THE ROYAL SOCIETY
PUBLISHING

Impacts of extreme summers on European ecosystems: a comparative analysis of 2003, 2010 and 2018

A. Bastos[1,†], Z. Fu[2], P. Ciais[2], P. Friedlingstein[3], S. Sitch[4], J. Pongratz[1,5], U. Weber[6], M. Reichstein[6], P. Anthoni[7], A. Arneth[7], V. Haverd[8], A. Jain[9], E. Joetzjer[10], J. Knauer[8], S. Lienert[11], T. Loughran[1], P. C. McGuire[12], W. Obermeier[1], R. S. Padrón[13], H. Shi[14], H. Tian[14], N. Viovy[2] and S. Zaehle[6]

[1]Department of Geography, Ludwig Maximilians Universität, Luisenstrasse 37, 80333 Munich, Germany
[2]Laboratoire des Sciences du Climat et de l'Environnement (LSCE), CEA-CNRS-UVSQ, UMR8212, 91191 Gif-sur-Yvette, France
[3]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK
[4]College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK
[5]Max Planck Institute for Meteorology, 20146 Hamburg, Germany
[6]Max Planck Institute for Biogeochemistry, 07745 Jena, Germany
[7]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research / Atmospheric Environmental Research, 82467 Garmisch-Partenkirchen, Germany
[8]CSIRO Oceans and Atmosphere, Canberra 2601, Australia
[9]Department of Atmospheric Sciences, University of Illinois, Urbana, IL 61801, USA
[10]Laboratoire Evolution et Diversite Biologique UMR 5174, CNRS Universite Paul Sabatier, Toulouse, France
[11]Climate and Environmental Physics, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern 3012, Switzerland
[12]Department of Meteorology, University of Reading, Earley Gate, Reading RG6 6BB, UK
[13]Department of Environmental Systems Science, Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland
[14]International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, 602 Duncan Drive, Auburn, AL 36849, USA

AB, 0000-0002-7368-7806; ZF, 0000-0001-7627-8824; PCM, 0000-0001-6592-4966; WO, 0000-0002-7094-8011; RSP, 0000-0002-7857-2549

In Europe, three widespread extreme summer drought and heat (DH) events have occurred in 2003, 2010 and 2018. These events were comparable in magnitude but varied in their geographical distribution and biomes affected. In this study, we perform a comparative analysis of the impact of the DH events on ecosystem CO_2 fluxes over Europe based on an ensemble of 11 dynamic global vegetation models (DGVMs), and the observation-based FLUXCOM product. We find that all DH events were associated with decreases in net ecosystem productivity (NEP), but the gross summer flux anomalies differ between DGVMs and FLUXCOM. At the annual scale, FLUXCOM and DGVMs indicate close to neutral or above-average land CO_2 uptake in DH2003 and DH2018, due to increased productivity in spring and reduced respiration in autumn and winter compensating for less photosynthetic uptake in summer. Most DGVMs estimate lower gross primary production (GPP) sensitivity to soil moisture during extreme summers than FLUXCOM. Finally, we show that the different impacts of the DH events at continental-scale GPP are in part related to differences in vegetation composition of the regions affected and to regional compensating or offsetting effects from climate anomalies beyond the DH centres.

This article is part of the theme issue 'Impacts of the 2018 severe drought and heatwave in Europe: from site to continental scale'.

1. Introduction

In the last two decades, Europe has been affected by extreme summer drought and heat (DH) events, each characterized by record-breaking climate anomalies

and large associated economic, social and environmental costs [1–3]. Both the 2003 and 2010 extreme summers in western Europe and western Russia, respectively, were caused by a combination of atmospheric circulation anomalies and land-atmosphere feedbacks [4–6], on top of anthropogenic climate change [7]. The summer in 2018 was exceptionally warm and dry in central and northern Europe, resulting in major losses in crop productivity [8] and in increased fire occurrence in high-latitude ecosystems [9].

The DH events in 2003, 2010 and 2018 (DH2003, DH2010 and DH2018, respectively) were similar in that all were characterized by preceding rainfall deficits and strong feedbacks between air temperature and soil-water anomalies [5,6]. Still, each event differed in: (i) their geographical distribution and biomes affected; (ii) the preceding climate conditions; and (iii) the resulting impacts on ecosystems. DH2018 affected higher latitude and less drought-prone regions compared to DH2003, which was centred in western Europe, including Mediterranean ecosystems [1]. DH2010 affected a very large area in western Russia including high-latitude forests and crop-lands and grasslands in temperate regions [3,10]. In DH2003 and DH2010, severe preceding rainfall deficits in late winter and spring contributed to soil desiccation and amplified the summer DH [3,6]. In 2003, increased evapotranspiration in spring also contributed to amplify drought conditions [11]. In 2018, the extreme warm and sunny conditions in spring and concurrent increase in evapotranspiration have contributed significantly to amplify the summer drought.

While warm and bright conditions are expected to stimulate photosynthesis in many temperate regions, especially when ecosystems are not water-limited [12], the associated increases in vapour-pressure deficit (VPD) and evapotranspiration rates impose progressively increased water-stress and lead to stomatal closure. Eventually plant productivity may collapse when DH are too intense and prolonged, or if high temperatures lead to leaf damage. Tree ring records document that warm summers (even extremely warm ones) in northern boreal Europe are associated with higher growth whereas in temperate and southern forests, the opposite is true [13].

For DH2003, observation-based and modelling studies agree on severe negative impacts on ecosystem productivity and net CO_2 uptake [1,14,15]. Schewe et al. [16] though have reported systematic underestimation of the negative impacts of DH2003 on ecosystem compared to satellite-based data. Both DH2010 and DH2018 were characterized by sharp north/south differences in ecosystem response to extreme summer heat and drying. In the case of DH2010, a strong decrease in productivity was found in observation-based data, with only a small northern area having enhanced summer productivity [10]. Flach et al. [10] prosed that differences in the observed impact of DH2010 on ecosystem productivity was related to differences in land cover and water-use efficiency over the growing season. Forests, predominant in the northern regions, use water more conservatively during drought [17] and have deeper roots, thereby being able to sustain increased photosynthesis for longer compared to crops and grasses, which predominate in the southern part of DH2010.

Given their similarity in extent and magnitude, but their differences in geography, biomes affected and preceding climate conditions, these three events constitute case studies that can improve our understanding of large-scale impacts of DH in the carbon cycle. In this study, we compare large-scale observations of ecosystem productivity over Europe from an eddy covariance (EC) measurement and remote-sensing observation-driven dataset and 11 Dynamic Global Vegetation Models (DGVMs), in order to address the following questions:

(a) Considering their different characteristics including different spatial patterns/extents, how did DH2003, DH2010 and DH2018 comparatively affect ecosystems' productivity at the European scale?
(b) How well do DGVMs simulate the carbon flux anomalies during DH2003, DH2010 and DH2018, compared to observation-based datasets?
(c) Can asymmetries in the observed responses to DH events be explained by different sensitivities to similar climate anomalies between different ecosystems?

2. Data

(a) Climate fields

In this study, climate fields are obtained from the ERA5 Reanalysis from the European Center for Medium-range Weather Forecast (ECMWF). We used 2 m surface temperature (T), total precipitation (P) and its partitioning into rain and snow fall (Rf, Sf, respectively), surface downward short-wave radiation (SWR), long-wave radiation (LWR) and sea-level pressure fields (SLP) at 0.25° spatial and hourly temporal resolution from January 1979 until December 2018. These variables were used as forcing for DGVM simulations and FLUXCOM runs. Additionally, volumetric soil-water content from levels 1–4 (0–289 cm) fields (referred to as soil moisture, SM) were used for drought assessment and DGVM evaluation. ERA5 has been shown to have improved performance over Europe relative to other datasets in precipitation [18], irradiance [19] and SM [20]. There is good agreement of the anomalies in SM over Europe reported by ERA5 and microwave satellite-based measurements from the SMOS sensor (shown in Bastos et al. [21]).

(b) FLUXCOM

FLUXCOM combines carbon and energy fluxes and meteorological measurements at site level from EC measurement sites with gridded spatial–temporal information from remote-sensing and meteorological datasets, using machine learning techniques (ML) to scale up these fluxes to global extents [22]. These estimates are purely data-driven and therefore independent to DGVM model output. In this study, we use the RS+METEO data based on the daily ERA5 meteorological fields for the period 1979–2018 described in Jung et al. [23]: a 6-member ensemble of gross primary production (GPP) and total ecosystem respiration (TER) from three ML methods (ANN, MARS, RF) together with two flux partitioning methods [24,25]. Net ecosystem production (NEP) is calculated for the three ML methods. The data are provided on a 0.5° regular latitude/longitude grid and daily time-steps. The FLUXCOM relies on a Water-Availability Index (WAI), which was calculated from the ERA5 fields at 0.25° latitude/longitude spatial resolution at daily time-steps.

The FLUXCOM dataset has been evaluated against site-level data [22] and independent estimates of global carbon fluxes [23]. FLUXCOM is known to show weak performance in estimating inter-annual variability in in situ carbon fluxes [22], and to systematically underestimate the magnitude of

variability [23]. However, Jung *et al.* [23] have shown a high correlation of NEP with observation-based estimates at a global scale. Moreover, Tramontana *et al.* [22] have shown that fluxes in boreal and temperate regions showed better agreement with observations. Given that this study focuses on large-scale anomaly patterns over continental Europe, we use FLUXCOM as an additional data-driven estimate of carbon fluxes. However, because of the known underestimation of the variability magnitude, we divide the NEP, GPP and TER fluxes by their respective variance.

(c) Dynamic global vegetation models

In addition to FLUXCOM, we used an ensemble of 11 DGVMs to assess carbon and water exchanges in response to drought: CABLE-POP [26], ISBA-CTRIP integrating ISBA-CTRIP [27], DLEM [28], ISAM [29], JSBACH [30], LPJ-GUESS [31], LPX-Bern [32], OCN [33], ORCHIDEE [34], ORCHIDEE-MICT [35] and SDGVM [36]. The models were forced with observed CO_2 concentration from Dlugokenky & Tans [37] and changing climate between 1979 and 2018 from ERA5 (see 'climate fields' section). A fixed land-cover map from 2010 was used as land-cover forcing, since we are interested in non-anthropogenic disturbances, and was obtained from the LUH2v2 h dataset [38]. For those models including nitrogen cycling (CABLE-POP, DLEM, JSBACH, LPJ-GUESS, LPX-Bern, OCN), nitrogen forcing from the NMIP project was used [28]. The model simulations were run for most models at 0.25° spatial and hourly temporal resolution for the European domain between 32–75°N and −11°–65°E, following a spin-up to equilibrate carbon pools.

In this study, we analyse GPP, TER, the net land-atmosphere carbon fluxes and SM from individual DGVMs. Since most models do not simulate fire emissions at a monthly time scale (except ISBA-CTRIP, JSBACH, ORCHIDEE-MICT), and the land-cover map is fixed, the net land-atmosphere carbon fluxes correspond to net ecosystem productivity (NEP) and are comparable to FLUXCOM. For DGVMs that ran at coarser spatial resolutions (ISBA-CTRIP, LPJ-GUESS, JSBACH, OCN), the simulated fields were resampled to the common 0.25° grid using an area conservative interpolation method. For each DGVM, daily and monthly GPP and SM were then extracted for EC site locations, in order to compare simulated values and their sensitivity to SM with site-level data (see below).

3. Methods

All datasets were first deseasonalized and detrended on a pixel basis by subtracting the mean seasonal cycle for the period 1979–2018 from daily/monthly fields and a long-term linear trend. Since DGVMs simulate soil layers of variable depths, the variance in simulated SM should differ between DGVMs, impacting the magnitude of SM_{anom}. Therefore, to reduce the influence of different soil schemes in SM_{anom}, standardized anomalies in SM from ERA5 and from DGVMs were used. The FLUXCOM dataset produced for this study shows too weak variance of anomalies in gross and net carbon fluxes, which is a known issue with this product [23]. The DGVMs also show considerable differences in the inter-annual variability amplitudes. Therefore, we analysed standardized anomalies in GPP, TER and NEP as in other studies based on this dataset [23]. As different datasets from FLUXCOM are available, they can provide a data-driven estimate of anomalies in carbon fluxes and their spread to compare with those simulated by DGVMs. Because of the standardization, the anomalies in GPP and TER do not necessarily add up to those in NEP.

We analyse for each individual model (i) the continental-scale spatial relationship between summer GPP and (SM,T) anomaly pairs individually for each DH event and for other years in the 2000s and (ii) the temporal sensitivity of daily fluxes to water-stress during extreme summers and for the remaining years in the 2000s. The spatial sensitivities to T and SM were estimated by an Ordinary Least-Squares fit of a multiple linear regression on each individual dataset for all the pixels in the continental domain, i.e. GPP = f(T,SM), where GPP, T and SM are arrays of size $1 \times n$, and n is the number of pixels in the European domain. Even though the multiple linear regression controls for correlation between variables, the strong correlation between T and SM adds uncertainty to the sensitivity estimates. The SM data used to estimate DGVM sensitivities were the simulated SM fields, while for FLUXCOM, the WAI [23] was used. We use the six runs from FLUXCOM as observation-based reference.

We then analysed GPP_{anom} in each DH event for pixels within different drought and land-cover classes, corresponding to SM_{anom} ranging between -3σ and 3σ, grouped into 0.5σ bins, and of forest or crop cover ranging from 0 to 100%, grouped into 10% bins The number of pixels within each group was also calculated to evaluate the incidence of each DH event in distinct land-cover types. Because FLUXCOM does not report GPP separately for different land-cover types, and since DGVMs also differ in the number of plant functional types represented, we assigned forest and crop cover classes to pixels using as reference the land-cover map used as DGVM forcing, i.e. LUH2v2 h map for the year 2010, which has been calibrated using the European Space Agency's CCI Land Cover product [39].

4. Results

(a) Impacts of the drought and heat events at continental scale

Anomalies in climate fields from ERA5 indicate strong warm and dry conditions along with increased solar radiation in western Europe in 2003 and in western Russia in 2010 (electronic supplementary material, figures S1 and S2). In 2018, the DH anomaly affected mainly central Europe and Scandinavia. While in 2003 and 2010, large sub-regions outside the centre of DH registered wetter and cooler summer conditions, in 2018 drought was more widespread, except for the Mediterranean region, and warmer than average conditions were registered in the whole continent.

We compare the summer and annual standardized NEP_{anom} (figure 1, top and bottom panels, respectively) from DGVMs with observation-based estimates from an ensemble of runs of the FLUXCOM data-driven product (filled box-plots). The corresponding absolute anomalies are shown in electronic supplementary material, figure S3. In summer, FLUXCOM and DGVMs estimate weakly negative NEP_{anom} in DH2003 and strongly negative NEP_{anom} in DH2010, both DH events associated with positive TER_{anom}, but differing on the sign of GPP_{anom}. In DH2018, FLUXCOM indicates negative NEP_{anom} linked with negative summer GPP_{anom} and neutral TER_{anom}, while DGVMs tend to estimate positive NEP_{anom}, mainly due to positive GPP_{anom}.

At the annual scale, DGVMs and FLUXCOM estimate negative NEP_{anom} in DH2010, mainly attributed to negative GPP_{anom}, and close to neutral anomalies in DH2003 and DH2018. DGVMs and FLUXCOM show, however, disagreements in the sign of gross and net fluxes in DH2003 and DH2018. The interquartile range of DGVM estimates is,
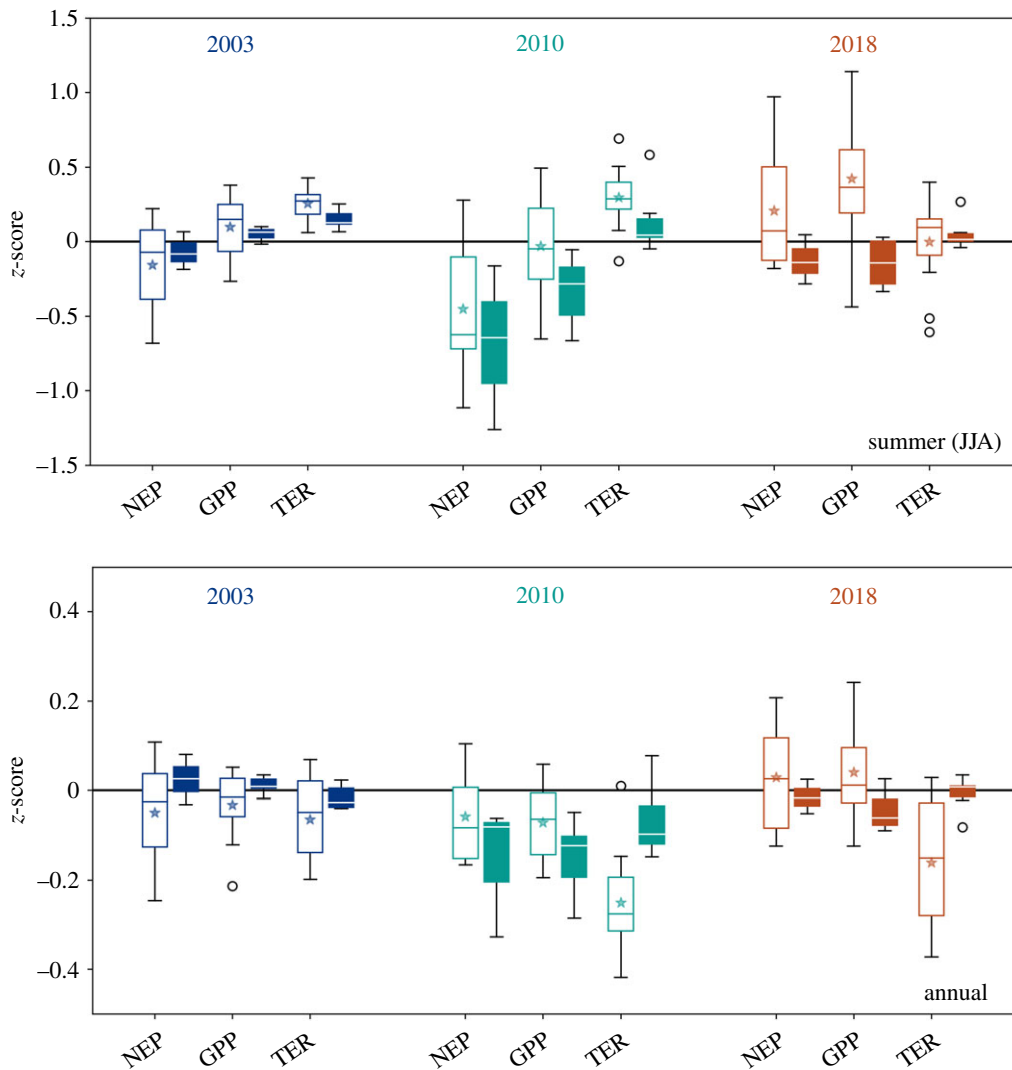
Figure 1. Summer and annual carbon flux anomalies. NEP_{anom}, GPP_{anom} and TER_{anom} in summers of 2003, 2010 and 2018 (blue, dark cyan and orange, respectively) estimated by the 11 DGVMs (white-filled boxplots) and from the data-driven FLUXCOM 6-member ensemble (colour-filled boxplots). Anomalies are reported in standardized units (divided by the standard deviation, z-scores). The stars indicate the MMEM and the circles show the outliers. (Online version in colour.)

though, often much higher than the estimated anomalies both in summer and annually, especially in 2003 and 2018. Therefore, the sign of anomalies cannot be robustly constrained by DGVMs.

(b) Spatial distribution of summer water and carbon anomalies

The multi-model ensemble mean (MMEM) simulates well the large-scale patterns and intensity of standardized summer SM_{anom} for all events, with high spatial correlation with ERA5 anomalies $R = [0.88, 0.90]$ for the three summers (electronic supplementary material, figure S2). The individual DGVMs tend to agree on the signs of anomalies, but those models simulating shallower soil layers have smaller absolute and standardized anomalies (maximum depth of soil-water balance of DGVMs from 0.5 m to greater than 6 m). The spread in DGVM estimates is mostly smaller than $\pm 1\sigma$, compared to SM_{anom} in the centres of the DH events below -1.5σ. As DGVMs, FLUXCOM was forced by ERA5, but a WAI was estimated independently as a predictor used to upscale flux tower data to the continent, WAI being calculated with a two-layer bucket model assuming a water holding capacity of 15 and 100 mm in upper and lower layer, respectively [22]. The spatial patterns of WAI also show

good, but slightly lower, agreement with ERA5 ($R = [0.82, 0.88]$, electronic supplementary material, figure S2).

We compare the spatial patterns of summer GPP (electronic supplementary material, figure S4) and NEP (electronic supplementary material, figure S5) from FLUXCOM and DGVMs. The MMEM and the mean of the six FLUXCOM datasets indicate negative GPP_{anom} in the central regions of DH events with the exception of some regions in Scandinavia during DH2018, which show above-average GPP. Despite the good spatial match of GPP_{anom} between the MMEM and FLUXCOM (spatial correlation $R = [0.71, 0.83]$), several DGVMs estimate opposite sign to FLUXCOM, especially in the transitional areas between DH centres and the surroundings in 2010 and 2018 (electronic supplementary material, figure S4, bottom panel). The spatial patterns of summer NEP_{anom} from the MMEM (electronic supplementary material, figure S5) show good agreement with FLUXCOM for DH2003 and DH2010 ($R = 0.81$ and 0.82, respectively), but poor agreement in DH2018 ($R = 0.51$), especially in Scandinavia. More models agree on the sign of the NEP_{anom} with FLUXCOM than of GPP_{anom} (more regions with greater than five models agreeing on the sign of anomalies), except in DH2018. In Scandinavia in DH2018, a majority of models estimate opposite sign of NEP_{anom} (positive) than that of FLUXCOM (negative).
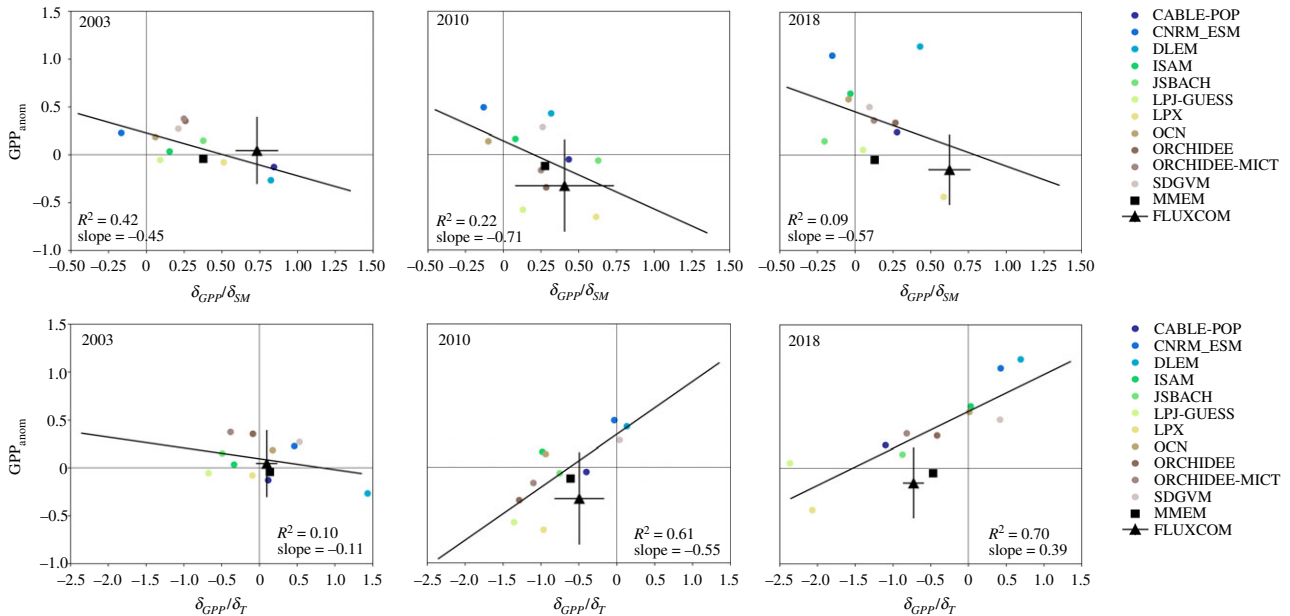
Figure 2. Simulated GPP$_{anom}$ by DGVMs and their sensitivities to SM (top) and temperature (T, bottom) compared to those from the data-driven product FLUXCOM. The markers show the values for individual DGVMs (coloured circles) and the MMEM (black square). The black line indicates the linear regression fit for the individual DGVM estimates, and the labels indicate the corresponding slope and R^2-values. The mean value of the 6-member ensemble from FLUXCOM is indicated by the triangle, and the ±1σ range by the error bars. (Online version in colour.)

(c) Sensitivity of gross primary productivity to drought and heat

The spatial sensitivity of summer GPP to SM and temperature (T) at continental scale was calculated using a multiple linear regression model fit on individual FLUXCOM runs and DGVM estimates and is shown in figure 2 (xx-axes) for the three DH events (see Methods), and compared to non-extreme summers in the 2000s (electronic supplementary material, figure S6). Simulated SM was used for DGVMs, while the analysis of FLUXCOM was based on the respective WAI. Sensitivities were calculated by performing a multiple linear regression using SM and T as predictors. A positive sensitivity (δGPP/δSM, δGPP/δT) means that GPP decreases when SM or temperature decrease.

All FLUXCOM runs estimate positive values of δGPP/δSM, yet stronger during DH2003 and DH2018 than during other summers since the year 2000 (compare slopes of regression lines in electronic supplementary material, figure S6). The DGVMs show weaker sensitivities of GPP to SM than FLUXCOM during non-extreme summers and a moderate increase during DH2003 and DH2010, with a decrease in DH2018. Furthermore, individual DGVMs show very different sensitivities to SM, with some models estimating negative δGPP/δSM values (i.e. higher GPP during drought conditions) even during DH events.

A positive sensitivity of GPP to T indicates a stimulation of productivity in warmer conditions, which is the case for non-extreme years in the 2000s. By contrast, both the MMEM and FLUXCOM indicate a decrease (i.e. increasingly negative) in the GPP sensitivity to T during the three DH events, which switches from slightly positive in DH2003 (and normal summers) to negative during DH2010 and DH2018. As in the case of SM though, individual DGVMs show a very large spread in the δGPP/δT values, especially during the DH events.

At continental scale, individual model differences in simulated GPP$_{anom}$ (figure 2) can be explained (R^2 of linear fit between DGVM estimates) by differences in the sensitivity of GPP to either T or SM, depending on the DH event. In DH2003, the GPP$_{anom}$ simulated by DGVMs show a stronger relationship to δGPP/δSM (explaining 42% of the between-model variability) compared to δGPP/δT (10% of the between-model variability). In both DH2010 and DH2018, differences in δGPP/δT are a better predictor of differences in modelled GPP between DGVMs (61% and 70% of the between-model variability, respectively) than δGPP/δSM (22% and 9% in DH2010 and DH2018, respectively). The slopes of the regression lines (figure 2) are also steeper in DH2010 and DH2018, indicating a greater contribution of δGPP/δT to differences in the simulated GPP$_{anom}$ (i.e. larger under- or overestimation of the sensitivity is associated with bigger differences in simulated GPP$_{anom}$). The MMEM is generally closer to FLUXCOM for the three events, while individual DGVMs show variable results between events.

(d) Land-cover contribution to gross primary productivity responses

The differences in continental-scale GPP$_{anom}$ and sensitivity to drought might be related with the distinct geographical distribution of each event, affecting regions with different land-cover composition, and imposing distinct climate anomalies in the surrounding regions. To isolate these effects, we calculate GPP$_{anom}$ for different SM anomaly classes, forest or crop cover fraction, and their corresponding extent for each DH event for FLUXCOM and DGVM MMEM (figure 3).

Generally, for the same drought class (individual columns with SM anomaly [−3, 0]), pixels with higher crop cover tend to show stronger negative GPP$_{anom}$. The inverse can be found for forest cover, with pixels with more than 80% forest cover showing less negative or even positive GPP$_{anom}$ for negative SM$_{anom}$ (DH2003 and DH2018). This is consistent with results from Flach et al. [17] and Walther et al. [40]. In DH2010, though, even regions with SM$_{anom}$ between +1 and +1.5σ (wet conditions) showed negative GPP$_{anom}$, especially in
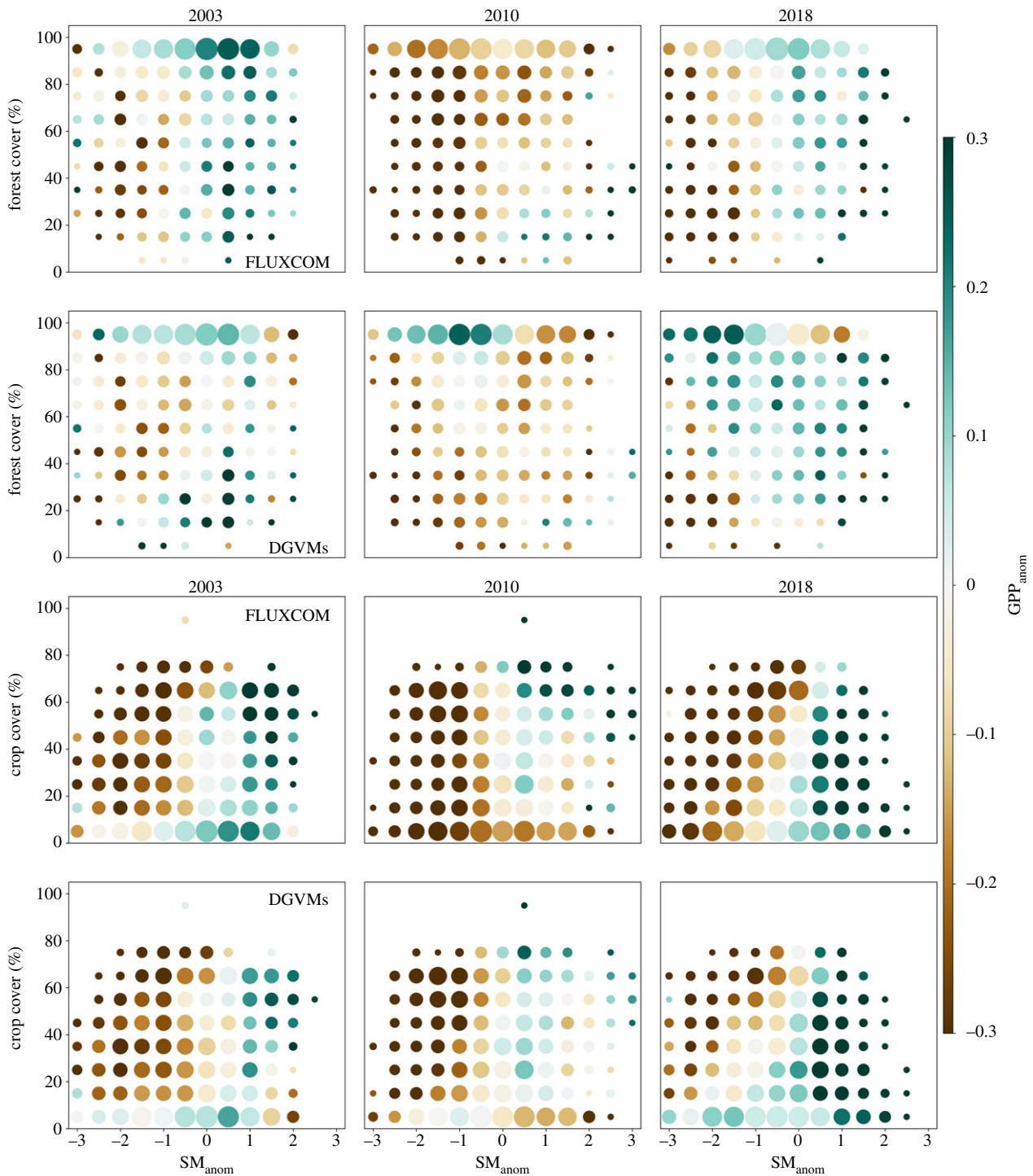
Figure 3. Impact of the DH events (left 2003, middle 2010, right 2018) in different land-cover types. The standardized GPP$_{anom}$ (colourmap) for the data-driven FLUXCOM product and for the MMEM of DGVMs are shown for different bins of SM$_{anom}$ and of forest (two top rows) and cropland (two bottom rows) cover fractions. The size of the markers is proportional to the number of pixels within a given class (all years have the same total number of pixels, corresponding to the land area). (Online version in colour.)

forest-dominated regions, explaining the strong negative GPP$_{anom}$ at continental scale in this event (figure 1). On the contrary, in DH2018, some of the classes with 50% or more forest cover registered positive GPP$_{anom}$ for negative SM$_{anom}$. Only regions with SM$_{anom}$ below −2σ showed strong negative GPP$_{anom}$ in the higher forest cover classes. In DH2018, the regions affected by strong drought conditions are mostly associated with crop fractions below 50%. Comparing with DH2018, drought conditions in DH2003 and DH2010 affected more pixels dominated by croplands (size of the markers). Wetter than average conditions affected less forest-dominated pixels in DH2018 and especially in DH2003.

DGVMs show similar results as FLUXCOM but tend to simulate weaker GPP$_{anom}$ than FLUXCOM in regions with higher forest cover and the opposite response to drought in pixels with 90–100% forest cover in DH2010 and DH2018.

5. Discussion

The three DH events in 2003, 2010 and 2018 imposed strong drought and extreme heat anomalies over large areas in Europe, all having a clear signal at the continental scale [3,21]. FLUXCOM and DGVMs estimate strong anomalies in carbon

fluxes at the continental-scale summer carbon fluxes during these events. However, DGVMs tend to show a large spread for NEP_{anom} and GPP_{anom} and even disagreement on the sign of anomalies (figure 1; electronic supplementary material, figures S4 and S5). Generally, the DGVMs tend to simulate higher summer GPP_{anom} (more positive or less negative) than FLUXCOM in all events (figure 1) and the MMEM shows moderate agreement in the spatial distribution of flux anomalies (electronic supplementary material, figures S4 and S5). Moreover, the differences between DGVMs and FLUXCOM in the spatial patterns in GPP and NEP_{anom} differ for the three DH events. The flux anomalies in response to DH2003 and DH2010 show better agreement between DGVMs and FLUX-COM at both continental and pixel scale than for DH2018 (electronic supplementary material, figures S4 and S5). In particular, the disagreement in the magnitude and sign of anomalies is particularly large in Scandinavia during DH2018. This might be because DGVMs do not simulate well drought-induced mortality, or because they underestimate soil organic carbon and consequently TER in the high latitudes. However, the disagreement might also be because FLUXCOM lacks information about carry-over effects from the previous spring warming, or because of its known limited performance in simulating inter-annual variability [22,23].

To assess the skill of both DGVMs and FLUXCOM in estimating the DH2018 anomalies in carbon fluxes, we compared their estimates of NEP and GPP with those of a compilation of 52 eddy covariance (EC) sites [41]. FLUXCOM shows generally better agreement with *in situ* NEP than GPP (electronic supplementary material, figure S7), while the opposite is found for DGVMs. For most sites, EC NEP or GPP anomalies are within the range of FLUXCOM and DGVMs. The results from the EC data seem to support positive GPP and anomalies in higher latitude sites, consistent with DGVMs, but in fewer high-latitude sites does this result in positive NEP anomalies, more consistent with FLUXCOM. Part of the disagreements can be explained by the coarse resolution of the datasets and the fact that the fluxes from EC sites correspond to a single ecosystem type, while the pixels of both FLUXCOM and DGVMs are mixed. FLUXNET sites are, therefore, not necessarily representative of the climate and composition of the grid cell modelled by DGVMs or FLUXCOM. For sites closely located but with different ecosystem types, the DGVMs and FLUX-COM estimates correspond roughly to the average anomaly over all ecosystem types. For an accurate comparison, DGVMs would need to be forced with the local environmental conditions for the FLUXNET sites.

Using global runs from DGVMs, Schewe *et al*. [16] proposed that all models underestimated the magnitude of GPP_{anom} in response to DH2003, compared to those from the MODIS (moderate resolution imaging spectroradiometer) product. However, the MODIS data are also partly modelled and do not simulate well the sensitivity to SM [42]. Even though we find a similar tendency for underestimation of negative GPP_{anom} (or overestimation of positive values in 2018) during DH events, the agreement between DGVMs and FLUXCOM is event dependent. Moreover, some DGVMs show values very close to those of the FLUXCOM mean, and some even report stronger negative standardized anomalies (figures 1 and 2). It is possible that part of the better agreement of estimated anomalies between DGVMs and FLUXCOM compared to the results of Schewe *et al*. [16] is due to the use of the most recent ERA5 forcing and

of high spatial and temporal resolution runs, which should better represent extremes in space and in time. Moreover, by using different runs of the FLUXCOM model we can report uncertainties in observation-based flux estimates, which is not the case for most remote-sensing products.

Because of the good agreement of simulated SM and of FLUXCOM WAI with ERA5 (electronic supplementary material, figure S2), the differences between the DGVMs and FLUXCOM are not likely to be due to poor simulation of the hydrology by DGVMs, but might rather indicate that DGVMs have too weak or too strong controls of water limitation on photosynthesis, or by missing the variable sensitivity of photosynthesis to specific climate conditions during droughts. Such differences may be due to differences across DGVMs and FLUX-COM in the storage capacity of moisture within the soil. Fu *et al*. [43] indicated that extreme drought affected the magnitude of the sensitivity of GPP from EC measurements to climate conditions, particularly SM, but the change in the sensitivity of productivity to drought was partly dependent on ecosystem type. Consistent with their results, FLUXCOM reports generally higher sensitivity of GPP to SM and more negative sensitivity to T at continental scale during the DH events as compared to other years (electronic supplementary material, figure S6). In FLUX-COM, the increase in $\delta GPP/\delta SM$ is particularly pronounced for DH2003 and DH2018, while the decrease in $\delta GPP/\delta T$ is stronger for DH2010 and DH2018. These changes in sensitivity to SM and T are consistent with a continental-scale transition from a temperature-limited regime to a water-limited regime during the extreme summers. Even though all data have been detrended, it is possible that the increasingly negative sensitivity of GPP_{anom} to T during summer is linked to progressively warmer background conditions imposing increasing water stress. Moreover, it has been shown that increased vapour pressure deficit in combination with reduced stomatal conductance under elevated CO_2 concentration might increase heat stress in plants, due to reduced evaporative cooling [44].

The relative strengths of $\delta GPP/\delta SM$ and $\delta GPP/\delta T$ indicate that continental-scale GPP response to the three events differs in the relative importance of T and SM: DH2003 was mainly driven by SM_{anom}, DH2018 mainly by temperature, and DH2010 by both. Consistently, the differences between DGVMs in simulated continental-scale GPP_{anom} for each event can be mainly explained by the modelled GPP sensitivity to SM in DH2003, and to T in DH2010 and DH2018. Because of the distinct role of T versus SM in explaining the large-scale anomalies in GPP, it is not possible to single out the best performing DGVMs, since a given DGVM may perform better for some DH events than others. However, the MMEM is generally more stable and shows better performance than individual models, which shows that using ensembles of DGVMs might reduce the impact of model errors in carbon flux estimates.

We find that the difference between the continental-scale impacts of the DH events on summer GPP and the changes in the large-scale sensitivity of GPP to SM can be traced back to the regional fingerprint of each event. DH2010 registered the lowest GPP at continental scale since the negative impacts of DH were combined with negative GPP_{anom} in many forest-dominated regions experiencing wet conditions (figure 3), for example, Scandinavia and parts of central Europe (electronic supplementary material, figures S2 and S4). In DH2018, the opposite was found, with some of the forest-dominated regions under moderate drought registering increased GPP (figure 3), in parts of Scandinavia and north-western Russia

(electronic supplementary material, figure S4). In most of Scandinavia and western Russia, ecosystems are temperature and light limited [12]. Therefore, wet conditions associated with more cloudiness in DH2010 lead to a negative impact on GPP rather than a positive one, and the drought conditions in DH2018 promoted increased productivity due to sunnier and warmer conditions.

Several studies have shown a relevant contribution of atmospheric circulation anomalies (e.g. blocking conditions) to the occurrence of temperature extremes in the northern hemisphere [3,4,45]. These persistent atmospheric circulation conditions affect not only the climate anomalies in the centres of DH events, but might also impose opposite anomalies in temperature or rainfall in the surrounding regions. Depending on whether these conditions are favourable for ecosystem productivity or not, they might lead to anomalies of the same sign of the DH events or offset the productivity losses in the DH regions. This is possibly the case for the wet conditions in Scandinavia in DH2010 and southern Europe in DH2018. Climate models suggest an eastward shift in blocking conditions in Europe [46]. Whether this could result in more frequent DH2010-like events in terms of European carbon balance remains unclear.

Despite negative (or weakly positive in 2018) NEP_{anom} in summer months, these extreme events were associated with increased CO_2 uptake at annual scale in both DGVMs and FLUXCOM. This can be explained by offsetting effects of both the preceding spring and the subsequent autumn seasons. On the one hand, all events were preceded by mild, although drier than average, springs which resulted in increased GPP and NEP, despite concurrent TER increase. Even if increased spring GPP might have contributed to amplify summer drought [11,21], it contributes to partly offset summer losses. Such offsetting effect of spring to summer had also been previously reported by Flach *et al.* [10] for some regions affected by DH2010, based on another set of FLUXCOM estimates. On the other hand, DGVMs predict that NEP recovered rather quickly following the extreme summers. The decrease in GPP during the extreme summer implies less substrate for decomposition during the months following the DH events, reducing TER during the release period. However, tree ring data indicate that unfavourable August–October climate conditions led to growth impairment over subsequent years in Europe [47]. The fast recovery in DGVMs may be an artefact because most models do not represent drought-induced mortality (included only in CABLE-POP and LPJ-GUESS) and possibly miss other long-term effects on growth.

If warm springs would be a precursor of extreme summers [5], then the positive effects of earlier growing-season start could be expected to offset the consequent impacts of summer DH. However, it is not clear if, with increasing temperatures, the negative effects of warming might start dominating the response of ecosystems [44]. Our sample of extreme events is too small to conclude about large-scale changes in the sensitivity of carbon fluxes to climate anomalies. Nevertheless, the progressively decreasing sensitivity of GPP

to T found for the three DH events might support the hypothesis of increasingly negative effects, for example, through increase in VPD, superimposed on year-to-year variability, or accumulated effect of more frequent disturbances.

6. Conclusion

In this study, we have compared the impacts of the three extreme summers in the past two decades on European ecosystems' CO_2 uptake. The events differed in regions affected, the climate anomalies preceding each event, as well as in the relative strength of drought versus heat anomalies. We find that the resulting impact of summer DH on continental carbon fluxes differs between events, but the relative contribution of GPP and TER to the net carbon balance cannot be constrained with confidence given the uncertainty of our estimates.

We find that, at continental scale, two of the three extreme summers did not result in strong negative anomalies in annual net CO_2 uptake, which can be explained in part by the offsetting effects of increased productivity from warm springs and decreased respiration in autumn and in part by regional compensation of anomalies. However, we also find that regional compensation in climate anomalies might act to reduce ecosystem productivity, depending on their geographical distribution and biomes affected, as was the case in 2010. Whether such seasonal and spatial compensation effects are limited to Europe (e.g. specific biome composition, large-scale atmospheric patterns) or can be found in other extreme events elsewhere is not clear.

We find a progressively stronger negative effect of heat anomalies on ecosystem productivity between each event, which might indicate a transition towards a more water-limited regime at continental scale, and a progressive increase of the negative effects of temperature on ecosystems.

References

1. Ciais PH *et al.* 2005 Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* **437**, 529–533. (doi:10.1038/nature03972)

2. Trigo RM, García-Herrera R, Díaz J, Trigo IF, Valente MA. 2005 How exceptional was the early August

2003 heatwave in France? *Geophys. Res. Lett.* **32**, L10701. (doi:10.1029/2005GL022410)

3. Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R. 2011 The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332**, 220–224. (doi:10.1126/science.1201224)

4. Schär C, Vidale PL, Luthi D, Frei C, Haberli C, Liniger MA, Appenzeller C. 2004 The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–336. (doi:10.1038/nature02300)

5. Miralles DG, Teuling AJ, van Heerwaarden CC, Vila-Guerau de Arellano J. 2014 Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat. Geosci.* **7**, 345–349. (doi:10.1038/ngeo2141)

6. Fischer EM, Seneviratne SI, Lüthi D, Schär C. 2007 Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophys. Res. Lett.* **34**, L06707. (doi:10.1029/2006GL027992)

7. Coumou D, Rahmstorf S. 2012 A decade of weather extremes. *Nat. Clim. Change* **2**, 491–496. (doi:10.1038/nclimate1452)

8. Beillouin D, Schauberger B, Bastos A, Ciais P, Makowski D. 2020 Impact of extreme weather conditions on European crop production in 2018. *Phil. Trans. R. Soc. B* **375**, 20190510. (doi:10.1098/rstb.2019.0510)

9. San-Miguel-Ayanz J *et al.* 2019 *Forest fires in Europe, Middle East and North Africa 2018.* EUR 29856 EN. Luxembourg: Publications Office of the European Union.

10. Flach M, Sippel S, Gans F, Bastos A, Brenning A, Reichstein M, Mahecha MD. 2018 Contrasting biosphere responses to hydrometeorological extremes: revisiting the 2010 western Russian heatwave. *Biogeosciences* **16**, 6067–6085. (doi:10.5194/bg-15-6067-2018)

11. Teuling AJ *et al.* 2013 Evapotranspiration amplifies European summer drought. *Geophys. Res. Lett.* **40**, 2071–2075. (doi:10.1002/grl.50495)

12. Seddon AW, Macias-Fauria M, Long PR, Benz D, Willis KJ. 2016 Sensitivity of global terrestrial ecosystems to climate variability. *Nature* **531**, 229–232. (doi:10.1038/nature16986)

13. Klesse S *et al.* 2018 A combined tree ring and vegetation model assessment of European forest growth sensitivity to interannual climate variability. *Glob. Biogeochem. Cycles* **2**, 1226–1240. (doi:10.1029/2017GB005856)

14. Reichstein M *et al.* 2007 Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Glob. Change Biol.* **13**, 634–651. (doi:10.1111/j.1365-2486.2006.01224.x)

15. Bastos A, Trigo C, Trigo R, Running SW. 2014 Analysing the spatio-temporal impacts of the 2003 and 2010 extreme heatwaves on plant productivity in Europe. *Biogeosciences* **11**, 3421. (doi:10.5194/bg-11-3421-2014)

16. Schewe J *et al.* 2019 State-of-the-art global models underestimate impacts from climate extremes. *Nat. Commun.* **10**, 1005. (doi:10.1038/s41467-019-08745-6)

17. Teuling AJ *et al.* 2010 Contrasting response of European forest and grassland energy exchange to heatwaves. *Nat. Geosci.* **3**, 722–727. (doi:10.1038/ngeo950)

18. Albergel C, Dutra E, Munier S, Calvet J-C, Munoz-Sabater J, de Rosnay P, Balsamo G. 2018 ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? *Hydrol. Earth Syst. Sci.* **22**, 3515–3532. (doi:10.5194/hess-22-3515-2018)

19. Urraca R, Huld T, Gracia-Amillo A, Martinez-de-Pison FJ, Kaspar F, Sanz-Garcia A. 2018 Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data. *Sol. Energy.* **164**, 339–354. (doi:10.1016/j.solener.2018.02.059)

20. Cerlini PB, Meniconi S, Brunone B. 2017 Groundwater supply and climate change management by means of global atmospheric datasets. Preliminary results. *Procedia Eng.* **186**, 420–427. (doi:10.1016/j.proeng.2017.03.245)

21. Bastos A *et al.* 2020 Direct and seasonal legacy effects of the 2018 heat and drought on European ecosystem productivity. *Sci. Adv.* **6**, eaba2724. (doi:10.1126/sciadv.aba2724)

22. Tramontana G *et al.* 2016 Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* **13**, 4291–4313. (doi:10.5194/bg-13-4291-2016)

23. Jung M *et al.* 2020 Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences* **17**, 1343–1365. (doi:10.5194/bg-17-1343-2020)

24. Reichstein M *et al.* 2005 On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Glob. Change Biol.* **11**, 1424–1439. (doi:10.1111/j.1365-2486.2005.001002.x)

25. Lasslop G *et al.* 2010 Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Glob. Change Biol.* **16**, 187–208. (doi:10.1111/j.1365-2486.2009.02041.x)

26. Haverd V, Smith B, Nieradzik L, Briggs P, Woodgate W, Trudinger C, Canadell JG, Cuntz M. 2018 A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis. *Geosci. Model Dev.* **11**, 2995–3026. (doi:10.5194/gmd-11-2995-2018)

27. Joetzjer E, Delire C, Douville H, Ciais P, Decharme B, Carrer D, Verbeeck H, De Weirdt M, Bonal D. 2015 Improving the ISBA (CC) land surface model simulation of water and carbon fluxes and stocks over the Amazon forest. *Geosci. Model Dev.* **8**, 1709–1727. (doi:10.5194/gmd-8-1709-2015)

28. Tian H *et al.* 2018 The global N₂O Model Intercomparison Project (NMIP). *Bull. Am. Meteorol. Soc.* **99**, 1231–1251. (doi:10.1175/BAMS-D-17-0212.1)

29. Meiyappan P, Jain AK, House JI. 2015 Increased influence of nitrogen limitation on CO_2 emissions from future land use and land use change. *Glob. Biogeochem. Cycles* **29**, 1524–1548. (doi:10.1002/2015GB005086)

30. Mauritsen T *et al.* 2018 Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM 1.2) and its response to increasing CO_2. *J. Adv. Model Earth Syst.* **11**, 998–1038. (doi:10.1029/2018MS001400)

31. Smith B, Warlind D, Arneth A, Hickler T, Leadley P, Siltberg J, Zaehle S. 2014 Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences* **11**, 2027–2054. (doi:10.5194/bg-11-2027-2014)

32. Lienert S, Joos F. 2018 A Bayesian ensemble data assimilation to constrain model parameters and land-use carbon emissions. *Biogeosciences* **15**, 2909–2930. (doi:10.5194/bg-15-2909-2018)

33. Zaehle S, Friend AD, Friedlingstein P, Dentener F, Peylin P, Schulz M. 2010 Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance. *Glob. Biogeochem. Cycles* **24**, GB1006. (doi:10.1029/2009GB003522)

34. Krinner G *et al.* 2005 A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Glob. Biogeochem. Cycles* **19**, GB1015. (doi:10.1029/2003GB002199)

35. Guimberteau M *et al.* 2017 ORCHIDEE-MICT (revision 4126), a land surface model for the high-latitudes: model description and validation. *Geosci. Model Dev. Discuss.* **2017**, 1–65. (doi:10.5194/gmd-2017-122)

36. Walker AP *et al.* 2017 The impact of alternative trait-scaling hypotheses for the maximum photosynthetic carboxylation rate (Vcmax) on global gross primary production. *New Phytol.* **215**, 1370–1386. (doi:10.1111/nph.14623)

37. Dlugokencky E, Tans P. 2018 *Trends in atmospheric carbon dioxide*. Natl Ocean Atmospheric Adm Earth Syst Res Lab NOAAESRL. See http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html.

38. Hurtt G *et al.* 2011 Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Clim. Change.* **109**, 117–161. (doi:10.1007/s10584-011-0153-2)

39. ESA. 2017 *Land cover CCI product user guide version 2*. Tech. Rep. See http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

40. Walther S, Duveiller G, Jung M, Guanter L, Cescatti A, Camps-Valls G. 2019 Satellite observations of the contrasting response of trees and grasses to variations in water availability. *Geophys. Res. Lett.* **46**, 1429–1440. (doi:10.1029/2018GL080535)

41. ICOS. 2020 *Drought-2018 ecosystem eddy covariance flux product for 52 stations in Europe in FLUXNET-*

Archive format. See https://www.icos-cp.eu/data-products/YVR0−4898.

42. Stocker BD, Zscheischler J, Keenan TF, Prentice IC, Seneviratne SI, Peñuelas J. 2019 Drought impacts on terrestrial primary production underestimated by satellite monitoring. *Nat. Geosci.* **12**, 264–270. (doi:10.1038/s41561-019-0318-6)

43. Fu Z *et al*. 2020 Sensitivity of gross primary productivity to climatic drivers during the summer drought of 2018 in Europe. *Phil. Trans. R. Soc. B* **375**, 20190747. (doi:10.1098/rstb.2019.0747)

44. Obermeier W *et al.* 2017 Reduced CO_2 fertilization effect in temperate C3 grasslands under more extreme weather conditions. *Nat. Clim. Change* **7**, 137. (doi:10.1038/nclimate3191)

45. Pfahl S, Wernli H. 2012 Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales. *Geophys. Res. Lett.* **39**, L12807. (doi:10.1029/2012GL052261)

46. Woollings T *et al.* 2018 Blocking and its response to climate change. *Curr. Clim. Change Rep.* **4**, 287–300. (doi:10.1007/s40641-018-0108-z)

47. Babst F, Carrer M, Poulter B, Urbinati C, Neuwirth B, Frank D. 2012 500 years of regional forest growth variability and links to climatic extreme events in Europe. *Environ. Res. Lett.* **7**, 045705. (doi:10.1088/1748-9326/7/4/045705)