010). *Impacts and adaptation response of infrastructure and communities to heatwaves: the southern Australian experience of 2009*. Queensland University of Technology: Report No. 192160915X.
- Reichstein, M. et al. (2007). Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Global Change Biology* 13 (3), 634–651.
- Reichstein, M. et al. (2013). Climate extremes and the carbon cycle. *Nature* 500 (7462), 287–295.
- Renard, B & M Lang (2007). Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources* 30 (4), 897–912.

- Rodner, E., B. Barz, Y. Guanche, M. Flach, M. D. Mahecha, P. Bodesheim, M. Reichstein, & J. Denzler (2016). Maximally Divergent Intervals for Anomaly Detection. *ICML Workshop on Anomaly Detection (ICML-WS)*, doi: 10.17871/BACI_ICML2016_Rodner.
- Roman, D. T., K. A. Novick, E. R. Brzostek, D. Dragoni, F. Rahman, & R. P. Phillips (2015). The role of isohydric and anisohydric species in determining ecosystem-scale response to severe drought. *Oecologia* 179 (3), 641–654.
- Rouault, G., J.-N. Candau, F. Lieutier, L.-M. Nageleisen, J.-C. Martin, & N. Warzée (2006). Effects of drought and heat on forest insect populations in relation to the 2003 drought in Western Europe. *Annals of Forest Science* 63 (6), 613–624.
- Rousseeuw, P. J. & M. Hubert (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (1), 73–79.
- Rousseeuw, P. J. & K Van Driessen (1990). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* 41, 212–223.
- Ruehr, N. K., A. Gast, C. Weber, B. Daub, & A. Arneth (2015). Water availability as dominant control of heat stress responses in two contrasting tree species. *Tree Physiology* 36, 164–178.
- Saatchi, S, S Asefi-Najafabady, Y Malhi, L. E. O. C. Aragao, L. O. Anderson, R. B. Myneni, & R Nemani (2013). Persistent effects of a severe drought on Amazonian forest canopy. *PNAS* 110, 565–570.
- Santos-Fernandez, E. (2013). *Multivariate Statistical Quality Control Using R*. SpringerBriefs in Statistics. New York Heidelberg Dordrecht London: Springer. DOI: 10.1007/978-1-4614-5453-3.
- Sarhadi, A, M. C. Ausin, M. P. Wiper, D Touma, & N. S. Diffenbaugh (2018). Multidimensional risk in a nonstationary climate: Joint probability of increasingly severe warm and dry conditions. *Science Advances*, eaau3487.
- Schär, C. & G Jendritzky (2004). Hot news from summer 2003. *Nature* 432, 559–560.
- Schewe, J. et al. (2019). State-of-the-art global models underestimate impacts from climate extremes. *Nature Communications*, doi: 10.1038/s41467-019-08745-6.
- Schoelzel, C. & P Friedrichs (2008). Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. *Nonlinear Processes in Geophysics* 15 (5), 761–772.
- Schwalm, C. R. et al. (2017). Global patterns of drought recovery. *Nature* 548 (7666), 202–205.
- Scott, D & C Lemieux (2010). Weather and Climate Information for Tourism. *Procedia Environmental Sciences* 1 (5), 146–183.
- Screen, J. A. & C Deser (2019). Pacific Ocean Variability Influences the Time of Emergence of a Seasonally Ice-Free Arctic Ocean. *Geophys. Res. Lett* 46 (4), 2222–2231.
- Seneviratne, S. I. et al. (2012). “Changes in climate extremes and their impacts on the natural physical environment”. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (IPCC SREX Report)*. Ed. by C. Field et al. Cambridge University Press, 109–230.
- Seneviratne, S. I., D. Lüthi, M. Litschi, & C. Schär (2006). Land–atmosphere coupling and climate change in Europe. *Nature* 443 (7108), 205–209.
- Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, & A. J. Teuling (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Science Reviews* 99 (3–4), 125–161.
- Seneviratne, S. I., M. G. Donat, B. Mueller, & L. V. Alexander (2014). No pause in the increase of hot temperature extremes. *Nature Climate Change* 4, 161–163.

- Sharples, J. J. (2009). An overview of mountain meteorological effects relevant to fire behaviour and bushfire risk. *International Journal of Wildland Fire* 18, 737–754.
- Shen, C., W.-C. Wang, Z. Hao, & W. Gong (2007). Exceptional drought events over eastern China during the last five centuries. *Climatic Change* 85 (3-4), 453–471.
- Siegmund, J. F., T. G. M. Sanders, I. Heinrich, E. van der Maaten, S. Simard, G. Helle, & R. V. Donner (2016). Meteorological Drivers of Extremes in Daily Stem Radius Variations of Beech, Oak, and Pine in Northeastern Germany: An Event Coincidence Analysis. *Frontiers in Plant Science*, doi: 10.3389/fpls.2016.00733.
- Sillmann, J, V. V. Kharin, X Zhang, F. W. Zwiers, & D Bronaugh (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres* 118 (4), 1716–1733.
- Simelton, E., E. D. G. Fraser, M. Termansen, P. M. Forster, & A. J. Dougill (2009). Typologies of crop-drought vulnerability: an empirical analysis of the socio-economic factors that influence the sensitivity and resilience to drought of three major food crops in China (1961–2001). *Environmental Science & Policy* 12 (4), 438–452.
- Sippel, S., J. Zscheischler, M. Heimann, F. E. L. Otto, J. Peters, & M. D. Mahecha (2015). Quantifying changes in climate variability and extremes: Pitfalls and their overcoming. *Geophys. Res. Lett* 42, 9990–9998.
- Sippel, S., J. Zscheischler, & M. Reichstein (2016). Ecosystem impacts of climate extremes crucially depend on the timing. *Proceedings of the National Academy of Sciences* 113 (21), 5768–5770.
- Sippel, S. et al. (2017). Warm winter, wet spring, and extreme response in ecosystem functioning on the Iberian Peninsula. *BAMS* 98, S80–S85.
- Sippel, S., M. Reichstein, X. Ma, M. D. Mahecha, H. Lange, M. Flach, & D. Frank (2018). Drought, Heat, and the Carbon Cycle: a Review. *Current Climate Change Reports* 4 (3), 266–286.
- Sitch, S et al. (2008). Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Global Change Biology* 14, 2015–2039.
- Smale, D. A. et al. (2019). global biodiversity and the provision of ecosystem services. *Nature Climate Change* 9, 306–312.
- Smetek, T. E. & K. W. Bauer (2007). Finding Hyperspectral Anomalies Using Multivariate Outlier Detection. *Proc. 2007 IEEE Aerosp. Conf*, 1–24.
- Smith, M. D. (2011). An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research. *Journal of Ecology* 99 (3), 656–663.
- Smoyer-Tomic, K. E., R Kuhn, & A hudson (2003). Heat Wave Hazards: An Overview of Heat Wave Impacts in Canada. *Natural Hazards* 28, 463–485.
- Stocker, B. D., J. Zscheischler, T. F. Keenan, I. C. Prentice, S. I. Seneviratne, & J. Peñuelas (2019). Drought impacts on terrestrial primary production underestimated by satellite monitoring. *Nature Geoscience* 12, 264–270.
- Stouffer, R. J. et al. (2006). Investigating the Causes of the Response of the Thermohaline Circulation to Past and Future Climate Changes. *Journal of Climate* 19, 1365–1387.
- Sui, Y., X. Lang, & D. Jiang (2014). Time of emergence of climate signals over China under the RCP4.5 scenario. *Climatic Change* 125 (2), 265–276.
- Swann, A. L. S., F. M. Hoffman, C. D. Koven, & J. T. Randerson (2016). Plant responses to increasing CO₂ reduce estimates of climate impacts on drought severity. *Proceedings of the National Academy of Sciences* 113 (36), 10019–10024.
- Tax, D. M. & R. P. W. Duin (2004). Support Vector Data Description. *Machine Learning* 54, 45–66.

- Terrence McCabe, J (1987). Drought and Recovery: Livestock Dynamics among the Ngisonyoka Turkana of Kenya. *Human Ecology* 15, 371–389.
- Teuling, A. J. et al. (2010). Contrasting response of European forest and grassland energy exchange to heatwaves. *Nature Geoscience* 3 (10), 722–727.
- Torres, R., C. De Michele, H. Laniado, & R. E. Lillo (2017). Directional multivariate extremes in environmental phenomena. *Environmetrics* 28 (2), e2428.
- Tramontana, G. et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* 13 (14), 4291–4313.
- Trenberth, K. E., A. Dai, G. van der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, & J. Sheffield (2013). Global warming and changes in drought. *Nature Climate Change* 4 (1), 17–22.
- Uexkull, N. von, M Croicu, H Fjelde, & H. Buhaug (2016). Civil conflict sensitivity to growing-season drought. *PNAS* 113, 12391–12396.
- Ukkola, A. M., M. G. De Kauwe, A. J. Pitman, M. J. Best, G Abramowitz, V Haverd, M Decker, & N Haughton (2016). Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environ. Res. Lett.* 11 (104012), doi:10.1088/1748-9326/11/10/104012.
- Vale, M. M., D. J. Moura, I. A. Naas, & D. F. Pereira (2010). Characterization of Heat Waves Affecting Mortality Rates of Broilers Between 29 Days and Market Age. *Brazilian Journal Of Poultry Science* 12, 279–285.
- Van Lanen, H. A. J. et al. (2016). Hydrology needed to manage droughts: the 2015 European case. *Hydrological Processes* 30 (17), 3097–3104.
- Vannitsem, S (2006). Statistical properties of the temperature maxima in an intermediate order Quasi-Geostrophic model. *Tellus A: Dynamic Meteorology and Oceanography* 59 (1), 80–95.
- Velde, M. van der, F. N. Tubiello, A. Vrieling, & F. Bouraoui (2011). Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. *Climatic Change* 113 (3-4), 751–765.
- Vliet, M. T. H. van & J. J. G. Zwolsman (2008). Impact of summer droughts on the water quality of the Meuse river. *Journal of Hydrology* 353 (1-2), 1–17.
- Vogel, M. M., J Zscheischler, R Wartenburger, D Dee, & S. I. Seneviratne (2019). Concurrent 2018 Hot Extremes Across Northern Hemisphere Due to Human-Induced Climate Change. *Earth's Future* 7 (6026), 692–703.
- Vrac, M & P. Naveau (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research* 43 (7), 716.
- Wegren, S. K. (2010). Food Security and Russia's 2010 Drought. *Eurasian Geography and Economics* 51, 1–17.
- Weiß, M. (2019). *Was die kahlen Kronen lehren.* In: *Süddeutsche Zeitung*. <https://www.sueddeutsche.de/wissen/waldsterben-duerre-baeume-buchen-fichten-1.4550243> accessed on August, 19th., 2019.
- Westerling, A. L., H. G. Hidalgo, D. R. Cayan, & T. W. Swetnam (2006). Warming and Earlier Spring INcrease Western U.S. Forest Wildfire Activity. *Science* 313 (5789), 940–943.
- Wetz, M. S. & D. W. Yoskowitz (2013). An 'extreme' future for estuaries? Effects of extreme climatic events on estuarine water quality and ecology. *Marine Pollution Bulletin* 69 (1-2), 7–18.
- White, T. A., B. D. Campell, P. D. Kemp, & C. L. Hunt (2000). Sensitivity of three grassland communities to simulated extreme temperature and rainfall events. *Global Change Biology* 6, 671–684.
- Williams, A. P. (2012). Temperature as a potent driver of regional forest drought stress and tree mortality. *Nature Climate Change* 3 (3), 292–297.

- Wolf, S. et al. (2016). Warm spring reduced carbon cycle impact of the 2012 US summer drought. *Proceedings of the National Academy of Sciences* 113 (21), 5880–5885.
- Yang, Y., R. J. Donohue, & T. R. McVicar (2016). Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water Resources Research* 52 (10), 8260–8276.
- Yeung, D.-Y. & C. Chow (2002). Parzen-window Network Intrusion Detectors. *Proceedings of the 16th International Conference on Pattern Recognition (IEEE)* 4, 385–388.
- Yi, K., D. Dragoni, R. P. Phillips, D. T. Roman, & K. A. Novick (2017). Dynamics of stem water uptake among isohydric and anisohydric species experiencing a severe drought. *Tree Physiology* 37 (10), 1379–1392.
- Yuste, J. C., J Penuelas, M Estiarte, J Garcia-Mas, S Mattana, R Ogaya, M Pujol, & J Sardans (2010). Drought-resistant fungi control soil organic matter decomposition and its response to temperature. *Global Change Biology* 17 (3), 1475–1486.
- Zaehle, S. & D. Dalmonech (2011). Carbon-nitrogen interactions on land at global scales: current understanding in modelling climate biosphere feedbacks. *Current Opinion in Environmental Sustainability* 3 (5), 311–320.
- Zander, K. K., W. J. W. Botzen, E. Oppermann, T. Kjellstrom, & S. T. Garnett (2015). Heat stress causes substantial labour productivity loss in Australia. *Nature Climate Change* 5 (7), 647–651.
- Zhou, L., Y. Xia, J. M. W. Brownjohn, & K. Y. Koo (2016). Temperature Analysis of a Long-Span Suspension Bridge Based on Field Monitoring and Numerical Simulation. *Journal of Bridge Engineering* 21 (1), 04015027.
- Zhou, S., Y. Zhang, A Park Williams, & P Gentine (2019). Projected increases in intensity, frequency, and terrestrial carbon costs of compound drought and aridity events. *Science Advances* 5, eaau5740.
- Zscheichler, J. et al. (2014). Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data. 28 (6), 585–600.
- Zscheichler, J., M. Reichstein, S Harmeling, A Rammig, E Tomelleri, & M. D. Mahecha (2014a). Extreme events in gross primary production: a characterization across continents. *Biogeosciences* 11 (11), 2909–2924.
- Zscheichler, J. & S. I. Seneviratne (2017). Dependence of drivers affects risks associated with compound events. *Science Advances* 3, 10.1126-sciadv.1700263.
- Zscheichler, J., M. D. Mahecha, S. Harmeling, & M. Reichstein (2013). Detection and attribution of large spatiotemporal extreme events in Earth observation data. *Ecological Informatics* 15 (C), 66–73.
- Zscheichler, J. et al. (2014b). A few extreme events dominate global interannual variability in gross primary production. *Environ. Res. Lett.* 9 (3), 035001.
- Zscheichler, J., M. Reichstein, J von Buttler, M. Mu, J. T. Randerson, & M. D. Mahecha (2014c). Carbon cycle extremes during the 21st century in CMIP5 models: Future evolution and attribution to climatic drivers. *Geophys. Res. Lett.* 41, 8853–8861.
- Zscheichler, J., R. Orth, & S. I. Seneviratne (2015). A submonthly database for detecting changes in vegetation-atmosphere coupling. *Geophys. Res. Lett.* 42, 9816–9824.
- Zscheichler, J. et al. (2018). Future climate risk from compound events. *Nature Climate Change* 8, 469–477.

List of Figures

1.1	'How exceptional was the year 2018?' Deviation of temperature and precipitation between 1881 - 2018 in Germany. Red (blue) colors indicate too hot (cold) temperatures, green (orange) colors indicate too wet (dry) years in terms of precipitation. Figure from DWD, 2018. . .	2
2.1	Detecting extreme events with a peak over threshold.	14
2.2	A random sample of multivariate normal distributed data to illustrate Hotelling's T^2 . Colors denote Hotellings's T^2 scores, black lines are isolines of similar mahalanobis distsance to the mean. Figure modified after Santos-Fernandez, 2013	17
2.3	Pairwise distances of observations at different time instances in the three dimensional space of temperature, surface moisture and radiation deviations from the mean seasonal cycle; depicted is one pixel affected by the Russian Heatwave 2010, for more information on variables see Chapter 5.	18
6.1	Change in the risk ratio of two multivariate normally distributed variables, compared to (univariate) risk assessment ignoring the covariance structure.	96
S1	Sensitivity of the regionalisation procedure assessed by changing the number of used principal components which are used for classification on a North-South transect covering Africa and Europe. In general, very low differences are found for detecting extreme events. However, some regions (e.g. West Africa, Albania) show slightly high deviations in the anomaly scores, when changing the number of principal components.	131

List of Tables

S1	Part of the RHW event one would potentially miss, when focussing on single variables only.	129
S2	Example of establishing a systematic typology of extreme events in three variables (temperature, radiation and water availability), and associated instantaneous negative impacts calculated as deviation of the fraction of photosynthetic absorbed active radiation from mean seasonal cycle (FAPAR). FAPAR is summarized as mean globally. The number in brackets denotes two different thresholds to detect extreme events with the applied detection scheme of chapter 5. The table shows that compounding droughts and heatwaves with concurrent extremely high radiation are associated with highest impacts (strongest reduction in FAPAR).	130

Übersicht der Manuskripte

Die folgenden Manuskripte sind Bestandteil dieser Dissertation:

- *Flach, M., Brenning, A., Gans, F., Reichstein, M., Sippel, S. & Mahecha, M. D. (2020):* Vegetation modulates the impact of climate extremes on gross primary production, *Biogeosciences Discuss.*, doi: 10.5194/bg-2020-80, in review.*
- *Flach, M., Sippel, S., Gans, F., Bastos, A., Brenning, A., Reichstein, M., & Mahecha, M. D. (2018):* Contrasting biosphere responses to hydrometeorological extremes: revisiting the 2010 western Russian Heatwave, *Biogeosciences*, 15, 6067–6085, doi: 10.5194/bg-15-6067-2018. (*Selected as Highlight Paper*)
- *Flach, M., Gans, F., Brenning, A., Denzler, J., Reichstein, M., Rodner, E., Bathiany, S., Bodesheim, P., Guanache, Y., Sippel, S., & Mahecha, M. D. (2017):* Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques, *Earth Syst. Dynam.*, 8, 677–696, doi: 10.5194/esd-8-677-2017.

* eine Bestätigung des Verlags folgt auf der nächsten Seite.

From: editorial@copernicus.org
Subject: bg-2020-80 (author) - manuscript accepted for discussion
Date: 23. March 2020 at 12:38
To: mflach@bgc-jena.mpg.de

E

Dear Milan Flach,

We are pleased to inform you that your following manuscript was accepted for public review and discussion in BGD, the scientific discussion forum of BG:

Title: Vegetation modulates the impact of climate extremes on gross primary production
Author(s): Milan Flach et al.
MS No.: bg-2020-80
MS Type: Research article
Iteration: Initial Submission
Special Issue: Understanding compound weather and climate events and related impacts (BG/ESD/HESS/NHESS inter-journal SI)

We kindly ask you to upload your manuscript as *.pdf file no later than 02 Apr 2020 at:
https://editor.copernicus.org/BG/production_file_upload/bg-2020-80

Your discussion paper will not be typeset. Instead, your author's *.pdf file will be used and a citation header will be added by us. This has to be previewed in the above-mentioned production file upload interface.

Please use full first names for all authors. Although references are still based on initials, we will use full first names on the title page of your paper (HTML and XML).

The manuscript preparation guidelines can be found at:
https://www.biogeosciences.net/for_authors/submit_your_manuscript.html

Before file upload, please consider submitting data sets, model code, or video supplements to reliable repositories, receive DOIs, and cite these assets in your manuscript including entries in the reference list.

To log in, please use your Copernicus Office user ID 302962.

You are invited to monitor the processing of your manuscript via your MS Overview:
https://editor.copernicus.org/BG/my_manuscript_overview

In case any questions arise, please do not hesitate to contact me.

Kind regards,

Natascha Töpfer
Copernicus Publications
Editorial Support
editorial@copernicus.org

on behalf of the BG Editorial Board

Erklärung zu den Eigenanteilen

Erklärung zu den Eigenanteilen des Promovenden sowie der weiteren Doktorandinnen/Doktoranden als Co-Autorinnen/-Autoren an den Publikationen und Zweitpublikationsrechten bei einer kumulativen Dissertation

Für alle in dieser kumulativen Dissertation verwendeten Manuskripte liegen die notwendigen Genehmigungen der Verlage ("Reprint permissions") für die Zweitpublikation vor.

Die Co-Autorinnen/-Autoren der in dieser kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung, als auch über die angegebenen Eigenanteile der weiteren Doktorandinnen/Doktoranden als Co-Autorinnen/-Autoren an den Publikationen und Zweitpublikationsrechten bei einer kumulativen Dissertation informiert und stimmen dem zu.

Die Anteile des Promovenden sowie der weiteren Doktorandinnen/Doktoranden als Co-Autorinnen/Co-Autoren an den Publikationen und Zweitpublikationsrechten bei einer kumulativen Dissertation sind in der Anlage aufgeführt.

Milan Flach

Promovend

Ort

Datum

Unterschrift

Ich bin mit der Abfassung der Dissertation als publikationsbasierte Dissertation, d.h. kumulativ, einverstanden und bestätige die vorstehenden Angaben.

Prof. Dr. Alexander
Brenning

Betreuer	Ort	Datum	Unterschrift
----------	-----	-------	--------------

Prof. Dr. Miguel D.
Mahecha

Betreuer	Ort	Datum	Unterschrift
----------	-----	-------	--------------

Angabe der Publikationsäquivalente

Erklärung zu den Eigenanteilen des Promovenden sowie der weiteren Doktorandinnen/Doktoranden als Co-Autorinnen/Co-Autoren an den Publikationen und Zweitpublikationsrechten bei einer kumulativen Dissertation

*Flach, M.*¹, Brenning, A²., Gans, F³., Reichstein, M⁴., Sippel, S.⁵ & Mahecha, M. D.⁶ (2020): Vegetation modulates the impact of climate extremes on gross primary production, *Biogeosciences Discuss.*, doi: 10.5194/bg-2020-80, in review.

Autor	1	2	3	4	5	6
Konzeption des Forschungsansatzes	X	X	X	X	X	X
Planung	X					X
Datenerhebung	X		X			
Datenanalyse und -interpretation	X	X				X
Schreiben des Manuskripts	X	X	X	X	X	X
Vorschlag Anrechnung Publikationsäquivalente	1					

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und unter Verwendung der angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt habe.

Hersbruck, 25. April 2020

Milan Flach

Chapter S1

Supplementary material

S1.1 Multivariate perspective compared to single variables

One very simple way to illustrate the advantage of the multivariate perspective on the Russian heatwave is to calculate the percentage of the anomaly one would miss compared to the multivariate combination of variables, when focussing on a single variable of interest.

The proportion one would potentially miss ranges between 13% (rH) to 89% (P) for the atmosphere (atmos-RHW) and 0% (GPP, LE) to 10% (FPAR) for the biosphere (bio-RHW) (Tab. S1). Thus, for the atmosphere anomaly, the multivariate combination of variables is essential to cover the whole anomalous area. In contrast, the biosphere variables already integrate over a wide range of processes and one could potentially focus on specific variables (GPP, LE), without missing parts of the anomaly.

Table S1 – Part of the RHW event one would potentially miss, when focussing on single variables only.

event / variable	T	R _g	P	rH	SM	GPP	LE	H	FPAR
atmos-RHW	0.52	0.36	0.89	0.13	0.31	0.19	0.18	0.23	0.56
bio-RHW	0.75	0.36	0.95	0.29	0.57	0	0	0.08	0.10

S1.2 An exemplary typology of extreme events

In chapter 5, 180 droughts and heatwaves were selected from all 1000 largest extreme events, due to their expected high impacts on vegetation's productivity. This restriction to droughts, heatwaves and combined droughts and heatwaves is not necessary with a different objective. An detailed analysis of any of the largest extreme events may reveal more details. One could detect extreme events in three variables, e.g.

temperature, radiation and some index of water availability. First one can distinguish many more nuanced differences of droughts and heatwaves, e.g. compounding droughts and heatwaves with simultaneously high radiation, as well as normal or low radiation. The same differentiation is possible with droughts (or heatwaves), splitted according to temperature (or water availability) (normal, low), and radiation (normal, low, high). Going even further, it is possible to develop an entire typology of extreme events, e.g. building on $3^3 = 27$ possible combinations (3 variables, each variable three categories: high, normal, low). Evaluating differences in the impacts e.g. on gross primary productivity and trying to explain those could facilitate more facets in the understanding of climate extreme events and their impacts (Table S2).

index	FAPAR (5%)	FAPAR (3%)	Temperature	Radiation	Water availability
9	-0.070	-0.070	high	high	low
6	-0.063	-0.069	high	normal	low
18	-0.061	-0.067	high	high	normal
8	-0.060	-0.065	normal	high	low
4	-0.056	-0.064	low	normal	low
14	-0.055	-0.049	normal	normal	normal
27	-0.052	-0.042	high	high	high
15	-0.052	-0.057	high	normal	normal
17	-0.051	-0.054	normal	high	normal
22	-0.050		low	normal	high
1	-0.050	-0.051	low	low	low
24	-0.047	-0.041	high	normal	high
13	-0.047	-0.050	low	normal	normal
19	-0.046	-0.046	low	low	high
5	-0.043		normal	normal	low
10	-0.041	-0.048	low	low	normal
26	-0.040	-0.043	normal	high	high
20	-0.039	-0.049	normal	low	high
2	-0.038	-0.042	normal	low	low
16	-0.037		low	high	normal
11	-0.032	-0.036	normal	low	normal
7	-0.032	-0.016	low	high	low
23	-0.031	-0.035	normal	normal	high
12	-0.016		high	low	normal
3	-0.013	-0.009	high	low	low
25	-0.010	-0.015	low	high	high
21	-0.010	-0.005	high	low	high

Table S2 – Example of establishing a systematic typology of extreme events in three variables (temperature, radiation and water availability), and associated instantaneous negative impacts calculated as deviation of the fraction of photosynthetic absorbed active radiation from mean seasonal cycle (FAPAR). FAPAR is summarized as mean globally. The number in brackets denotes two different thresholds to detect extreme events with the applied detection scheme of chapter 5. The table shows that compounding droughts and heatwaves with concurrent extremely high radiation are associated with highest impacts (strongest reduction in FAPAR).

S1.3 Sensitivity of the regionalisation procedure

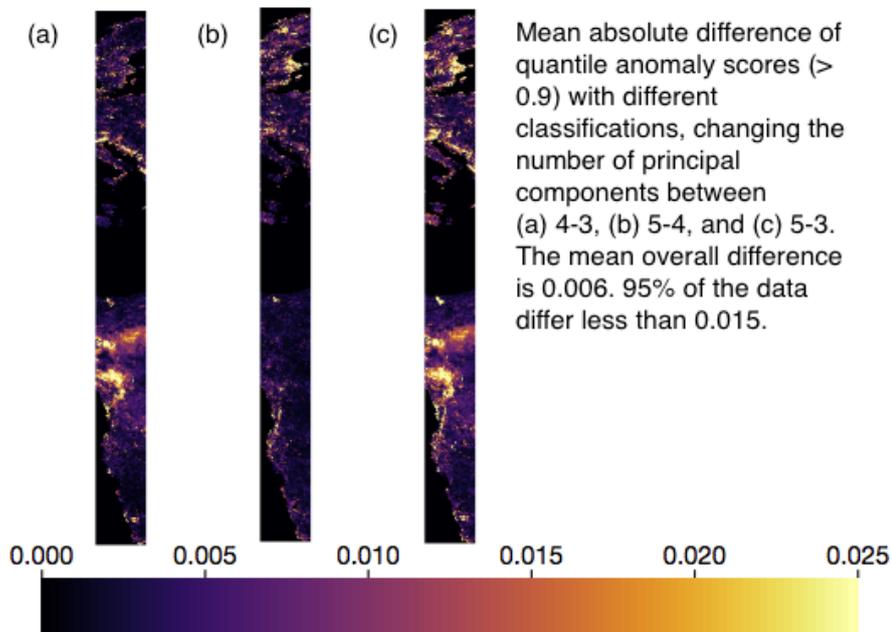


Figure S1 – Sensitivity of the regionalisation procedure assessed by changing the number of used principal components which are used for classification on a North-South transect covering Africa and Europe. In general, very low differences are found for detecting extreme events. However, some regions (e.g. West Africa, Albania) show slightly high deviations in the anomaly scores, when changing the number of principal components.

Chapter S2

Supplementantary material of chapter 3

Supplement of Earth Syst. Dynam., 8, 677–696, 2017
<https://doi.org/10.5194/esd-8-677-2017-supplement>
© Author(s) 2017. This work is distributed under
the Creative Commons Attribution 3.0 License.



Earth System
Dynamics  Open Access

Supplement of

Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques

Milan Flach et al.

Correspondence to: Milan Flach (milan.flach@bgc-jena.mpg.de)

The copyright of individual parts of the supplement might differ from the CC BY 3.0 License.

Supplementary Material 1 Parameterization of Recurrences and Kernel Density Estimation

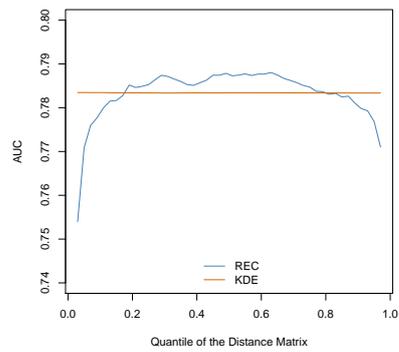


Figure S1. We test different choices of σ (Recurrences (*REC*)) or ε (for Kernel Density Estimation (*KDE*)) in a small simulation (500 repetitions) trying to detect a BaseShift. σ (or ε , respectively) is varied between the 0.05 and 0.95 quantile of the distribution of values of the distance matrix. The Area Under the receiver operator characteristics Curve (*AUC*) is computed for each parameterization. Results exhibit constant *AUC* values for *KDE* within the testing range of σ . In contrast *REC* is more sensitive to the choice of ε , although it might yield slightly higher *AUC* values in case of optimal chosen ε .

Supplementary Material 2 Effect of different data properties on the algorithms

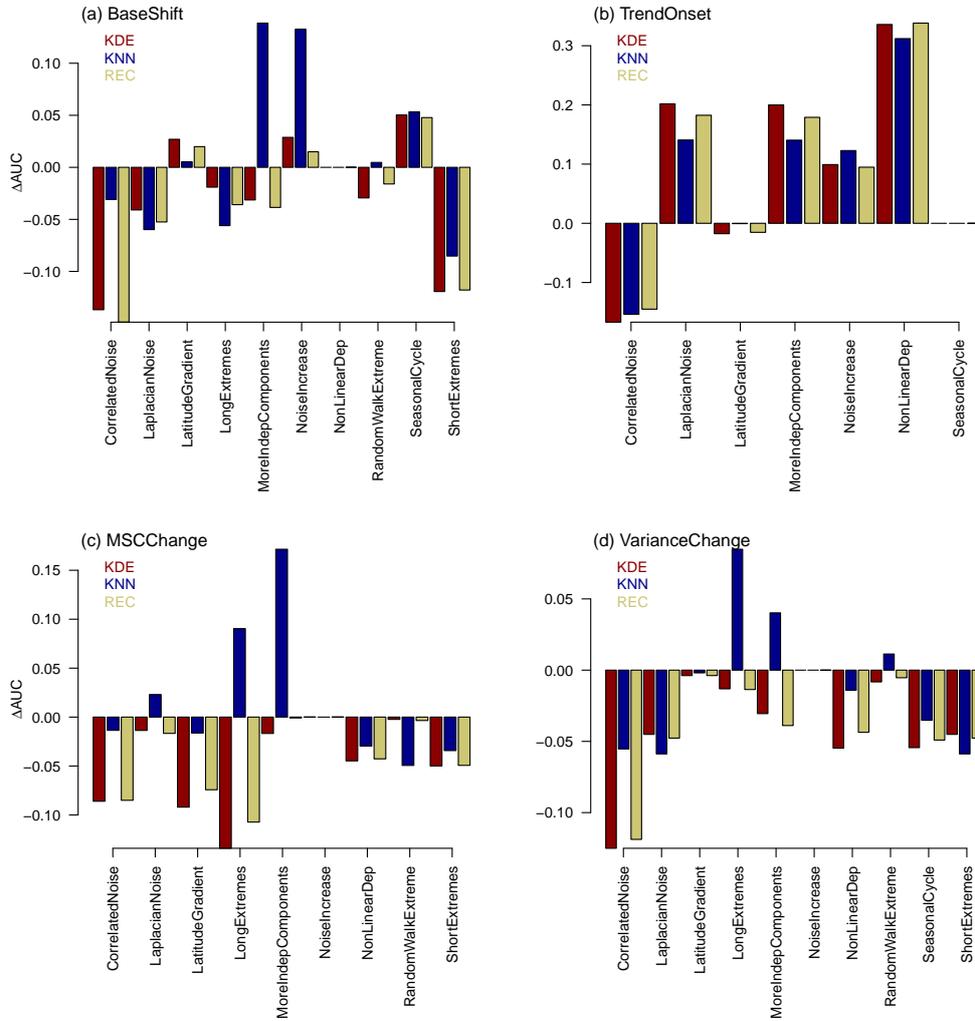


Figure S2. Effect of different data properties on the 3 best detection algorithms (KDE, REC, KNN-Gamma) presented as AUC difference to the UNIV control for the event types (a-d). Details for each algorithm reveal that KNN-Gamma is often less affected by 'difficult' data properties like CorrelatedNoise or MoreIndepComponents, i.e. KNN-Gamma is more adaptive than the other two algorithms.

Chapter S3

Supplementary material of chapter 5

Supporting Information for "Vegetation modulates the impact of climate extremes on gross primary production" by Flach et. al.

Case Studies

To illustrate the range of relative drought and heat anomalies, we report on different high and low impact extreme events in our data base in the following.

Case Study: European Heatwave 2003

The European heatwave 2003 (classified here as compounding drought and heatwave) started in April, peaked on the 14th of July and finished in September with a maximal duration of 104 days (Figure 1). The affected volume was $33 \cdot 10^6 \text{ km}^2\text{d}$. Total reduction in GPP is estimated here as $-11 \text{ gC m}^{-2}\text{month}^{-1}$, which is less than reported in (Ciais et al. 2005) ($-28 \text{ gC m}^{-2}\text{month}^{-1}$). Differences can be seen especially for forested ecosystems in the low mountain ranges (Vosges, Thuringian Forests, Black Forest, Ardennes, Rhenish Massif, Taunus) and a larger area with enhanced productivity around the Alps (Figure 1). Comparing our results with eddy covariance site level data (Ciais et al. 2005, Reichstein et al. 2007) from literature reveals that site level data and Fluxcom-RS agree on the direction of the GPP anomalies for DE-Hai (both negative), FR-Hes (both negative), FR-Lbr (both positive), and disagree for BE-Vie, DE-Tha (both positive by FLUXCOM-RS, negative at the tower site).

Case Study: US drought and heatwave 2012

The US drought and heatwave in 2012 started in April, peaked on June, 29th, and ended mid October with a maximal duration of 144 days. The affected volume was $12 \cdot 10^6 \text{ km}^2\text{d}$. Total reduction in GPP is estimated here as $-15 \text{ gC m}^{-2}\text{month}^{-1}$. However, on average forests increase their productivity by 13% ($22 \text{ gC m}^{-2}\text{month}^{-1}$). The one common affected tower site from literature (US-MMS) (Wolf et al. 2016) is in agreement with FLUXCOM-RS (both agree on a negative impact) (Figure 2).

Case Study: Russian Heatwave 2010

The Russian heatwave 2010 is one of the most impacting events in our data base (Fig. 3, classified here as compounding drought and heatwave).

It started in June, peaked on July, 19th, and ended in September with a maximal duration of 80 days. The affected volume was $87 \cdot 10^6 \text{ km}^2d$. The total reduction in GPP is estimated here as 72 TgC , roughly comparable to (Bastos et al. 2014) (90 TgC). We estimated losses in agricultural ecosystems to be 22% compared to the normal GPP average, which is comparable e.g. to estimates from (Wegren 2010) (grain harvest reduction of one third). However, we also estimate forests to enhance their productivity by 5% during the event. We could not find eddy covariance tower based estimates of carbon losses in literature.

Case Study: European heatwave 2018

The summer drought and heatwave in Mid-Europe 2018 started in June, peaked on July, 24th and lasted until August (Fig. 4). It was associated with a relative reduction in GPP by 10 % (6%) in agricultural (other) ecosystems. However, forests on average still are still productive as usual ($\pm 0 \%$). These differences between forests and agricultural ecosystems are also reported in (Buras et al. 2019).

Case Study: Drought in Brazil 2012

The drought in Brazil 2012 is another high impact event in our data base (Fig. 5). It started in April, peaked on May, 2nd, and lasted until September, with a maximum duration of 136 days. The affected volume was $24 \cdot 10^6 \text{ km}^2d$. Total reduction in carbon is estimated to be -36 TgC . We are not aware of any tower or model based comparison of this number. The affected area is in agreement with maps published in (Marengo et al. 2013).

Case Study: Horn of Africa 2009

The drought at the greater Horn of Africa 2009 (Fig. 6 is classified here as compounding drought and heatwave). It started in July, peaked on August, 14th, and lasted until September. The affected volume was $23 \cdot 10^6 \text{ km}^2d$. Total reduction in GPP is estimated to be 20 TgC . There have been no eddy covariance towers in the affected region 2009, but the event is also reported in (Nicholson 2014).

Case Study: Indian heatwave 2009

The Indian heatwave 2009 is one of the least impacting (most enhanced GPP) events in our data base (Fig. 7). It started in May, peaked at August, 1st. and lasted until September. The affected volume of the event is $45 \cdot 10^6 \text{ km}^2d$. The peak of the event is 53 days before peak of the growing season. Although it is classified as a relative drought and heatwave, there is still water available for large areas during the event (Fig. 7 (d), 25%, mean surface moisture during event). The conditions are associated with increased

productivity by 6 % (12%) in agricultural (forest) ecosystems and enhances total GPP by 13 TgC compared to the other years. The event presents the strongest enhancement of GPP of a single event in our data base. We can find warnings related to drought reduced agricultural yields in literature (Neena et al. 2011). However, the Organisation for Economic Co-operation and Development (OECD) reports a 4% increase in the wheat production for India 2009 (OECD 2009).

Case Study: Siberian heatwave 2011

The Siberian heatwave in 2011 another example of the least impacting event in our data base (Fig. 8). It starts in May, peaks on June, 1st. and lasts until the end of June. The affected volume is $31 \cdot 10^6 km^2d$. The peak of the heatwave is 41 days before peak growing season and it is has a maximum duration of 40 days, but shows a rather short duration for large parts of the event (Fig. 8). Water seems to be available during the heatwave (27%, mean surface moisture during event). These conditions are associated with an enhancement of GPP by 34 % (29 %) in forests (other) ecosystems (43 TgC) in other ecosystems. The event illustrates that a relative detection of drought and heat anomalies also includes events which are not severely affecting vegetation but can be beneficial for vegetation especially in Northern latitudes for which temperature and surface moisture conditions are still moderate.

Case Study: China drought 2011

The drought (and heatwave) in Southern China 2011 is one of the events in our data base (Fig. 9) showing strong differences among a East-West gradient. It starts in July 2006, peaks on August, 21th and lasts until September. The affected volume is $38 \cdot 10^6 km^2d$. Maximum duration is 72 days, 31 days after peak growing season. Eastern agricultural (-7 TgC) areas are strongly affected by the event in contrast to other ecosystems (+7 TgC) in the western parts of the affected area. These dipole-like structures can also be seen in maps of mean temperature and surface moisture during the event. Forests are partly associated with enhanced productivity, especially in regions with high temperatures and available surface moisture. This is interpreted as a positive reaction to the available radiation during the heatwave part of the event (Song et al. 2019).

East Europe 2015

The summer drought and heatwave affecting Europe with a focus on East Europe in 2015 started in June, peaked on August, 2nd and lasted until September. It was relatively short lasting for most of the affected area (< 20 days), and probably therefore affecting productivity not so strongly (Total

impact: -0.5 TgC agriculture, $+5 \text{ TgC}$ forests), although the affected volume was comparable to other high-impact extreme events ($21 \cdot 10^6 \text{ km}^2\text{d}$).

Amazon 2010

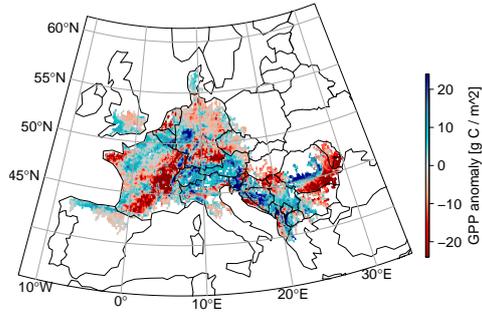
The Amazon drought 2010 is one of the well-known high-impact events of the last decade. It is classified here as a compounding drought and heatwave. It started in October 2009, peaked on January, 25th and lasts until March with a maximal duration of 184 days. It affected both, forests, and other ecosystems which reduced GPP by 7%, 9% respectively. The affected volume ($107 \cdot 10^6 \text{ km}^2\text{d}$) is among the highest in our data base (Fig. 11).

References

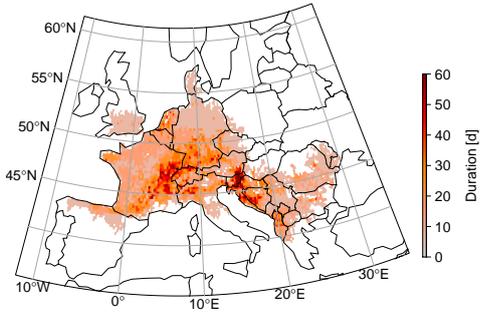
- Bastos, A., Gouveia, C. M., Trigo, R. M. & Running, S. W. (2014). Analysing the spatio-temporal impacts of the 2003 and 2010 extreme heatwaves on plant productivity in Europe, *Biogeosciences* **11**(13): 3421–3435.
- Buras, A., Rammig, A. & Zang, C. S. (2019). Quantifying impacts of the drought 2018 on european ecosystems in comparison to 2003, *Biogeosciences Discussions* **2019**: 1–23.
URL: <https://www.biogeosciences-discuss.net/bg-2019-286/>
- 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. & Valentini, R. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature* **437**(7058): 529–533.
- Marengo, J. A., Alves, L. M., Soares, W. R., Rodriguez, D. A., Camargo, H., Riveros, M. P. & Pabló, A. D. (2013). Two Contrasting Severe Seasonal Extremes in Tropical South America in 2012: Flood in Amazonia and Drought in Northeast Brazil, *Journal of Climate* **26**(22): 9137–9154.
- Neena, J. M., Suhas, E. & Goswami, B. N. (2011). Leading role of internal dynamics in the 2009 Indian summer monsoon drought, *Journal of Geophysical Research* **116**: D13103.
- Nicholson, S. E. (2014). Journal of Arid Environments, *Journal of Arid Environments* **103**(c): 71–79.
- OECD (2009). Oecd data: Crop production, In: <https://data.oecd.org/agroutput/crop-production.htm> accessed on July, 4th 2019.
- Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S., Viovy, N., Cramer, W., Granier, A., Ogée, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., Grünwald, T., Heimann, M., Heinesch, B., Knohl, A., Kutsch, W., Loustau, D., Manca, G., Matteucci, G.,

- Miglietta, F., Ourcival, J.-M., Pilegaard, K., Pumpanen, J., Rambal, S., Schaphoff, S., Seufert, G., Soussana, J. F., Sanz, M. J., Vesala, T. & Zhao, M. (2007). Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis, *Global Change Biology* **13**(3): 634–651.
- Song, L., Li, Y., Ren, Y., Wu, X., Guo, B., Tang, X., Shi, W., Ma, M., Han, X. & Zhao, L. (2019). Divergent vegetation responses to extreme spring and summer droughts in Southwestern China, *Agricultural and Forest Meteorology* **279**: doi: 10.1016/j.agrformet.2019.107703.
- Wegren, S. K. (2010). Food Security and Russia’s 2010 Drought, *Eurasian Geography and Economics* **51**: 1–17.
- Wolf, S., Keenan, T. F., Fisher, J. B., Baldocchi, D. D., Desai, A. R., Richardson, A. D., Scott, R. L., Law, B. E., Litvak, M. E., Brunsell, N. A., Peters, W. & van der Laan-Luijkx, I. T. (2016). Warm spring reduced carbon cycle impact of the 2012 US summer drought, *Proceedings of the National Academy of Sciences* **113**(21): 5880–5885.

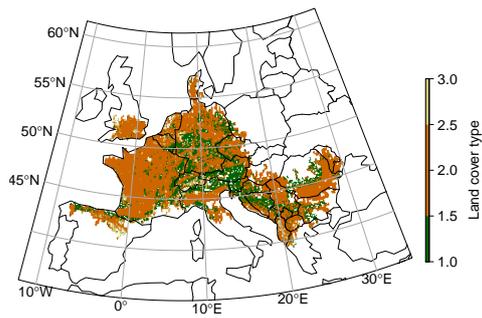
(a) GPP anomaly



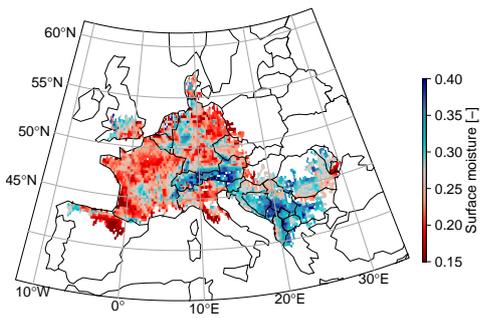
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

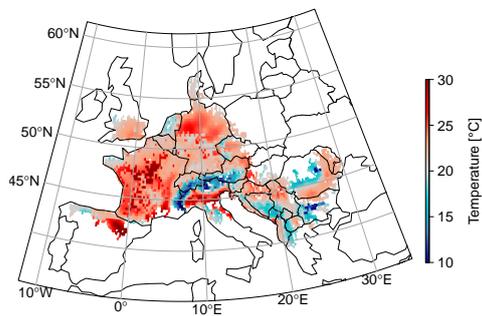
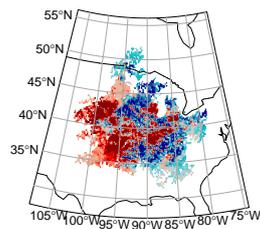
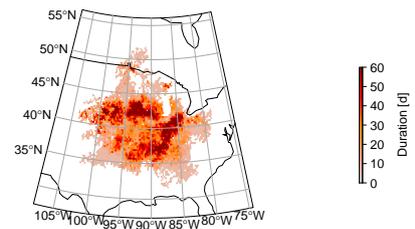


Figure 1: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the European heatwave 2003.

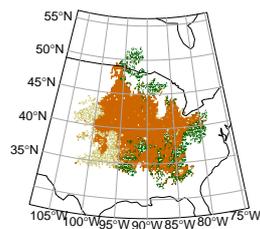
(a) GPP anomaly



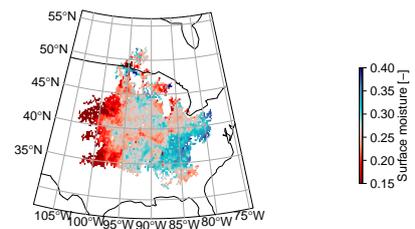
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

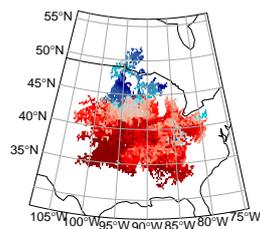
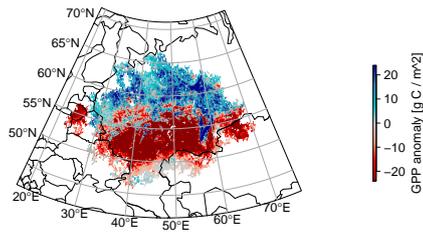
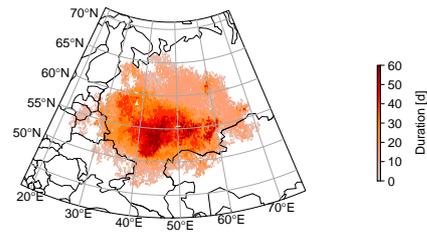


Figure 2: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the US 2012 drought and heatwave.

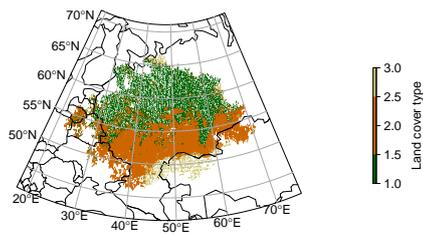
(a) GPP anomaly



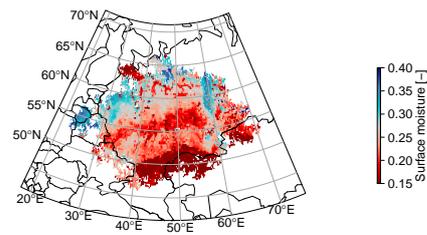
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

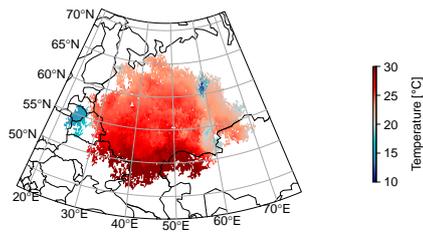
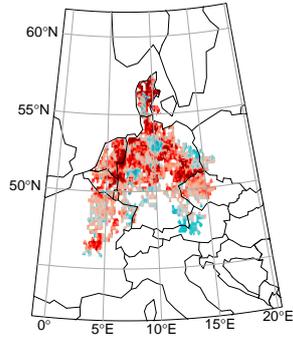
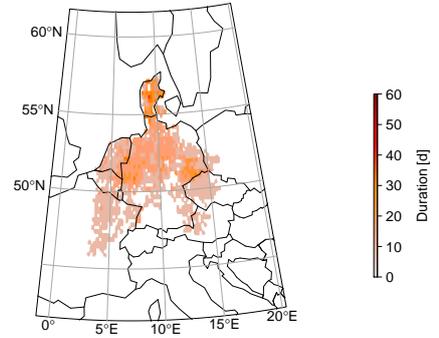


Figure 3: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the Russian heatwave 2010.

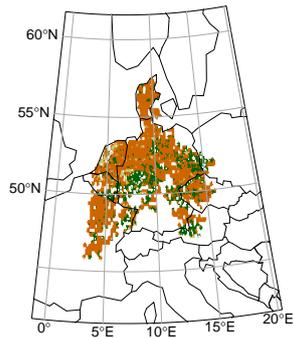
(a) GPP anomaly



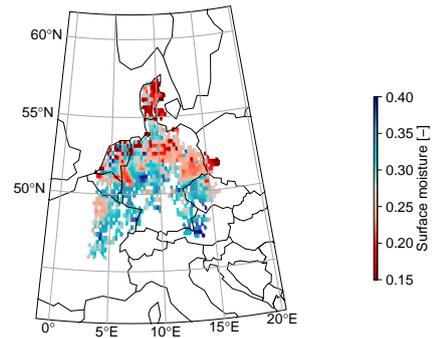
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

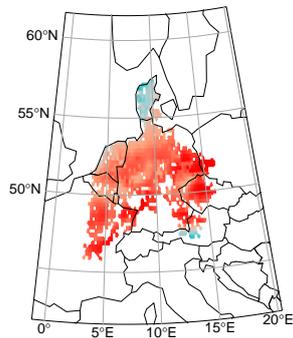
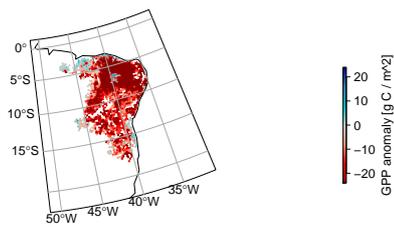
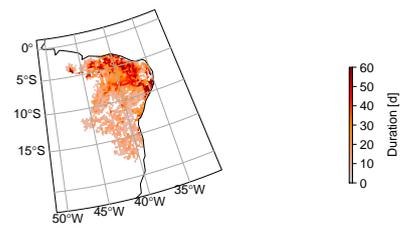


Figure 4: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the European heatwave 2018.

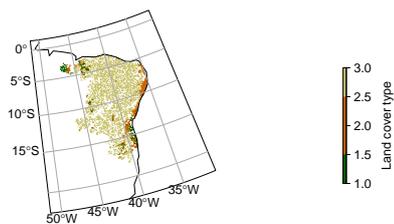
(a) GPP anomaly



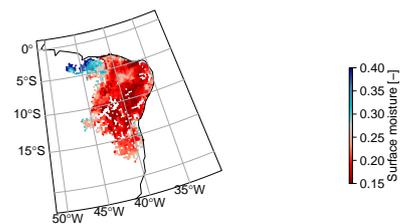
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

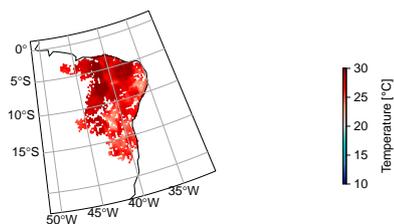
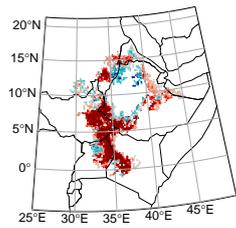
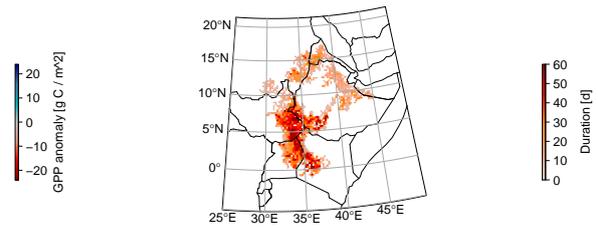


Figure 5: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the drought in Brazil 2012.

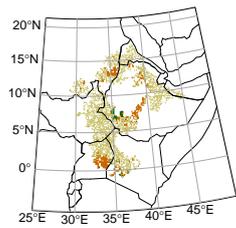
(a) GPP anomaly



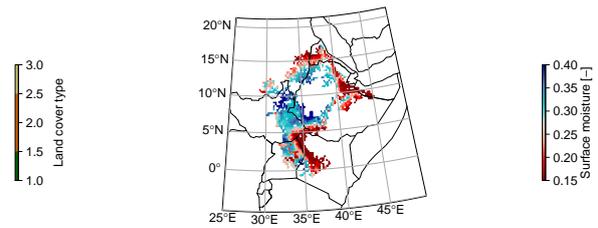
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

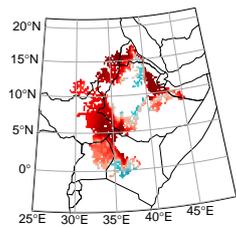
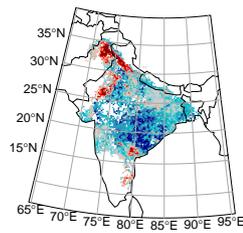
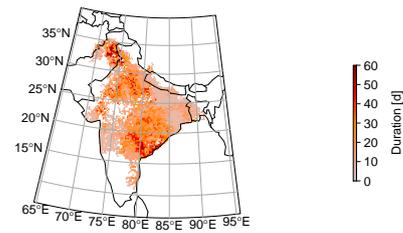


Figure 6: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the drought at the greater Horn of Africa 2009.

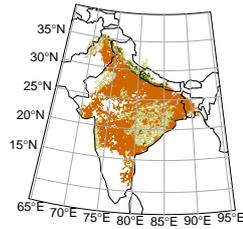
(a) GPP anomaly



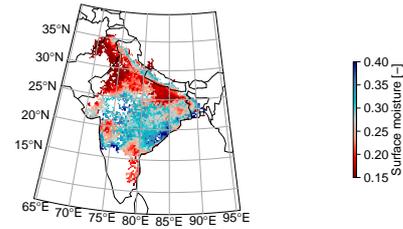
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

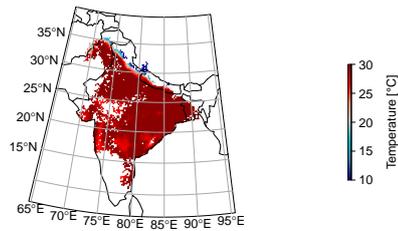
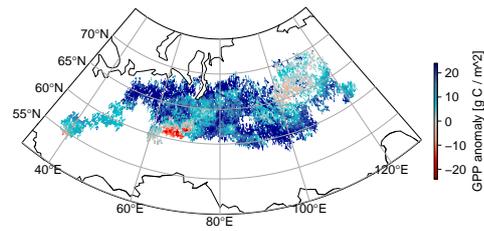
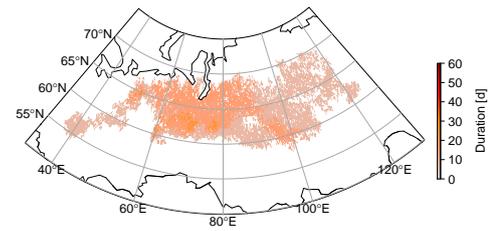


Figure 7: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the Indian heatwave 2009.

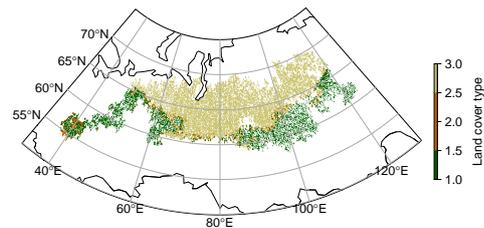
(a) GPP anomaly



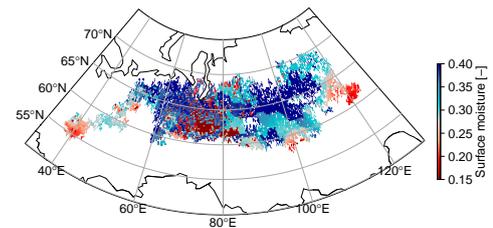
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

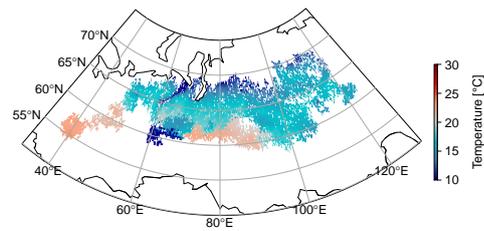
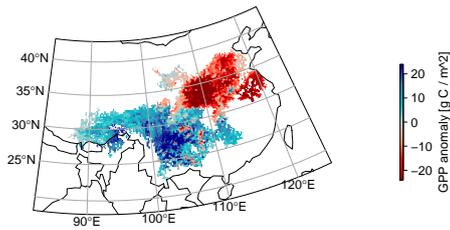
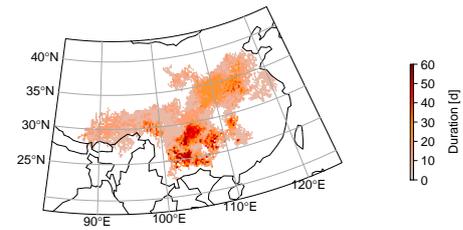


Figure 8: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the Siberian heatwave 2011.

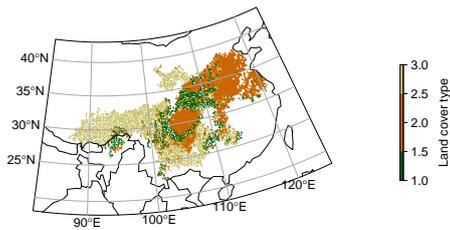
(a) GPP anomaly



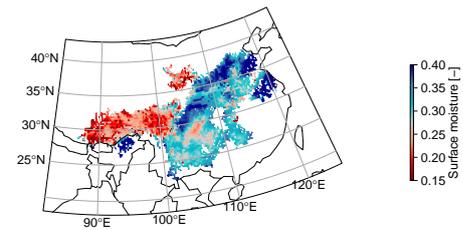
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

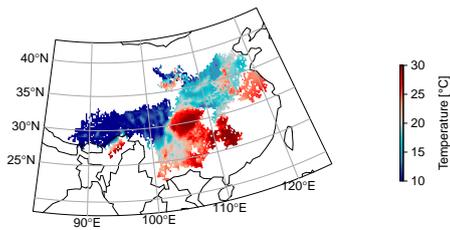
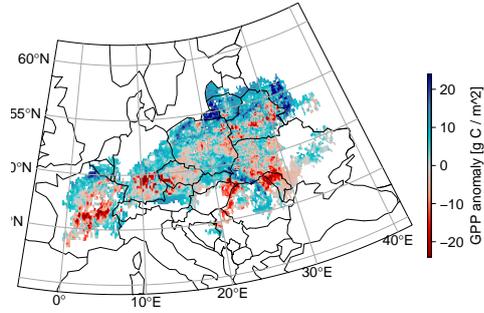
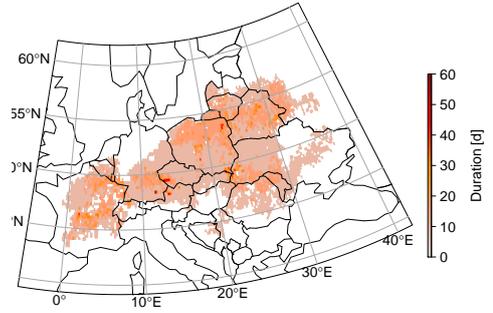


Figure 9: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the China drought 2011.

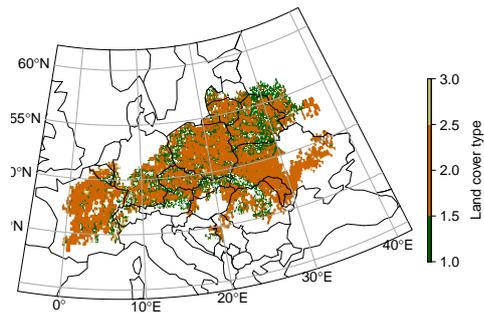
(a) GPP anomaly



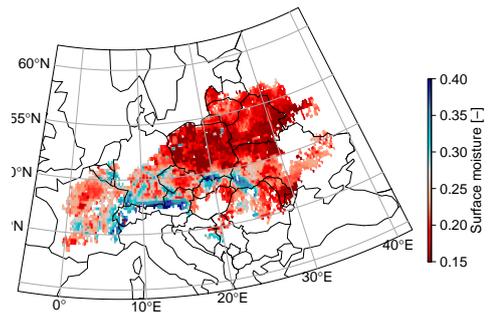
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

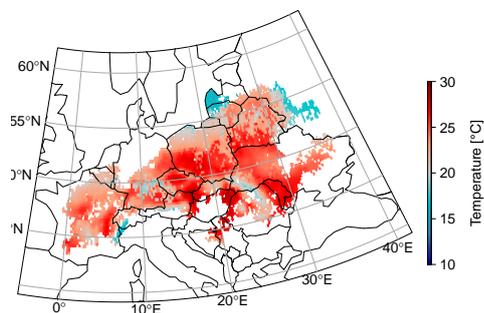
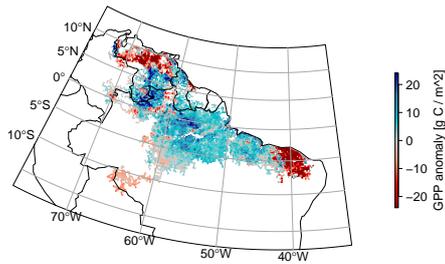
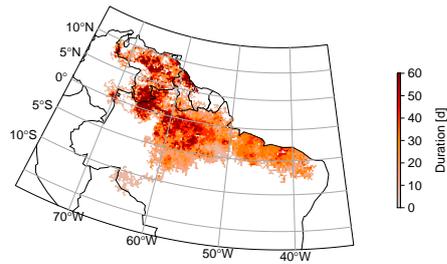


Figure 10: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the Drought and heatwave in East Europe 2015.

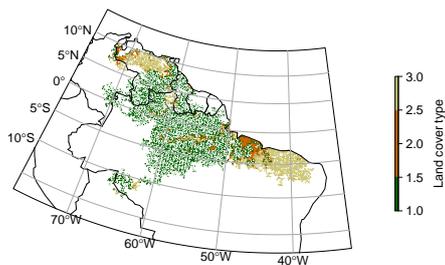
(a) GPP anomaly



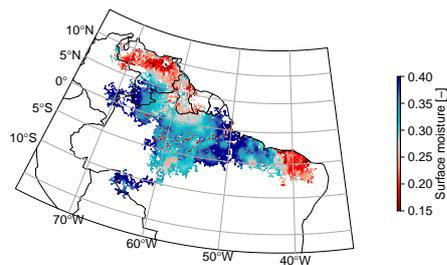
(b) Duration



(c) Ecosystems



(d) Surface moisture



(e) Temperature

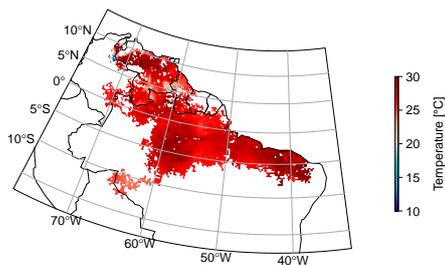


Figure 11: (a) Impact on GPP anomalies, (b) duration, (c) affected ecosystem types, (d) surface moisture conditions, and (e) temperature conditions associated with the Amazon drought 2010.