Edge Detection Reveals Abrupt and Extreme Climate Events

SEBASTIAN BATHIANY^a

Wageningen University and Research, Wageningen, Netherlands

JOHAN HIDDING

Netherlands eScience Center, Amsterdam, Netherlands

MARTEN SCHEFFER

Wageningen University and Research, Wageningen, Netherlands

(Manuscript received 18 June 2019, in final form 13 April 2020)

ABSTRACT

The most discernible and devastating impacts of climate change are caused by events with temporary extreme conditions (“extreme events”) or abrupt shifts to a new persistent climate state (“tipping points”). The rapidly growing amount of data from models and observations poses the challenge to reliably detect where, when, why, and how these events occur. This situation calls for data-mining approaches that can detect and diagnose events in an automatic and reproducible way. Here, we apply a new strategy to this task by generalizing the classical machine-vision problem of detecting edges in 2D images to many dimensions (including time). Our edge detector identifies abrupt or extreme climate events in spatiotemporal data, quantifies their abruptness (or extremeness), and provides diagnostics that help one to understand the causes of these shifts. We also publish a comprehensive toolset of code that is documented and free to use. We document the performance of the new edge detector by analyzing several datasets of observations and models. In particular, we apply it to all monthly 2D variables of the RCP8.5 scenario of the Coupled Model Intercomparison Project (CMIP5). More than half of all simulations show abrupt shifts of more than 4 standard deviations on a time scale of 10 years. These shifts are mostly related to the loss of sea ice and permafrost in the Arctic. Our results demonstrate that the edge detector is particularly useful to scan large datasets in an efficient way, for example multimodel or perturbed-physics ensembles. It can thus help to reveal hidden “climate surprises” and to assess the uncertainties of dangerous climate events.

1. Introduction

Like many other sciences, the Earth sciences are seeing an explosion in the amount of data from models and observations (Overpeck et al. 2011). This development calls

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-19-0449.s1>.

^a Current affiliation: Climate Service Center Germany (GERICS), Helmholtz-Zentrum Geesthacht, Hamburg, Germany.

Corresponding author: Sebastian Bathiany, sebastian.bathiany@hzg.de

for the creation of novel tools in order to automatically detect, track, and classify specific patterns in climate datasets (Ganguly et al. 2014; Faghmous and Kumar 2014). An important class of such patterns are extreme climate fluctuations that either persist (abrupt shift to another climate state) or decay after a short time (extreme events). While they are among the most devastating manifestations of climate change, it is often unclear how likely their occurrence will be for a certain level of global mean warming.

Abrupt climate shifts have often been described as “tipping points” (Lenton et al. 2008), comparable to a capsizing canoe, which irreversibly leaves its stable equilibrium once the canoeist leans too far to the side. The risk of triggering large-scale climate tipping points in the future, with devastating impact on society, is still a research topic involving much speculation. However, it is plausible that on a local scale, natural thresholds in the climate system like the melting point of ice or the wilting point of plants

DOI: 10.1175/JCLI-D-19-0449.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

can indeed cause abrupt change. Also, at the edge of deserts, monsoonal rain belts, or ice-covered areas, abrupt shifts can occur when climate zones shrink or expand. Consequently, abrupt shifts do occur in future climate scenarios: in the comprehensive climate models used for the recent report of the Intergovernmental Panel on Climate Change (IPCC), [Drijfhout et al. \(2015\)](#) identified 43 abrupt shifts on a decadal time scale that are larger than four standard deviations of the internal climate variability.

Extreme events like heat waves, heavy rainfall events, or droughts are typically treated as an entirely separate research topic because they are internally generated by the climate system, persist only up to a few weeks, and are not apparent in climate data averaged over longer time scales. Climate change is already affecting the frequency distribution of extreme weather events: an increasing number of extreme events have been detected in the recent past, and in many cases they have been attributed to human influence ([Coumou and Rahmstorf 2012](#); [Otto et al. 2018](#)).

Common methods for detecting abrupt or extreme events include machine learning approaches ([Monteleoni et al. 2013](#); [Liu et al. 2016](#); [Reichstein et al. 2019](#)), as well as predefined anomaly detection methods that identify extreme events as values that are far from most other data points ([Flach et al. 2017](#)). With regard to data with spatial dimensions only, like photographs or satellite data, machine vision approaches have been developed. The detection of edges is among the most fundamental steps in machine vision tools for identifying and recognizing objects or movement. A milestone in their development was the method of [Canny \(1986\)](#), which has remained a state-of-the-art approach until today. For example, related methods were used to infer the distribution of vegetation ([Kent et al. 2006](#); [Sun 2013](#)) clouds ([Dim and Takamura 2013](#)), sea ice ([Mortin et al. 2012](#)), and ocean vortices ([Dong et al. 2011](#)) from satellite images.

The problem of detecting abrupt change in time series has been approached with a large variety of methods ([Basseville and Nikiforov 1993](#); [Csörgö and Horváth 1997](#)). In climate science, changepoint detection algorithms have a long history in the analysis of observations with the aim of finding artificial step changes caused by changes in measurement devices ([Ducré-Robitaille et al. 2003](#); [Reeves et al. 2007](#)). Such changepoint detection has also been applied to detect real shifts in the observed climate ([Beaulieu et al. 2012](#)), in ecosystems ([Andersen et al. 2009](#)), the economy ([Silva and Teixeira 2008](#)), and many other systems. Another approach is to use the Bayesian paradigm to estimate the probability of abrupt shifts in climate records of hurricanes, heavy rainfall, and heat waves ([Chu and Zhao 2011](#)). With regard to abrupt shifts in simulated data, [Drijfhout et al. \(2015\)](#) investigated annually averaged output of selected variables

from current Earth system models. Their approach was not fully automatic, but relied on a visual inspection of geographical maps, and a selection of conspicuous regions to produce time series.

It has remained a challenge to detect events using their spatial as well as time components. A typical approach consists in identifying temporal anomalies with a peak-over-threshold approach first, followed by a flood-fill algorithm that checks the grid points' spatial connectivity in order to find the connected components ([Lloyd-Hughes 2012](#); [Zscheischler et al. 2013](#); [Prabhat et al. 2012](#)). However, the detection of events is only as good as the signal-to-noise ratio allows on a data-point level, and the point-by-point analysis is computationally expensive.

Although historically the motivations behind changepoint detection in time series and edge detection in images have been different, these approaches are technically closely related because both detect sharp step changes in data ([Radke et al. 2005](#)). In this article, we consequently combine both worlds (edge detection in images and changepoint detection in time series) by applying edge detection to space as well as time dimensions. We achieve this by generalizing [Canny's \(1986\)](#) edge detector to more than two dimensions (including space and time). This method has the advantage that spatiotemporal connectivity is taken into account for the detection itself, instead of being checked afterward, which allows one to separate the meaningful spatiotemporal features associated with tipping points or extremes from the noise on smaller scales. In contrast to the methods mentioned above, we hence provide a single tool that can detect features in space and time, and that can detect extreme as well as abrupt events. We complement the approach with a number of climate-specific tools such as space–time calibration, a quantification of abruptness, and a number of other diagnostics.

The article is structured as follows: In [section 2](#), we introduce [Canny's \(1986\)](#) original edge detector, and explain how we adjust it to the specific requirements of climate data. In [section 3a](#), we complement the edge detector with a new method to quantify an event's abruptness in time and provide a statistical assessment of this method in [section 3b](#). [Section 4](#) outlines our approach to calibrating the aspect ratio of space and time dimensions. In [section 5](#), we show several examples, analyzing spatial edges of sea ice and forests, and atmospheric rivers in the east Pacific ([section 5a](#)). [Appendices A](#) and [B](#) provide additional technical information for the analysis of vegetation (as an example of how we treat missing data) and sea ice. [Section 5b](#) shows an assessment of the detection of abrupt shifts based on an artificial test case. In [section 5c](#), we explain how the tool can be applied to identify extreme events, and we demonstrate its performance by detecting

FIG. 1. The major steps of the Canny edge detection algorithm. (top left) The original image, and (top right) the change after the first two steps: smoothing (seen in the color shading, which represents the data) and gradient calculation (represented by the arrows). (bottom left) Nonmaximum suppression maintains only the maxima of the gradients, and (bottom right) hysteresis thresholding keeps only the strongest connected edges.

several recent heat waves in global reanalysis data, with [appendix C](#) providing technical details. In [section 6](#), we finally apply the tool to a large number of models and variables from the CMIP5 archive ([Taylor et al. 2012](#)), and thus present the first automatic detection of abrupt shifts in Earth system models (technical details are in [appendix D](#)). In [section 7](#) we discuss the experiences we have made and provide our overall conclusions. The software of our new tool—including the full code, installation instructions, example notebooks, documentation, and the software used to generate all results shown in this article—is freely available. [Appendix E](#) describes the diagnostics that are included in our tool but not discussed at length in the article, and [appendix F](#) provides information on how the software and analysis scripts can be obtained, installed, and used.

2. Edge detection in space for climate data

In this section we briefly outline the four fundamental steps of [Canny's \(1986\)](#) original edge detector: smoothing,

differentiation, nonmaximum suppression, and hysteresis thresholding. [Figure 1](#) shows an example of these steps. For more explanations with regard to the choices of methods in this section, we refer the reader to [Canny \(1986\)](#). Here, we particularly focus on the additions we made to the algorithm in order to allow the analysis of typical climate datasets. All parameters we introduce along the way, the strategy for choosing their values, and the values chosen for analyzing data discussed in this article are summarized in [Table 1](#).

a. Smoothing

Taking derivatives amplifies noise present in the data, which would lead to the detection of features that are small-scale spurious events restricted to a few grid cells. To remove this noise, we first smooth the data on a chosen scale σ_i in a specific dimension i with a Gaussian kernel. This smoothing step is illustrated as the change in the colored field from the upper-left to the upper-right panel in [Fig. 1](#). While Canny's edge detector was designed for images with equidistant

TABLE 1. Parameters of the edge detector, strategies for selecting their values, and specific choices we made about options and values for each of the datasets we used (including the test case in section 5b). Index t stands for time; x stands for space. The order of the analysis steps follows the control flow of the algorithm in contrast to the sections in the main text. In the hysteresis thresholding, S^* refers to the combination of largest gradient in space and time [based on Eq. (3)].

Analysis step	Parameters	Strategies for decision on parameter values
Tapering	Number of grid cells n_i to be covered by the flat filter in each dimension i ; number of iterations for repeated tapering N	Trial and error: use aim that no edges remain over areas with missing values and at the boundaries between valid and missing data Sea ice observations: no tapering used LAI observations: $n_t = 0$, $n_x = 25$, and $N = 5 \times 10^3$ Heat waves (ERA5; temperature): no tapering used CMIP5 (ocean and land variables): $n_t = 0$, $n_x = 5$, and $N = 50$ Test case: no tapering used Atmospheric rivers (ERA5; water vapor): no tapering used
Smoothing	One smoothing scale σ_i per dimension i ; here, σ_t is for time and is σ_x for space	User chooses subjectively on what scale edges shall be detected Sea ice observations: $\sigma_t = 1$ yr and $\sigma_x = 120$ km LAI observations: $\sigma_t = 10$ yr and $\sigma_x = 200$ km Heat waves (ERA5; temperature): $\sigma_t = 5$ days and $\sigma_x = 50$ km CMIP5 simulations: $\sigma_t = 10$ yr and $\sigma_x = 100$ km Test case: $\sigma_t = 10$ years and $\sigma_x = 300$ km (if not specified otherwise in supplemental Table S2) Atmospheric rivers (ERA5; water vapor): $\sigma_t = 1$ h and $\sigma_x = 100$ km
Calibration	Aspect ratio between space and time γ	Option 1: pick weights oneself Sea ice observations: $\gamma = 10^{10} \text{ km yr}^{-1}$ LAI observations: $\gamma = 10^6 \text{ km yr}^{-1}$ Heat waves (ERA5; temperature): $\gamma = 10^{-5} \text{ km yr}^{-1}$ Option 2: γ tuned to make a certain percentile of space and time gradients in the dataset identical (not used in this article) Option 3: γ tuned to make a certain percentile of space and time gradients in a reference dataset identical CMIP5: 75th percentile Test case: maximum Atmospheric rivers (ERA5; water vapor): $\gamma = 10^{10} \text{ km h}^{-1}$
Hysteresis thresholding	Upper threshold h_u ; lower threshold h_l	Option 1: set thresholds based on knowledge about typical gradients to be detected Option 2: based on percentiles of the signal in the data itself Sea ice observations: 95th and 50th percentiles LAI observations: 75th and 50th percentiles Heat waves (ERA5; temperature): 99.99th and 95th percentiles Atmospheric rivers (ERA5; water vapor): 95th and 90th percentiles Option 3: Based on percentiles of the signal in a reference dataset CMIP5 simulations: S^* and $S^*/2$ Test case: S^* and $S^*/2$ (if not specified otherwise in supplemental Table S3)
Abruptness quantification	Transition chunk length c_{trans} , max chunk length c_{max} , and min chunk length c_{min}	Depends on what time scale is most relevant for the problem at hand Sea ice observations: not used LAI observations: not used Heat waves (ERA5; temperature): $c_{\text{trans}} = 30$ days, $c_{\text{max}} = 460$ days (5 summers), and $c_{\text{min}} = 184$ days (2 summers) CMIP5 simulations: $c_{\text{trans}} = 2$ yr, $c_{\text{max}} = 30$ yr, and $c_{\text{min}} = 15$ yr Test case: $c_{\text{trans}} = 2$ yr, $c_{\text{max}} = 30$ yr, and $c_{\text{min}} = 15$ yr Atmospheric rivers (ERA5; water vapor): not used

geometry (like photographs), we allow the smoothing scale to differ between the dimensions for two reasons:

- The dimensions can have different units (space vs time)
- Climate data are often presented in latitude–longitude coordinates. To correct for the spherical distortion, we scale the kernel width with $1/\cos(\varphi)$, where φ is the geographic latitude (Fig. S1 in the

online supplemental material). Moreover, we allow for specifying axes of periodicity like the geographical longitude or the time of the year.

b. Differentiation with the Sobel operator

We compute derivatives using the Sobel operator, creating a field of gradients from the data (i.e., computing

the arrows in Fig. 1, top right). As Canny (1986) shows, the Sobel operator is the mathematical operation that results in the largest signal-to-noise ratio for detecting step edges. In two dimensions, the Sobel operator can be described as a convolution of the data with the 3×3 matrix

$$\mathbf{S}_x = \frac{1}{8} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}. \quad (1)$$

The matrix $\mathbf{S}_y$ for computing the Sobel operator in the y direction is the transpose of $\mathbf{S}_x$. Each operator is moved along both dimensions of the 2D grid, resulting in a field of derivatives. As for the smoothing, we account for the geometric distortion of the Earth grid by dividing the output of the Sobel operator with $\cos(\varphi)$. The procedure works the same way for any number n of dimensions larger than 2; the matrix above then becomes an n -dimensional operator [see Topa and Schalkoff (1989) for a similar approach and an explanation on how to define directions in discrete gridded space].

c. Nonmaximum suppression

In this step, all pixels in the space–time grid are labeled as part of an edge where the derivative of the data in the direction of the local gradient has a maximum. The neighboring points (in the direction of the gradient) are labeled as having no edge.

The result of this procedure is a mask that indicates the position of edges in space and time. Hence, this step turns continuous images into a sketch-like representation that only retains the edges but no data in between (Fig. 1, lower left).

d. Hysteresis thresholding

To remove spurious edges and keep only the most prominent patterns, we apply hysteresis thresholding. All edges where the gradient magnitude is greater than an upper threshold h_u are labeled as strong edges. All edges with a gradient magnitude smaller than h_u but still larger than a lower threshold h_l are labeled as weak edges. All strong edges are kept as edges in the final result. Weak edges are selected only if they are connected to a strong edge. This consideration of connectivity suppresses spurious small-scale features (as can be seen in the two lower figures in Fig. 1) and is an important prerequisite to detect spatiotemporally coherent events. The choice of threshold values can be based on the empirical distribution of gradients in the data or on a reference dataset.

e. Missing data points

Many climate datasets have gaps, for example because of missing observations or because variables are only

defined either on land or over the ocean. We therefore introduce a simple tapering method, which is described in appendix A.

Overall, our approach introduces all properties of spatial edge detection to time series: It is retrospective (not sequential) but is able to detect multiple change-points (above the spatial scale selected for smoothing). It is also several orders of magnitude faster than change-point detection methods and thus much more useful for large amounts of data.

3. Quantifying abruptness

a. A piecewise linear approach

The appeal of edge detection is its ability to separate spatiotemporal features from the noise very efficiently. However, it does not quantify how sharp these edges are on smaller scales (i.e., how abrupt a transition would appear to the eye). In the following, we present a simple but effective way to quantify the abruptness of an event. This quantification is performed after the edges have been detected as described in the previous section. When detecting extreme (instead of abrupt) events (see section 5c), the abruptness can be interpreted as the product between the magnitude and the extent of an extreme event.

Our quantifier of abruptness stands in the tradition of creating local linear approximations to the data that are disrupted by changepoints (Zeileis et al. 2003; Zhao et al. 2016; Reeves et al. 2007; Beaulieu et al. 2012). As an example, Fig. 2a shows a temperature time series from the model MPI-ESM-LR at a grid point in the Pacific sector of the Arctic Ocean. The abrupt shift in the early twenty-second century is associated with abrupt sea ice loss on a relatively large scale (Bathiany et al. 2016), bringing the warmer open ocean water in direct contact with the air and also damping the temperature variability. The time resolution here is yearly because the time series shows the April monthly average for each year.

The steps to calculate abruptness are as follows:

- 1) We pick the moment in time with the largest time component of the Canny signal (vertical red dashed line, identified as an edge by Canny's edge detector).
- 2) We remove a few years of data on both sides of this transition. The number of years to pick is the first of three parameters of the method (parameter c_{trans} ; see Table 1). Here we take a value of two years. The data point on the edge itself is always removed.
- 3) To either side of this created data gap we pick a certain number of data points. The number of years

FIG. 2. Quantifying the abruptness of shifts: (a) Time series of April temperature at one grid point in MPI-ESM-LR. The black line shows the original data, the blue line is the smoothed data (in space and time), and the red vertical line indicates the year with the largest time gradient in the smoothed data. (b) Calculation of abruptness based on two data chunks (black and red lines) around the transition shown in (a). Straight lines show the linear regression lines for both separate chunks and their extrapolation up to the transition point (dotted part of the lines). The blue arrow indicates the difference between the intercepts; black and red arrows indicate standard deviations σ of both data chunks. Abruptness is calculated by using Eq. (2) and the three quantities given in (b).

chosen is the second parameter, $c_{\max}$. Our default choice is to pick 30 years to each side.

- 4) In case at least one of the data chunks on both sides is shorter than a threshold $c_{\min}$ (the third parameter; default: 15 years), we do not quantify this abrupt shift. In the example of Fig. 2 this is not the case, but it can happen if a shift is very close to the beginning or end of a time series. It is then impossible to tell if the shift is part of the natural variability or rather would persist if the time series were prolonged.
- 5) We apply a linear regression to both data chunks individually. The difference of the intercepts on the vertical axis (called the offset in Fig. 2b) is a measure of the amplitude of the abrupt shift.
- 6) We calculate the standard deviation for both data chunks (σ_1 and σ_2 in Fig. 2b). It is important to note

here that we do not remove the trends first (i.e., we compute the standard deviations for the original data, not the residuals from the regression). This choice leads to a reduced abruptness in the case of a large overall rate of change in the data but is not a very essential one because its effect is much smaller than the difference between the abrupt cases we identified.

- 7) We compute the average of the two standard deviations. In case of different lengths of the data chunks (see step 4), we do not weight the standard deviations with the available data points to compute this average because a large variability to only one side of the transition is enough to destroy the impression of an abrupt shift, regardless if this part of the time series is shorter.

- 8) We define the abruptness as the ratio between the offset and the mean of these standard deviations, in close analogy to the statistical concept of a signal-to-noise ratio:

$$\text{abruptness} = \text{offset} / [1/2 \times (\sigma_1 + \sigma_2)]. \quad (2)$$

In the example shown in Fig. 2, the result is 4.67.

b. Statistical uncertainty

As a final assessment of performance, we quantify the uncertainty range of the measured abruptness for several typical examples. To avoid making assumptions about the data-generating process (such as Gaussian distributions of the regression residuals, or absence of autocorrelation), we choose a Monte Carlo approach and generate a large number of realizations for each of our examples. In each example, a linear change in the mean (linear trend, representing climate change) is superimposed by a step change with a specific step size (the abrupt shift) from one point in time to the next. The abrupt transition is always known a priori because here we only assess the abruptness quantification, not the detection of the events already achieved with Canny's algorithm.

We create several test cases, with the following choices:

- Zero mean versus a linear trend in the time series with slope 0.05 (change in mean from 0 to 10 within the 200 time points).
- The abrupt shift occurs either in the middle of the time series, or at a given time step right after the start of the time series.
- The abrupt shift occurs from one time step to the next as a default, but it can be spread out linearly across a number of time steps τ on both sides of the center point.
- We add one of two types of background noise: 1) stochastic white noise and 2) temperature data from the control simulation of the complex climate model MPI-ESM-LR. Both are standardized to have 0 mean and a variance of 1.

To obtain the noise from the climate model, we cut the 1000-yr-long preindustrial control simulation of MPI-ESM-LR into five parts, choosing April only. As the model grid has 192×96 grid points, we obtain 92 160 realizations of noise in total. For the stochastic white noise produced with a random number generator, we use the same number of realizations. As climate variability differs between locations, the realizations from the climate model do not have an identical spectrum, but in their entirety represent a typical spectrum of monthly temperature variability on Earth.

We plot the distributions of measured abruptness for all cases, for various step sizes of the abrupt shifts from 0 to 10, and for two different choices for each parameter, $c_{\max}$ and

c_{trans} (Fig. S2 in the online supplemental material). In the cases with white noise, the typical width of the distributions (capturing $\sim 99\%$ of realizations) is about 4, while uncertainty increases somewhat with increasing abruptness. The overlap of distributions for step sizes 0 and 2 is around 10%–20% at most in all cases with immediate shifts ($\tau = 0$) and at least 30 data points on each side of the shift (Fig. S2), indicating that an abruptness of 2 or larger is typically statistically distinguishable from 0. The superposition of the noise and the abrupt shift by a linear trend does not affect the uncertainty of abruptness, but it does affect its mean value. In the absence of a trend, the expected abruptness is identical to the step size. In case of an additional linear trend, the abruptness decreases with the slope. This occurs because the slope is part of the standard deviation of the data chunks on both sides. Taking the general trend into account is a deliberate design of our measure of abruptness; abrupt shifts in systems with permanent rapid change are hence downgraded.

When considering the more realistic climate variability from the model, uncertainties do not change substantially. Although the temporal correlations in the data make it harder to constrain the linear regression, these correlations are small enough to not affect the results.

In our simple examples, the larger $c_{\max}$ is, the less uncertain the estimated abruptness becomes. However, in more realistic data, nonlinear trends may become relevant further away from the abrupt shift. $c_{\max}$ should be set to a time scale within which the linear assumption is still expected to hold—for climate projections, 30 years is a reasonable choice.

There are essentially two effects to be considered when choosing the cutoff length c_{trans} .

- The uncertainty slightly increases for larger c_{trans} because the available chunks of the time series are then further apart (although they still have the same length), which amplifies uncertainties in the estimated slope.
- In the climate shift from the complex model shown in Fig. 2, it is apparent that transitions do not happen immediately, but still need some time. This is reflected in our idealized experiments with a transition region, controlled by parameter τ . The cutoff length c_{trans} should not be smaller than that time scale. Otherwise, points belonging to the transition from one state to the next are considered in the linear regression, and the estimator of abruptness is negatively biased (rows 5 and 6 in Fig. S2).

In the situation that events occur close to the beginning or the end of the time series, uncertainties become very large because of the much smaller number of points constraining the slope (last two rows in Fig. S2). We

hence set a minimum chunk length of 15 years in the analysis of climate models presented in [section 6](#).

To avoid using other abrupt changes in the regressions that might be close to the shift under consideration, we exclude the data points of and beyond the neighboring abrupt shift. However, using a smoothing scale σ_t that is not much smaller than $c_{\max}$ already prevents cases with several edges occurring close to each other in time.

4. Calibration of aspect ratio between space and time

In contrast to Canny's original approach, our edge detector deals with data that are represented on a grid with spatial as well as time dimensions (e.g., longitude, latitude, and the year). Derivatives in time have different units than those in the spatial directions. It is therefore not as straightforward to define the magnitude of a gradient vector as for the original Canny detector that deals with spatial dimensions only. To compare space and time units, we introduce a parameter γ , which represents the aspect ratio between space and time, in units of kilometers per year. Given the space and time components of the derivatives obtained from the Sobel operator, S_x and S_t , we compute the total Sobel signal S (gradient in space–time) from

$$S = [(\gamma S_x)^2 + S_t^2]^{1/2}. \quad (3)$$

This step is applied after computing the Sobel derivatives in each dimension ([section 2b](#)) but before applying the nonmaximum suppression and hysteresis thresholding, because the latter steps require the computation of a total signal. The value of γ has to be chosen by the user. If only spatial edges shall be detected (the classical application of the Canny edge detector, like examples in [sections 3](#) and [4](#)), all weight is put on space (i.e., $\gamma \rightarrow \infty$). In contrast, if all that matters to the user are abrupt shifts in time, regardless of their spatial structure, all weight is put on time ($\gamma \rightarrow 0$). If one is interested in the spatiotemporal structure of the events, both parts need to be considered.

There are several possibilities to achieve a meaningful calibration of γ in practice. For example, one can construct the distributions of spatial and temporal gradients from the data and choose γ in such a way that a certain percentile of both distributions is identical. In this calibration, γ essentially tells us how many kilometers one needs to travel in order to see a change in a certain variable that is identical to the change during one year at a fixed location. Similarly, in case of a moving spatial edge (like the sea ice edge or the tree line moving north over time), γ tells us how fast the spatial edge would have to move across a fixed location in order to produce the same step change in one year. In the case of climate

change simulations, the calibration can be based on the gradients in the preindustrial control simulation, which represents the natural climate variability without human interference. We will present an example of this unsupervised calibration in [section 5b](#).

5. Applications and performance assessment

To demonstrate the ability of the above algorithms to detect meaningful features, we apply our tool to several datasets. First, we follow the original motivation of edge detectors to find sharp gradients in space only. Second, we create a test case with edges in space and time, and assess the ability of the edge detector to calibrate correctly using a control dataset, and to separate the edges from the noise. Third, we also document its ability to identify historical heat waves as extreme events in reanalysis data.

a. Detection of spatial edges in satellite data and reanalysis

In the analysis of spatial features, we analyze satellite observations of sea ice concentration and leaf-area index (LAI), and reanalysis data of atmospheric humidity. [Figure 3](#) provides an example for each of these cases. The technical details of the detection of edges in LAI and sea ice are provided in [appendices A](#) and [B](#), respectively. Although we focus on the spatial component of edges only, the time of occurrence is still provided by our tool, which can provide an impression of the spatial variability of an edge's position ([Fig. 3b](#)), or a continuous movement in time ([Fig. 3d](#)).

In the following, we focus on the detection of atmospheric rivers. Atmospheric rivers are long filaments of large water vapor fluxes ([Newell et al. 1992](#)), which play an important role for the water vapor transport in mid-latitudes ([Ralph et al. 2004](#)). At the west coast of North America, landfalling atmospheric rivers can contribute to heavy rainfall events ([Dettinger et al. 2011](#)). Several approaches to detect and distinguish them from other weather events have been proposed ([Racah et al. 2017](#); [Byna et al. 2011](#); [Liu et al. 2016](#)). [Ralph et al. \(2004\)](#) established a proxy for the detection of atmospheric rivers based on a threshold criterion for the value of the vertically integrated water vapor (>2 cm), and criteria for the minimal length (2000 km) and maximum width (1000 km), which were narrowed down by [Neiman et al. \(2008\)](#) to include only landfalling rivers between 32.5° and 52.5°N. An algorithm according to this definition is implemented in the detection toolkit TECA ([Prabhat et al. 2012](#)).

Using ERA5 reanalysis data from the Copernicus climate data store ([Copernicus Climate Change Service 2017](#)), we apply our edge detector to the problem of

FIG. 3. Spatial edges in climate data detected with Canny's approach: (a) Leaf area index from satellite observations (time mean 1981–2011), (b) fraction of times during this period for which a spatial edge is detected at a certain location, (c) satellite observations of sea ice cover for September 2015, (d) detected edges in sea ice for 1979–2015 (the color marks the latest year for which the edge is detected at this position), (e) vertically integrated water vapor (mm) in the west Pacific, showing an atmospheric river at 1200 UTC 26 Apr 1998, and (f) the result of applying the edge detector to the field shown in (e).

finding atmospheric rivers and compare the results with [Dettinger et al. \(2011, hereinafter D11\)](#), who extended Neiman's dataset to provide a “ground truth” for atmospheric rivers based on their definition. Like D11, our results are based on the daily vertically integrated water vapor field in the northwest Pacific Ocean in the years 1998–2007, with the condition of landfall between 32.5° and 52.5°N. The main differences of our approach to D11 are as follows. First, we apply a threshold criterion to the field of maximum spatial gradients, instead of the absolute value of water vapor (using the 95th and 90th percentile of gradients; see [Table 1](#)). Second, we do not apply specific criteria with regard to the shape of the detected rivers, but make a comparison based on different thresholds of the total length of detected edges connected to the coast (see [Table S1](#) in the online supplemental material). Because of gaps in the satellite record, and because the dates with the largest gradient do not always coincide with the dates of largest absolute water vapor, we allow a dating mismatch of 1 or 2 days.

Because of the different definitions, the comparison should not be regarded as an evaluation of one method against the other but rather as a proof of concept that shows the potential of edge detection. The most important thing is that the rivers' edges are not sharp enough along their entire length to be fully counted, and therefore we would miss many of D11's events if we applied a similar length criterion (2×2000 km, accounting for both sides of the river). An improved calibration with new criteria (e.g., considering the water vapor flux field, and/or making more

sophisticated choices of the threshold values) will probably improve the detection rate. Nonetheless, there is a clear co-occurrence of dates between D11 and our detected events, as can be seen in [Fig. S3](#) of the online supplemental material, which shows the length of our detected edges superimposed by the events listed in D11. For a minimum length of edges of 1000 km or less, around 90% of the dates reported in D11 are also identified by our approach ([Table S1](#)). It is also obvious that, in total, our algorithm detects more events than D11, which shows in the small percentages of detected dates that are also listed in D11. The fact that this fraction increases with increasing length of the detected features (despite the decrease in the total number of detected events) shows that our more general definition also captures edges that are smaller than D11.

There are also several large features that we detect but that are not reported in D11. The original water vapor field corroborates that these events are real, and are thus potential false negatives in D11, which demonstrates the potential of edge detection to find previously overlooked features. [Figure 3e](#) provides one example of such an event. Insofar, our results complement the findings of [Byna et al. \(2011\)](#), who discuss potential false positives in D11 data, and who detected 81% of the events reported in D11 based on the same criteria.

b. Performance assessment of detecting edges in space and time

In the following, we present an artificial test case with predefined edges in space and time. The test case demonstrates how the method is able to detect abrupt shifts in time

FIG. 4. Test-case scenario with artificial edges: (a) zonal mean data depending on year and latitude (edge 1 is a stationary spatial edge, edge 2 is a spatial edge moving north, edge 3 is a tipping point, and edge 4 indicates the spatial boundaries of tipped area), (b) time series at one grid point at the latitude indicated by the blue and red arrows in (a) along with time series from the reference dataset piControl at 18°S [location of red arrow in (a)], and (c) Abrupt shifts found by the edge detector and the year of their occurrence.

despite the presence of strong spatial edges (due to the calibration). To generate the data, we use a grid with 240×120 grid points in the longitudinal and meridional direction, and monthly time resolution spanning the years 2006–2100 (like in a CMIP5 climate change scenario). We generate artificial data with a zonally uniform deterministic component and added white noise with a constant mean of 0 and standard deviation of 1, but with spatiotemporally different realizations. We introduce several edges into the deterministic component of the dataset (Fig. 4a):

- Edge 1—a stationary spatial edge at approximately 65°S. The edge separates the region to the north with a mean of 0 and the region to the south with mean 100.
- Edge 2—a spatial edge separating the default region with zero mean and the region north of this edge with a mean of -3 . This edge is located at 36°N in 2006 and moves north linearly in time until it reaches 72°N in 2100.
- Edge 3—a tipping point in the Southern Hemisphere tropics: In 2053 (half of the time), the variable sud-

denly jumps from 0 to a mean of $+3$ at all grid points between -30°S and the equator.

- Edge 4—as a consequence of the tipping point (edge 3), a spatial edge is introduced between the tipped area and the surroundings after the collapse.

We add background noise in two different ways:

- 1) as stochastic white noise with noise level 1, and
- 2) as a realization of internal climate variability, using the preindustrial control simulation from the complex climate model MPI-ESM-LR.

We also create a reference dataset (similar to the piControl simulation in CMIP5) in which edges 1 and 2 are also present but edge 2 stays stationary at 35°N instead of shifting north over time.

We illustrate the calibration for the reference dataset of the test case in Fig. 5 (for the more realistic case of the complex climate model MPI-ESM-LR, see Fig. S4 of the online supplemental material), by showing the distribution of gradients in space and time before and after calibration. Because of the simplicity of the test case in

FIG. 5. Calibration of space-to-time aspect ratio for artificial test data. Each blue dot stands for one grid cell that is part of an edge. Its position on the vertical axis is the time gradient; its position on the horizontal axis is the spatial gradient. For each of these two types of dimensions, the red lines show the maximum absolute value on both sides of the scale. Shown are (a) spatial vs temporal gradients in the original data in physical units and (b) gradients in the recalibrated data (i.e., after converting space into time). The outer and inner dark curves in (b) mark the upper and lower hysteresis thresholds, respectively.

which the data scatter around a few discrete values, the gradients in the scatter diagram of gradients are oriented along a few lines (whose number is larger than the edges in the original data because the calibration is applied after smoothing the data). We choose the maximum of the gradient as the target for calibration; that is, after calibration the maxima are supposed to be the same in space and time (Fig. 5b). As the calibration in Fig. 5 is based on the “preindustrial” reference dataset, all points with a discernible spatial component result from edge 1 (which here occurs as several distinct lines because of the spatial smoothing and the limited spatial resolution of the dataset). Because of the calibration, edge 1 is rescaled in a way that it does not hide the abrupt shifts in time occurring farther north, although it is much sharper in space in terms of physical units. In contrast to the reference dataset, the “climate change” dataset involves more changes over time, but not larger spatial edges. As a consequence, edge 2 and edge 3 are detected by the edge detector because the total signal exceeds the upper threshold. The upper threshold of the hysteresis thresholding we chose here, S^* , is the gradient that would result from combining [as in Eq. (3)] the largest time gradient and the largest spatial gradient of the dataset. This gradient is typically somewhat larger than the largest gradient in the data because the maximum derivatives in space and time generally do not co-occur in space and time. In Fig. 5b S^* is represented by the outer black curve.

Using the test case, we can quantitatively assess the detection of abrupt shifts. For both noise types, we investigate cases with a standard deviation (noise level) of 0.1 and 1. We quantify the success of the detection by

computing true positive rates and false positive rates for all time points in two separate spatial regions:

- 1) everything north of 10°N (i.e., the area edge 2 is moving over, and outer regions without abrupt shifts) and
- 2) the tipping point region between 30°S and the equator.

As Fig. 4c illustrates for a stochastic white noise with noise level 1, $\sigma_t = 10$ years, and $\sigma_x = 300$ km, the edge detector is able to identify all grid cells with edges correctly while only detecting very few false positives. The false positives are associated with background noise and tend to occur at grid cells next to the true positives because of the condition of connectivity in the hysteresis thresholding step.

We apply a range of smoothing scales for time scales between 0 and 10 years, length scales between 0 and 1000 km, and a variety of upper and lower thresholds. The results of this assessment are shown in Tables S2 and S3 in the online supplemental material. The smoothing scales are crucial parameters for separating the spatiotemporal signal from the noise. As the noise’s correlation scale in space and time is typically shorter than the scale of the features to be detected, larger smoothing scales tend to yield higher detection rates. However, in case of too much spatial smoothing, the edges become too blurred to be identified at the right space and time coordinates but are still found. True positive rates tend to be higher in the tipping-point area than for the moving spatial edge. This is due to the more complicated structure of the edge, with different latitudes collapsing at different times. When using modeled climate variability instead of white noise, the precision of the tropical tipping point decreases. In these low latitudes, where there

FIG. 6. Relation between extreme and abrupt events using the Russian heat wave in 2010 as an example: (top) normalized temperature T_n during June–August from the ERA5 reanalysis in western Russia and (bottom) cumulative sum of T_n .

is a large ocean area, climate variability involves higher spatial and temporal correlations. This makes it more difficult to reduce the noise by smoothing. Optimal results are typically obtained with $\sigma_t = 10$ years and $\sigma_x = 300$ km. We point out that the low true positive rates of ~ 0.5 in Table S2 for modeled noise with noise level 1 are merely due to a misplacement error. Because of the low-frequency climate variability, the detected time point of the abrupt shift is spread out over a few years.

Because of the simplicity of the test case, the results hardly depend on the choice of the hysteresis thresholds (Table S3). As long as the upper threshold is between the maximum of the background noise and the maximum signal, true positives are always close to 100% (i.e., edge detector identifies all “true” pixels with edges). At the same time, the false positive rate is usually of the order of magnitude 10^{-5} – 10^{-3} ; that is, fewer than 0.1% of the detected pixels are not part of the actual edge in the noise-free ground-truth dataset.

c. Extreme event detection—Application to heat waves in reanalysis data

As outlined in the introduction, the time scales and mechanisms typically differ between abrupt and extreme

events. However, the problem of detecting them phenomenologically is closely related: While the differentiation of a time series turns an abrupt shift into an extreme event, integration results in turning extreme events into abrupt events (Fig. 6). For instance, the cumulative sum is a popular approach to the problem of detecting shifts in the mean value of time series (Basseville and Nikiforov 1993, 35–41). The larger the amplitude and the larger the duration of an extreme event are, the larger the abrupt shift in the integral becomes. In this regard, detecting abrupt events and detecting extreme events are two sides of the same coin.

To demonstrate that our new edge detector is suitable for revealing extreme events, we apply it to the surface air temperature from ERA5 reanalysis data (see appendix C for technical details). We aim to detect extreme events in summer temperature (i.e., heat waves), which are among the weather events that caused the largest impacts and that received the largest attention in the last decades. Figure 7a shows the abruptness for each grid cell where at least one edge (extreme event in the original data) occurred. At grid cells where several events occurred, Fig. 7a shows the

FIG. 7. Extreme temperature events in ERA5 for summer (June–August) between 2000 and 2018: (a) abruptness in the cumulative sum time series of the normalized temperature T_n and (b) the year in which the event with the largest abruptness (i.e., largest change in the temperature integral) occurs.

largest of the abruptness values. Figure 7b shows the year in which this largest event occurred. Time series for selected grid points are shown in Fig. S5 of the online supplemental material. A comparison of these results with the recent heat waves discussed in the literature confirms that our approach reliably identifies the major heat waves from the last 20 years. Over western Russia, the Russian heat wave of 2010 (Flach et al. 2018) appears as the green area in Fig. 7b. The European heat wave of 2003 (García-Herrera et al. 2010) and the 2004 heat wave in Alaska (Sulikowska et al. 2019) show up as blue areas. The event over the central and eastern Pacific results from the 2015/16 El Niño event.

6. Data mining for abrupt events in the CMIP5 model archive

While the previous examples were mainly driven by the motivation to demonstrate and test the performance of the edge detector, we now go one step further and present the first automatic data mining for abrupt shifts in monthly data of phase 5 of the Coupled Model Intercomparison Project (CMIP5;

Taylor et al. 2012). We picked all two-dimensional (not vertically resolved) variables that were available for the RCP8.5 scenario in CMIP5 (see Table 2). In total, our analysis comprises 78 variables from 37 climate change simulations (see appendix D for technical details). It can be understood as an advancement of the study by Drijfhout et al. (2015, hereinafter D15), which did not use automatic detection algorithms. In contrast to D15, we scan all individual months (instead of annual means) and all variables made available by the modeling groups (instead of a selection of key variables).

As a first check, we establish whether the abrupt shifts found by D15 in the same ensemble also appear in the results of our data-mining output. This is indeed the case, and our diagnostic figures confirm the spatiotemporal structure of the events found by D15—without any additional manual analysis steps. Events that do not reoccur in our analysis are typically linked to internal variability, like decadal switches between ice-covered and open ocean. Since these events are also present in the preindustrial period, the calibration and hysteresis thresholding, by design, prevents finding them in the future scenario.

TABLE 2. Scanned CMIP5 variables as defined by the Climate Model Output Rewriter (CMOR) protocol. For each realm, the columns list the variable name, the description of the variable, its unit, and the number of models for which output from the RCP8.5 as well as the piControl simulation was available, and how many models produced shifts above a certain abruptness. Italic font indicates that only annual means were scanned instead of monthly means.

CMOR variable name	Variable description	Unit	No. of models scanned	Models with edges found	Models with edges of abruptness > 2	Models with edges of abruptness > 4	Models with edges of abruptness > 10
Atmosphere realm							
cct	Air pressure at convective cloud top	Pa	24	22	13	4	2
ci	Fraction of time convection occurs	1	18	18	18	7	0
clivi	Ice water path	kg m^{-2}	33	20	16	0	0
clt	Total cloud fraction	%	34	23	19	5	0
clwvi	Condensed water path	kg m^{-2}	32	26	17	1	0
hfls	Surface upward latent heat flux	W m^{-2}	34	29	25	8	0
hfss	Surface upward sensible heat flux	W m^{-2}	34	27	21	13	0
hurs	Near-surface relative humidity	%	28	23	18	9	2
huss	Near-surface specific humidity	1	29	29	27	5	0
pr	Precipitation	$\text{kg m}^{-2} \text{ s}^{-1}$	34	31	28	0	0
prc	Convective precipitation	$\text{kg m}^{-2} \text{ s}^{-1}$	1	0	0	0	0
prsn	Snowfall flux	$\text{kg m}^{-2} \text{ s}^{-1}$	34	28	17	3	0
prw	Water vapor path	kg m^{-2}	34	34	33	0	0
ps	Surface air pressure	Pa	33	3	2	0	0
rlds	Surface downwelling longwave radiation	W m^{-2}	33	33	32	16	0
rldscs	Surface downwelling clear-sky longwave radiation	W m^{-2}	30	30	28	7	0
rlus	Surface upwelling longwave radiation	W m^{-2}	32	32	31	21	7
rlut	TOA outgoing longwave radiation	W m^{-2}	34	26	18	0	0
rlutcs	TOA outgoing clear-sky longwave radiation	W m^{-2}	34	29	26	16	1
rsds	Surface downwelling shortwave radiation	W m^{-2}	34	33	31	11	0
rsdscs	Surface downwelling clear-sky shortwave radiation	W m^{-2}	32	32	31	26	3
rsus	Surface upwelling shortwave radiation	W m^{-2}	34	34	33	11	0
rsuscs	Surface upwelling clear-sky shortwave radiation	W m^{-2}	31	31	30	25	17
rsut	TOA outgoing shortwave radiation	W m^{-2}	10	9	9	8	2
sci	Fraction of time shallow convection occurs	1	15	15	14	9	0
sfcWind	Near-surface wind speed	m s^{-1}	28	13	10	2	0
tas	Near-surface air temperature	K	35	34	33	17	2
tasmax	Daily max near-surface air temperature	K	34	32	31	15	2
tasmin	Daily minimum near-surface air temperature	K	33	33	32	17	2
tauu	Surface downward eastward wind stress	Pa	34	13	6	0	0
tauv	Surface downward northward wind stress	Pa	34	12	7	0	0
ts	Surface temperature	K	34	34	33	21	5
uas	Eastward near-surface wind	m s^{-1}	26	9	3	0	0
vas	Northward near-surface wind	m s^{-1}	27	2	1	0	0

TABLE 2. (Continued)

CMOR variable name	Variable description	Unit	No. of models scanned	Models with edges found	Models with edges of abruptness > 2	Models with edges of abruptness > 4	Models with edges of abruptness > 10
Ocean realm							
chl	Total chlorophyll mass concentration at surface	kg m^{-3}	13	12	12	7	1
dpco2	Delta pCO ₂	Pa	10	9	9	6	1
epc100	Downward flux of particle organic carbon	$\text{mol m}^{-2} \text{s}^{-1}$	13	11	11	6	1
fgco2	Surface downward CO ₂ flux	$\text{kg m}^{-2} \text{s}^{-1}$	19	18	17	11	1
frc	Downward carbon flux at ocean bottom	$\text{mol m}^{-2} \text{s}^{-1}$	5	5	5	2	1
intdic	Dissolved inorganic carbon content	kg m^{-2}	12	10	9	7	0
intpp	Primary organic carbon production by all types of phytoplankton	$\text{mol m}^{-2} \text{s}^{-1}$	14	12	12	8	1
m1otst	Ocean mixed layer thickness defined by sigma <i>T</i>	m	12	7	6	4	1
omlmax	Monthly maximum ocean mixed layer thickness defined by mixing scheme	m	18	7	7	7	2
pbo	Sea water pressure at sea floor	dbar	19	16	11	3	0
ph	Sea water pH at surface	1	13	11	9	1	1
sos	Sea surface salinity	psu	27	17	14	7	3
talk	Total alkalinity at surface	mol m^{-3}	14	8	7	4	2
tos	Sea surface temperature	K	32	19	11	9	0
zooc	Zooplankton carbon concentration at surface	mol m^{-3}	11	9	9	7	1
zoocmisc	Other zooplankton carbon concentration at surface from additional zooplankton component concentrations alone (e.g., micro- and meso-)	mgC m^{-3}	1	1	1	0	0
zos	Sea surface height above geoid	m	32	19	14	11	0
Ocean ice realm							
sic	Sea ice area fraction	%	38	34	31	30	22
sit	Sea ice thickness	m	38	25	23	12	5
snc	Surface snow area fraction	%	22	22	21	16	11
snd	Snow depth on sea ice	m	29	19	16	9	3
Land surface realm							
baresoilFrac	Bare soil fraction	%	2	0	0	0	0
cSoil	Carbon mass in soil pool	kg m^{-2}	24	22	20	16	5
cVeg	Carbon mass in vegetation	kg m^{-2}	22	19	19	17	5
fFire	Carbon mass flux into atmosphere due to CO ₂ emission from fire	$\text{kg m}^{-2} \text{s}^{-1}$	14	11	10	2	1
fVegLitter	Total carbon mass flux from vegetation to litter	$\text{kg m}^{-2} \text{s}^{-1}$	14	14	13	6	2
fVegSoil	Total carbon mass flux from vegetation directly to soil	$\text{kg m}^{-2} \text{s}^{-1}$	4	3	3	3	2
gpp	Carbon mass flux out of atmosphere due to gross primary production on land	$\text{kg m}^{-2} \text{s}^{-1}$	24	22	21	18	8
grassFrac	Natural grass fraction	%	2	2	2	2	1
lai	Leaf area index	1	27	25	21	16	9
mrro	Total runoff	$\text{kg m}^{-2} \text{s}^{-1}$	29	28	24	4	1
mrros	Surface runoff	$\text{kg m}^{-2} \text{s}^{-1}$	30	30	28	10	3
mrso	Total soil moisture content	kg m^{-2}	31	24	19	18	10

TABLE 2. (Continued)

CMOR variable name	Variable description	Unit	No. of models scanned	Models with edges found	Models with edges of abruptness > 2	Models with edges of abruptness > 4	Models with edges of abruptness > 10
mrsos	Moisture in upper portion of soil column	kg m^{-2}	27	20	16	12	5
nbp	Carbon mass flux out of atmosphere due to net biospheric production on land	$\text{kg m}^{-2} \text{s}^{-1}$	19	17	17	12	3
npp	Carbon mass flux out of atmosphere due to net primary production on land	$\text{kg m}^{-2} \text{s}^{-1}$	23	20	19	13	7
ra	Carbon mass flux into atmosphere due to autotrophic (plant) respiration on land	$\text{kg m}^{-2} \text{s}^{-1}$	23	21	20	16	8
rGrowth	Carbon mass flux into atmosphere due to growth autotrophic respiration on land	$\text{kg m}^{-2} \text{s}^{-1}$	12	12	12	10	5
rh	Carbon mass flux into atmosphere due to heterotrophic respiration on land	$\text{kg m}^{-2} \text{s}^{-1}$	23	23	21	14	4
rMaint	Carbon mass flux into atmosphere due to maintenance autotrophic respiration on land	$\text{kg m}^{-2} \text{s}^{-1}$	10	9	9	9	3
tran	Transpiration	$\text{kg m}^{-2} \text{s}^{-1}$	27	24	21	11	5
treeFrac	Tree cover fraction	%	14	12	9	9	8
Land ice realm							
snc	Snow area fraction	%	22	22	21	16	11
snm	Surface snowmelt	$\text{kg m}^{-2} \text{s}^{-1}$	21	21	17	10	3
snw	Surface snow amount	kg m^{-2}	30	26	23	16	12

D15 reported a 38% chance of finding at least one abrupt shift in any variable in a certain model for the RCP85 simulation (see their supporting information). Counting in a similar way (i.e., taking all gradient maxima regardless of their abruptness), our results would essentially yield a 100% chance. This discrepancy has three major reasons:

- 1) *Temporal resolution*—We analyzed all individual months of the year, whereas **D15** scanned annual means and only analyzed the subannual scale in case an abrupt shift was detected. Many abrupt shifts, however, appear only in specific months or seasons, which prevented detecting any event in annual mean data.
- 2) *Comprehensiveness*—We scanned all available from the CMIP5 archive, whereas **D15** focused on a few selected key variables. Moreover, our automatic approach is much more thorough than their visual inspection in the sense that no events that satisfy our criteria can be missed.
- 3) *Difference in methods*—Our minimum requirement to find an event is the exceedance of the spatiotemporal gradient above a typical gradient in the preindustrial

control simulation. Because of the additional change in forcing in the climate change simulations relative to the control simulations, this occurs frequently, especially for very slow variables. For fast variables (variables with a large variability on the scale of a few years), the climate change signal does not exceed the natural variability on the chosen (decadal) time scale. In contrast, **D15** formulated fairly strict criteria that the time series needed to pass to be counted (such as showing a gradient value of 4 times the preindustrial mean).

We therefore filter our results according to certain thresholds of abruptness, to successively retain the events that are most abrupt. **Figure 8** shows the frequency of all detected events with abruptness larger than 4 (for comparison, see the time series in **Fig. 2**, which has an abruptness of 4.67). To count a model for an abrupt shift at a specific grid cell, we require that any variable shows such a shift in any month (or the annual mean) regardless of which variable or month shows the abrupt shift. This procedure allows the best comparison to **D15**. We additionally analyzed what fraction of variables showed abrupt shifts, and repeated these analyses for specific seasons, but

FIG. 8. Fraction of models with abrupt events with abruptness larger than 4 in the RCP8.5 simulations of the CMIP5 ensemble, considering all variables and all months.

the spatial pattern is always similar (not shown). Table 2 lists the number of models with abrupt shifts for each of the 78 variables we scanned.

In agreement with D15, the Arctic Ocean is the major hotspot of abrupt shifts (Fig. 8), which is shown by essentially all models. The contribution of individual variables in Table 2 confirms that sea ice is the reason for the Arctic shifts, which in turn affects shortwave fluxes and temperature (because the warmer open water comes into contact with the atmosphere), and hence longwave fluxes. Interestingly, most abrupt shifts happen in winter (December–February) and spring (March–May). This fact confirms the hypothesis that winter ice will disappear more abruptly than summer ice, which was analyzed in a subset of the models (Bathiany et al. 2016).

On land, the Himalaya and the tundra region are most vulnerable, which is related to the loss of snow coverage and melting soils, again affecting shortwave fluxes. Obviously, the melting point of water is a highly important threshold causing abrupt transitions in climate, at least on a local to regional scale, over the ocean as well as over land. It has to be noted also that many models do not have a representation of frozen soil. Insofar, the occurrence of abrupt shifts over high-latitude land areas may even be underestimated. On the other hand, one may argue that the soil physics are still only very crudely represented even in the other models, and that results may change in improved models.

As also noted in D15, there are few abrupt shifts in the tropics, and model agreement is also very low. Note, however, that some of the vegetation changes in low latitudes, especially developments associated with Amazon dieback, are drastic relative to preindustrial variability but are too

slow to be picked up as abrupt shifts on the basis of our parameter choices.

7. Discussion and conclusions

From the results above, we conclude that edge detection is an effective method to detect and analyze the spatiotemporal structure of abrupt and extreme climate events. In particular, we succeeded to detect and quantify abrupt shifts in an automatic, reproducible, and objective way (given certain parameter choices). The unified treatment of space and time arguably requires a somewhat more abstract thinking than other tools. However, our approach provides one general framework to detect and diagnose robust nonlinear features in space and time.

As we have shown for the calibration and selection of thresholds, users can be guided in the adequate choice of the parameters based on statistics of the data at hand (or a reference dataset). One interesting extension would be making the hysteresis thresholds spatially explicit, for example, when focusing on the time dimension, by setting them to 3 and 2 standard deviations of the natural variability at each grid point. In our applications, it was sufficient to use global thresholds because the climate variability on the decadal time scale we focused on does not drastically differ between locations. However, in datasets with substantial heteroscedasticity (heterogeneous variability), spatially explicit thresholds could prevent that abrupt shifts in regions with small background variability are overshadowed by the noise in other regions. Moreover, our approach so far is univariate despite the fact that many variables are entangled

with each other due to direct physical interactions. Although it would not make sense to define another dimension that stretches over different variables, approaches to multivariate detection of extremes (Flach et al. 2017) might inspire preprocessing strategies that would ultimately support multivariate edge detection.

In our definition of abruptness, we followed the traditional strategy of aiming for a large signal-to-noise ratio. However, alternative approaches are possible. For example, abruptness could also be understood as the length (in space or time) of the transition relative to the length of the episodes (or spaces) of stasis before and/or after the shift. Whereas our current approach is tied to fixed parameters that determine the time scales of these regimes, such an approach would be even more independent of time scales, whether a transition takes picoseconds or centuries.

All these examples indicate that our framework is flexible and allows for many possible extensions and improvements in the future. Although we focused on two space dimensions and one time dimension, the tool can be applied to even more dimensions. It could then help to detect and characterize nonlinear features in three space dimensions of atmosphere or ocean models, like sudden stratospheric warmings (Chao 1985; Butler et al. 2015), jet streams (Kern et al. 2018a), water-mass boundaries, heat waves in the ocean (Hobday et al. 2016), or the dynamics of fronts in the atmosphere (Kern et al. 2018b).

If extended with a sophisticated algorithm that can classify the distribution of gradients (see appendix E with supplemental Figs. S7 and S8 for illustration), it may become possible to infer information about the mechanisms behind tipping points automatically. For example, the gradual retreat of the sea ice edge might be distinguishable from a large-scale synchronous change (tipping point). Similarly, a gradual spatial retreat of the monsoon could be distinguished from a delay in its onset, or from a large-scale “monsoon failure” from one year to the next. These classes of change are often associated with different physical mechanisms, for example the functioning and sensitivity of circulation patterns, intrinsic physical thresholds, or self-amplifying feedbacks.

The orientation of edges in space and time can therefore help to decide on hypotheses on the mechanism behind an abrupt shift. Since the problem at hand is comparable to recognizing objects on photographs, an application that machine learning algorithms have been developed for successfully, machine learning could be a promising approach to this problem. A challenge in the particular case of annual cycles is the fact that months and years are not independent dimensions but constitute one spiralling dimension. Including the time of year in addition to the year itself would need to take this

topological feature into account. A remaining caveat lies in the fact that it will also become more difficult in higher dimensions to visualize the results. Nonetheless, these challenges are not specific to the edge detection approach, which is a promising strategy to tackle these problems, and should be further developed in this direction.

A particular advantage of our tool is its speed. Relative to typical statistical tests that apply to time series (Zhao et al. 2016; Beaulieu et al. 2016) it is faster by several orders of magnitude. Scanning the monthly CMIP5 data took approximately 1 day per terabyte on a simple laptop, an amount that would be practically impossible to analyze using statistical tests, and which took approximately 40 h with the (subjective and hence nonreproducible) method used in Drijfhout et al. (2015). In the future, this speed could very easily be increased again by several orders of magnitude by applying graphical processing units (GPUs) because GPUs are particularly suited to machine-vision algorithms like edge detection. We emphasize that the larger the data at hand are, and the larger the number of datasets to be analyzed is, the more useful data-mining tools like edge detectors will become.

We believe that the largest potential of the method lies in the context of constraining uncertainties, a research aim that is particularly promising in the case of abrupt shifts. Despite their societal relevance, our knowledge about the risks of future abrupt climate shifts is far from robust. Several important aspects are highly uncertain: future greenhouse gas emissions (scenario uncertainty), the current climate state (initial condition uncertainty), the question whether and how to model specific processes (structural uncertainty), and what values one should choose for parameters appearing in the equations (parametric uncertainty). Such uncertainties can be explored using ensemble simulations. For example, by running many simulations with different combinations of parameter values a perturbed-physics ensemble can address how parameter uncertainty affects the occurrence of extreme events (Clark et al. 2006). This strategy can be particularly beneficial for studying abrupt events as well since abrupt shifts are associated with region-specific processes, whereas models are usually calibrated to produce a realistic global mean climate at the expense of regional realism (Mauritsen et al. 2012; McNeall et al. 2016). The currently available model configurations are therefore neither reliable nor sufficient to assess the risk of abrupt shifts (Drijfhout et al. 2015). It is hence very plausible that yet-undiscovered tipping points can occur in climate models.

Given the huge amount and complexity of data associated with ensemble simulations, data-mining tools are an indispensable tool for this analysis and similar research strategies. We hope that our approach can inspire more developments in this direction that would help to

constrain uncertainties about extreme and abrupt climate events and would ultimately provide a more reliable basis for decision-making.

Acknowledgments. The work resulting in this article was part of the path-finding project “Data-mining tools for abrupt climate change” funded by the Netherlands eScience Centre (NLESC). Author Bathiany acknowledges funding for his previous position at Wageningen University and Research (WUR) from the Netherlands Earth System Science Centre (NESSC). We are grateful for the comments of four anonymous referees who helped us to improve the paper. Bathiany also acknowledges J. Canny’s inspiring illustration of a Dalek (Canny 1986, Fig. 11 therein), and the musical achievements of Franz Schubert, which kept the authors going during the creation of the software.

APPENDIX A

Missing Data—Application to Observations of the Leaf Area Index

Here we outline our approach for dealing with missing data points, using a satellite dataset of leaf-area index (leaf area per grid cell area), shown in Figs. 3a and 3b. The dataset is a fusion of AVHRR and MODIS data. For technical details see Liu et al. (2012) and other references available at the site (<http://icdc.cen.uni-hamburg.de/1/daten/land/globallai.html>) from where we downloaded the data. The dataset ranges from 1981 to 2011, with half-monthly sampling until February 2000 and 8-daily sampling thereafter, and it has a spatial resolution of approx. 0.0727° . As for the sea ice data, we first remapped the data on a coarser grid, with a resolution of somewhat less than 0.5° , to obtain a file size that we can handle more easily.

Because the standard algorithms we used cannot handle missing values, we devised a specific routine to avoid having the boundaries of regions with missing data (the coasts in the LAI example) show up as artificial edges. It is not sufficient to just fill data gaps with a certain number, like 0 for instance, because strong edges would then be detected at the coasts that can hide more subtle edges over land or ocean.

Hence, we apply a tapering routine to the data before all other steps of the analysis, in case missing values exist. The tapering routine is an iterative procedure consisting of two steps. Given a smoothing scale σ_{taper} , we repeat the following steps N times:

- 1) We smooth the entire dataset using a flat filter of width σ_{taper} . This removes sharp gradients every-

where (land and ocean in the LAI example; see Fig. S6 in the online supplemental material).

- 2) We replace the unmasked pixels (land points in the LAI example) with the original values. This way, the ocean does not affect the land results, but edges over ocean disappear.

The filter requires three arguments. The first is the data. The second is one number per dimension (here time, latitude, and longitude), which states over what number of grid cells the filter extends in this dimension. The minimum value for each is 3 (otherwise it would not smooth anything in a symmetric way). The third argument is the number of times one repeats this filtering. Thereafter, the original data are used again for the defined grid cells; hence, the tapered data are only used for the missing grid cells. If one decreases the size of the flat filter, one should increase the number of steps to obtain satisfying results.

In the example of the LAI dataset, one can see that, by repeatedly smoothing the data, values for LAI on land diffuse into the oceans (Fig. S6). This mitigates effects of smoothing over regions where no value for the LAI is defined. For the best results one should choose a small σ_{taper} and a large N ; however, taking a larger σ_{taper} will lead to a faster convergence in large masked regions. For the LAI dataset, we picked a 25×25 pixel flat kernel for smoothing, repeating $N = 5,000$ times (see Table 1).

APPENDIX B

Application to Spatial Edges in Sea Ice Observations

We use the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF) team sea ice concentration climate data record (Laverne et al. 2019), obtained online (<http://icdc.cen.uni-hamburg.de/1/daten/cryosphere/seaiceconcentration-osisaf.html>). The dataset has a spatial resolution of $25 \text{ km} \times 25 \text{ km}$. As explained above, the edge detector requires a rectangular grid with latitude and longitude as coordinates. We therefore interpolate the data bilinearly on such a grid with a resolution of approx. 0.5° . The dataset covers the period from 1979 to 2015, typically with a satellite image every two days in the first nine years, and every day thereafter (apart from some data gaps). We compute monthly averages and pick only the data from September, the month that typically has the smallest extent of Arctic sea ice.

As for all other datasets, the parameter values used in the edge detection are given in Table 1. The threshold values for the two hysteresis thresholds are based on the percentiles of the distribution of (spatial) gradients in the smoothed dataset (using all years), using the 95th and 50th percentiles. The

result is not sensitive to this choice because the ice edge is a very prominent feature with large gradients as compared with areas with almost total ice cover or open ocean.

APPENDIX C

Detection of Heat Waves in Reanalysis Data

We obtained ERA5 reanalysis data from the Copernicus Climate Data Store ([Copernicus Climate Change Service 2017](#)). We selected June, July, and August for 2000–18. For every day in these time windows, we downloaded hourly data for 0000, 0600, 1200, and 1800 UTC and then averaged these four times of each day to obtain daily averages. In similarity to [Zscheischler et al. \(2013\)](#), we subtract the time mean value and then normalize the daily anomalies with the temporal standard deviation at each grid point. Last, we bring the results on a T42 grid (approximately 2.8° at the equator) by bilinear interpolation to keep the amount of data sufficiently small to be easily handled. We then integrate the data over time as shown in [Fig. 6](#) and apply the edge detector with the settings given in [Table 1](#). The spatial smoothing scale of 50 km is below the grid resolution and does not have an additional smoothing effect on purpose. The smoothing scale in time, σ_t , is 5 days. Spatial gradients are not considered for the quantification of gradients in this application (see [section 4](#) and [Table 1](#) for further explanations). We choose hysteresis thresholds of 99th and 95th percentile, respectively, to only detect the most extreme events.

APPENDIX D

Choices of CMIP5 Variables and Models

a. Variables

We analyzed all available variables with two spatial dimensions (surface variables and vertical integrals) from the monthly output in CMIP5, for all 12 months with one exception: Variables related to the composition and biogeographical distribution of land vegetation and land carbon pools react so slowly to changing atmospheric conditions that the annual cycle is negligible (often, these variables are only updated yearly in the models). We therefore only scanned annual means instead of monthly means for the variables cSoil, cVeg, treeFrac, grassFrac, and baresoilFrac (printed in italics in [Table 2](#)).

b. Models

We analyzed all models for which at least one of these variables was available for both the piControl and RCP8.5 simulation, namely the following 37 climate

models: ACCESS1.0, ACCESS1.3, BCC_CSM1.1, BCC_CSM1.1(m), BNU-ESM, CanESM2, CCSM4, CESM1(BGC), CESM1(CAM5), CESM1(CAM5.1-FV2), CMCC-CESM, CMCC-CM, CMCC-CMS, CNRM-CM5, CSIRO Mk3.6.0, EC-EARTH, FGOALS-g2, FIO-ESM, GFDL CM3, GFDL-ESM2G, GFDL-ESM2M, GISS-E2-H, GISS-E2-H-CC, GISS-E2-R, GISS-E2-R-CC, HadGEM2-CC, HadGEM2-ES, INM-CM4.0, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC5, MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, NorESM1-M, and NorESM1-ME.

c. Land–sea masks

For all land or ocean variables, we always applied the land–sea mask on the native grid of each model and variable before analyzing the data, including the tapering method explained in [appendix A](#). Many of the climate models' ocean grids have a complicated geometry, for example because their poles are shifted over land to avoid numerical problems. The edge detector currently cannot handle such curvilinear geometry. For these grids, we therefore remapped the data of each model onto the atmosphere/land grid of the same model using a bilinear interpolation before we apply the edge detector. As usual, the parameter choices are listed in [Table 1](#).

APPENDIX E

Additional Diagnostics

The standard output of our *hypercc* tool provides a number of additional figures that help to analyze the results. We illustrate these features in Figs. S7–S10 in the online supplemental material using the abrupt sea ice loss in MPI-ESM-LR as an example (as in [section 3a](#)). These features are the following:

- Diagrams of the gradient components in space and time, where the size and color represent abruptness as defined in [section 3a](#)—Figure S7a shows the results for the test case from [section 5b](#); Fig. S8a does so for MPI-ESM-LR. As the test case shows, one can ideally distinguish tipping-point behavior from edges with a substantial spatial component. In three additional, similar scatterplots, we color coded the year (Figs. S7b and S8b), latitude (Figs. S7c and S8c) and longitude (Figs. S7d and S8d) of the shifts to see when and where they occur.
- A geographical map of the excess time gradient at each grid cell that is part of at least one edge (Fig. S9a)—The excess time gradient is calculated as the maximum absolute value of the difference between the time component of gradients and its time mean value.

- A geographical map of the largest abruptness at each grid cell that is part of an edge (Fig. S9b).
- A geographical map showing the year of the transition that is associated with this abruptness at each grid point (Fig. S9c).
- A time series at the grid point with the largest abruptness (similar to Fig. 2a).
- A time series with the histograms of the total magnitude of spatiotemporal gradients in the preindustrial control simulation (Fig. S10a), and the climate change scenario (Fig. S10b)—The latter gives an indication how sharp the peak is and how large it is relative to the two thresholds based on the preindustrial control data.

APPENDIX F

Installation and Documentation of Developed Software

All our code is available online (<https://github.com/abrupt-climate>) and is free to download and reuse. It contains installation instructions and documented scripts for producing most of the figures in this article. The github project abrupt-climate has three repositories: *notebooks*, *hypercc*, and *hyper-canny*.

The repository *notebooks* contains a file *Installation_Instructions.txt* that explains how to install the edge detection software and a file *requirements.txt* that is used during the installation to make sure that required packages are installed. The repository *notebooks* also contains the following Jupyter notebooks programmed in Python3, which were used to create the figures in this article:

- *Analysis_CMIP5.ipynb* documents how the CMIP5 data was calibrated and analyzed.
- *ERA5_T2m_extremes.ipynb* was used to analyze ERA5 data and create the results shown in section 6.
- *Observations_LAI.ipynb* was used to analyze the LAI data (section 4).
- *Observations_seaice.ipynb* was used to analyze the sea ice observations (section 3).
- *Testcase_abrupt.ipynb* was used to analyze the test-case data (sections 7 and 8).
- *canny_example_circle.ipynb* was used to create Fig. 1.
- *tutorial.ipynb* is a short tutorial (which uses the file *mogreps.py*) that explains some basic technical features of the fundamental steps of the edge detection (section 2).
- *ERA5_atmospheric_rivers.ipynb* was used to detect atmospheric rivers as discussed in section 5a.

The repository *hypercc* was used to automatically scan the CMIP5 archive without any graphical user interface. It contains a folder *scripts* with the following UNIX shell scripts:

- *create_testcase_abruptshifts.ksh* was used to create the data for the test case in sections 7 and 8 (requires Matlab) and
- *scan_CMIP5.ksh* was used to scan the monthly CMIP5 data.

The subfolder *evaluation* contains all scripts used to assess our methods used in sections 3b, 5a, and 5b.

The repository *hyper-canny* is automatically installed and used when following the installation instructions in the repository *notebooks*. No manual actions are required in this repository.

The complete output of the CMIP5 scan that is presented in section 6, with combined diagnostic figures for each model and variable and *netcdf* files of abruptness and years of events, is available online (<https://figshare.com/s/61ed8ed3f9a919b3faf5>).

REFERENCES

- Andersen, T., J. Carstensen, E. Hernández-García, and C. M. Duarte, 2009: Ecological thresholds and regime shifts: Approaches to identification. *Trends Ecol. Evol.*, **24**, 49–57, <https://doi.org/10.1016/j.tree.2008.07.014>.
- Basseville, M., and I. V. Nikiforov, 1993: *Detection of Abrupt Changes: Theory and Application*, Prentice Hall, 528 pp.
- Bathiany, S., D. Notz, T. Mauritsen, V. Brovkin, and G. Raedel, 2016: On the potential for abrupt Arctic winter sea-ice loss. *J. Climate*, **29**, 2703–2719, <https://doi.org/10.1175/JCLI-D-15-0466.1>.
- Beaulieu, C., J. Chen, and J. L. Sarmiento, 2012: Change-point analysis as a tool to detect abrupt climate variations. *Philos. Trans. Roy. Soc.*, **370A**, 1228–1249, <https://doi.org/10.1098/RSTA.2011.0383>.
- , R. Killick, S. Taylor, and H. Hullait, 2016: Package *EnvCpt*—Detection of structural changes in climate and environment time series. <https://cran.r-project.org/web/packages/EnvCpt/EnvCpt.pdf>.
- Butler, A. H., D. J. Seidel, S. C. Hardiman, N. Butchart, T. Birner, and A. Match, 2015: Defining sudden stratospheric warmings. *Bull. Amer. Meteor. Soc.*, **96**, 1913–1928, <https://doi.org/10.1175/BAMS-D-13-00173.1>.
- Byna, S., Prabhat, M. F. Wehner, and K. J. Wu, 2011: Detecting atmospheric rivers in large climate datasets. *Proc. Second Int. Workshop on Petascale Data Analytics: Challenges and opportunities (PDAC'11)*, Seattle, WA, PDAC, 7–14.
- Canny, J., 1986: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-8**, 679–698, <https://doi.org/10.1109/TPAMI.1986.4767851>.
- Chao, W. C., 1985: Sudden stratospheric warmings as catastrophes. *J. Atmos. Sci.*, **42**, 1631–1646, [https://doi.org/10.1175/1520-0469\(1985\)042<1631:SSWAC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<1631:SSWAC>2.0.CO;2).
- Chu, P., and X. Zhao, 2011: Bayesian analysis for extreme climatic events: A review. *Atmos. Res.*, **102**, 243–262, <https://doi.org/10.1016/j.atmosres.2011.07.001>.
- Clark, R. T., S. J. Brown, and J. M. Murphy, 2006: Modeling Northern Hemisphere summer heat extreme changes and their uncertainties using a physics ensemble of climate sensitivity experiments. *J. Climate*, **19**, 4418–4435, <https://doi.org/10.1175/JCLI3877.1>.

- Copernicus Climate Change Service, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), accessed 14 May 2019, <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>.
- Coumou, D., and S. Rahmstorf, 2012: A decade of weather extremes. *Nat. Climate Change*, **2**, 491–496, <https://doi.org/10.1038/nclimate1452>.
- Csörgö, M., and L. Horváth, 1997: *Limit Theorems in Change-Point Analysis*. John Wiley and Sons, 414 pp.
- Dettinger, M. D., F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan, 2011: Atmospheric rivers, floods and the water resources of California. *Water*, **3**, 445–478, <https://doi.org/10.3390/w3020445>.
- Dim, J. R., and T. Takamura, 2013: Alternative approach for satellite cloud classification: Edge gradient application. *Adv. Meteor.*, **2013**, 584816, <https://doi.org/10.1155/2013/584816>.
- Dong, C., F. Nencioli, Y. Liu, and J. C. McWilliams, 2011: An automated approach to detect oceanic eddies from satellite remotely sensed sea surface temperature data. *Geosci. Remote Sens. Lett.*, **8**, 1055–1059, <https://doi.org/10.1109/LGRS.2011.2155029>.
- Drijfhout, S., and Coauthors, 2015: Catalogue of abrupt shifts in Intergovernmental Panel on Climate Change climate models. *Proc. Natl. Acad. Sci. USA*, **112**, E5777–E5786, <https://doi.org/10.1073/pnas.1511451112>.
- Ducré-Robitaille, J.-F., L. A. Vincent, and G. Boulet, 2003: Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.*, **23**, 1087–1101, <https://doi.org/10.1002/joc.924>.
- Faghmous, J. H., and V. Kumar, 2014: A big data guide to understanding climate change: The case for theory-guided data science. *Big Data*, **2**, 155–163, <https://doi.org/10.1089/big.2014.0026>.
- Flach, M., and Coauthors, 2017: Multivariate anomaly detection for Earth observations: A comparison of algorithms and feature extraction techniques. *Earth Syst. Dyn.*, **8**, 677–696, <https://doi.org/10.5194/esd-8-677-2017>.
- , S. Sippel, F. Gans, A. Bastos, A. Brenning, M. Reichstein, and M. D. Mahecha, 2018: Contrasting biosphere responses to hydrometeorological extremes: Revisiting the 2010 western Russian heatwave. *Biogeosciences*, **15**, 6067–6085, <https://doi.org/10.5194/bg-15-6067-2018>.
- Ganguly, A. R., and Coauthors, 2014: Toward enhanced understanding and projections of climate extremes using physics-guided data mining techniques. *Nonlinear Processes Geophys.*, **21**, 777–795, <https://doi.org/10.5194/npg-21-777-2014>.
- García-Herrera, R., J. Díaz, R. M. Trigo, J. Luterbacher, and E. M. Fischer, 2010: A review of the European summer heat wave of 2003. *Crit. Rev. Environ. Sci. Technol.*, **40**, 267–306, <https://doi.org/10.1080/10643380802238137>.
- Hobday, A. J., and Coauthors, 2016: A hierarchical approach to defining marine heatwaves. *Prog. Oceanogr.*, **141**, 227–238, <https://doi.org/10.1016/j.pocean.2015.12.014>.
- Kent, M., R. A. Moyeed, C. L. Reid, R. Pakeman, and R. Weaver, 2006: Geostatistics, spatial rate of change analysis and boundary detection in plant ecology and biogeography. *Prog. Phys. Geogr.*, **30**, 201–231, <https://doi.org/10.1191/030913306pp477ra>.
- Kern, M., T. Hewson, F. Sadlo, R. Westermann, and M. Rautenhaus, 2018a: Robust detection and visualization of jet-stream core lines in atmospheric flow. *IEEE Trans. Vis. Comput. Graph.*, **24**, 893–902, <https://doi.org/10.1109/TVCG.2017.2743989>.
- , —, A. Schaeffer, R. Westermann, and M. Rautenhaus, 2018b: Interactive 3D visual analysis of atmospheric fronts. *IEEE Trans. Vis. Comput. Graph.*, **25**, 1080–1090, <https://doi.org/10.1109/tvcg.2018.2864806>.
- Lavergne, T., and Coauthors, 2019: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea ice concentration climate data records. *Cryosphere*, **13**, 49–78, <https://doi.org/10.5194/tc-13-49-2019>.
- Lenton, T. M., H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, 2008: Tipping elements in the Earth's climate system. *Proc. Natl. Acad. Sci. USA*, **105**, 1786–1793, <https://doi.org/10.1073/pnas.0705414105>.
- Liu, Y., R. Liu, and J. M. Chen, 2012: Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data. *J. Geophys. Res.*, **117**, G04003, <https://doi.org/10.1029/2012JG002084>.
- , and Coauthors, 2016: Application of deep convolutional neural networks for detecting extreme weather in climate datasets. <http://arxiv.org/abs/1605.01156>.
- Lloyd-Hughes, B., 2012: A spatio-temporal structure-based approach to drought characterisation. *Int. J. Climatol.*, **32**, 406–418, <https://doi.org/10.1002/joc.2280>.
- Mauritsen, T., and Coauthors, 2012: Tuning the climate of a global model. *J. Adv. Model. Earth Sci.*, **4**, M00A01, <https://doi.org/10.1029/2012MS000154>.
- McNeill, D., J. Williams, B. Booth, R. Betts, P. Challenor, A. Wiltshire, and D. Sexton, 2016: The impact of structural error on parameter constraint in a climate model. *Earth Syst. Dyn.*, **7**, 917–935, <https://doi.org/10.5194/esd-7-917-2016>.
- Monteleoni, C., G. A. Schmidt, and S. McQuade, 2013: Climate informatics: Accelerating discovering in climate science with machine learning. *Comput. Sci. Eng.*, **15**, 32–40, <https://doi.org/10.1109/MCSE.2013.50>.
- Mortin, J., T. M. Schroder, A. W. Hansen, B. Holt, and K. C. McDonald, 2012: Mapping of seasonal freeze-thaw transitions across the pan-Arctic land and sea ice domains with satellite radar. *J. Geophys. Res.*, **117**, C08004, <https://doi.org/10.1029/2012JC008001>.
- Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger, 2008: Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the west coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeorol.*, **9**, 22–47, <https://doi.org/10.1175/2007JHM855.1>.
- Newell, R. E., N. E. Newell, Y. Zhu, and C. Scott, 1992: Tropospheric rivers?—A pilot study. *Geophys. Res. Lett.*, **19**, 2401–2404, <https://doi.org/10.1029/92GL02916>.
- Otto, F. E. L., S. Philip, S. Kew, S. Li, A. King, and H. Cullen, 2018: Attributing high-impact extreme events across timescales—A case study of four different types of events. *Climatic Change*, **149**, 399–412, <https://doi.org/10.1007/s10584-018-2258-3>.
- Overpeck, J. T., G. A. Meehl, S. Bony, and D. R. Easterling, 2011: Climate data challenges in the 21st century. *Science*, **331**, 700–702, <https://doi.org/10.1126/science.1197869>.
- Prabhat, O. Rübel, S. Byna, K. Wu, F. Li, M. Wehner, and W. Bethel, 2012: TECA: A parallel toolkit for extreme climate analysis. *Procedia Comput. Sci.*, **9**, 866–876, <https://doi.org/10.1016/j.procs.2012.04.093>.
- Racah, E., C. Beckham, T. Maharaj, S. E. Kahou, Prabhat, and C. Pal, 2017: Extreme weather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *NIPS '17: Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, NIPS, 3405–3416, <https://dl.acm.org/doi/pdf/10.5555/3294996.3295099>.
- Radke, R. J., S. Andra, O. Al-Kofahi, and B. Roysam, 2005: Image change detection algorithms: A systematic survey. *IEEE Trans. Image Process.*, **14**, 294–307, <https://doi.org/10.1109/TIP.2004.838698>.

- Ralph, F. M., P. J. Neiman, and G. A. Wick, 2004: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Mon. Wea. Rev.*, **132**, 1721–1745, [https://doi.org/10.1175/1520-0493\(2004\)132<1721:SACAOO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1721:SACAOO>2.0.CO;2).
- Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteor. Climatol.*, **46**, 900–915, <https://doi.org/10.1175/JAM2493.1>.
- Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204, <https://doi.org/10.1038/s41586-019-0912-1>.
- Silva, E. G., and A. A. C. Teixeira, 2008: Surveying structural change: Seminal contributions and a bibliometric account. *Struct. Change Econ. Dyn.*, **19**, 273–300, <https://doi.org/10.1016/j.strueco.2008.02.001>.
- Sulikowska, A., Walawender, J. P., and Walawender, E., 2019: Temperature extremes in Alaska: Temporal variability and circulation background. *Theor. Appl. Climatol.*, **136**, 955–970, <https://doi.org/10.1007/s00704-018-2528-z>.
- Sun, J., 2013: Exploring edge complexity in remote-sensing vegetation index imageries. *J. Land Use Sci.*, **9**, 165–177, <https://doi.org/10.1080/1747423X.2012.756071>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Topa, L. C., and R. J. Schalkoff, 1989: Edge detection and thinning in time-varying image sequences using spatio-temporal templates. *Pattern Recognit.*, **22**, 143–154, [https://doi.org/10.1016/0031-3203\(89\)90061-7](https://doi.org/10.1016/0031-3203(89)90061-7).
- Zeileis, A., C. Kleiber, W. Kraemer, and K. Hornik, 2003: Testing and dating of structural changes in practice. *Comput. Stat. Data Anal.*, **44**, 109–123, [https://doi.org/10.1016/S0167-9473\(03\)00030-6](https://doi.org/10.1016/S0167-9473(03)00030-6).
- Zhao, C., Y. Cui, X. Zhou, and Y. Wang, 2016: Evaluation of performance of different methods in detecting abrupt climate changes. *Discrete Dyn. Nat. Soc.*, **2016**, 5898697, <https://doi.org/10.1155/2016/5898697>.
- Zscheischler, J., M. D. Mahecha, S. Harmeling, and M. Reichstein, 2013: Detection and attribution of large spatiotemporal extreme events in Earth observation data. *Ecol. Inform.*, **15**, 66–73, <https://doi.org/10.1016/j.ecoinf.2013.03.004>.