



**HAL**  
open science

## State-of-the-art global models underestimate impacts from climate extremes

Jacob Schewe, Simon Gosling, Christopher Reyer, Fang Zhao, Philippe Ciais,  
Joshua Elliott, Louis Francois, Veronika Huber, Heike Lotze, Sonia I  
Seneviratne, et al.

### ► To cite this version:

Jacob Schewe, Simon Gosling, Christopher Reyer, Fang Zhao, Philippe Ciais, et al.. State-of-the-art global models underestimate impacts from climate extremes. *Nature Communications*, 2019, 10 (1), pp.1005. 10.1038/s41467-019-08745-6 . hal-02895259

**HAL Id: hal-02895259**

**<https://hal.science/hal-02895259>**

Submitted on 17 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ARTICLE

<https://doi.org/10.1038/s41467-019-08745-6>

OPEN

# State-of-the-art global models underestimate impacts from climate extremes

Jacob Schewe  et al.<sup>#</sup>

Global impact models represent process-level understanding of how natural and human systems may be affected by climate change. Their projections are used in integrated assessments of climate change. Here we test, for the first time, systematically across many important systems, how well such impact models capture the impacts of extreme climate conditions. Using the 2003 European heat wave and drought as a historical analogue for comparable events in the future, we find that a majority of models underestimate the extremeness of impacts in important sectors such as agriculture, terrestrial ecosystems, and heat-related human mortality, while impacts on water resources and hydropower are over-estimated in some river basins; and the spread across models is often large. This has important implications for economic assessments of climate change impacts that rely on these models. It also means that societal risks from future extreme events may be greater than previously thought.

---

Correspondence and requests for materials should be addressed to J.S. (email: [schewe@pik-potsdam.de](mailto:schewe@pik-potsdam.de)).

<sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.

Estimation of the total damages caused by climate change requires a quantification of climate impacts across a large range of economic and societal sectors. These sectors include agriculture<sup>1</sup>, water resources<sup>2</sup>, energy supply and demand<sup>3</sup>, human health<sup>4</sup> and ecosystem services<sup>5</sup>. There are approaches that integrate damages across sectors, such as the highly idealised damage functions used in integrated assessment modelling<sup>6</sup>, but also more sophisticated, coupled economic modelling frameworks that combine individual sectoral models<sup>7–9</sup>. However, these approaches are centred on gradual changes in physical and biophysical indicators—such as crop yields or water resources—and largely ignore the impacts of extreme climate and weather events.

This is a serious research gap because such events cause enormous damages<sup>10</sup>. For addressing it, sectoral impact models must be able to credibly represent the impacts of extreme events. The goal of this paper is to test whether this is the case in the complex, process-based impact models that are routinely being applied in global-scale climate impact assessment<sup>1–3</sup>. While these models may be too costly to integrate them directly in cross-sectoral economic models, they are the benchmark for any simpler models. More generally, these complex impact models, in conjunction with global climate models, form the basis of much of our current knowledge about future global climate change impacts, as reflected in the Intergovernmental Panel on Climate Change reports, for instance. Whether they capture extreme events well is therefore a key concern even beyond the application in economic assessments.

And yet, it is not known how well the current suite of models can reproduce the multi-sectoral impacts of a given climatic extreme event. Global process-based impact models have been evaluated in terms of average quantities and sometimes in terms of inter-annual or intra-annual variability<sup>11–14</sup>, but their performance under extreme conditions has rarely been tested at large spatial scale<sup>15</sup>, and never—to our knowledge—in a multi-sector setting. And since events that are very rare today may become much more frequent in the future<sup>16</sup>, testing for variability alone may not be enough.

Here, we choose the 2003 European heat wave and drought (EHWD) event as a test case. The EHWD was substantially stronger than previously observed events; it severely impacted several important sectors across a large geographical area, and its impacts are relatively well documented. We examine the impacts of the EHWD in a large ensemble of state-of-the-art impact models covering agriculture, water resources, terrestrial and marine ecosystems, energy, and human health, for the first time in a common modelling framework. For each of these sectors, we identify key observed impacts of the 2003 EHWD reported in the literature and/or recorded in public databases, and examine how closely the models—driven by observations-based climate data—reproduce those impacts. As a common impact metric, we choose the deviation of 2003 from the historical average, adjusted for long-term trends, and normalised by the historical standard deviation (except for human health; see Methods). We thereby circumvent potential biases in the baseline or the average inter-annual variability, and instead focus on the models' ability to pick out the anomalous 2003 event from the rest of the time series.

The Results section first provides a climatological analysis of the 2003 EHWD, and then presents impact model results for each sector. In the Discussion section, we summarise and evaluate our findings across sectors, and discuss their implications for integrated assessments of climate change impacts and for future model development.

## Results

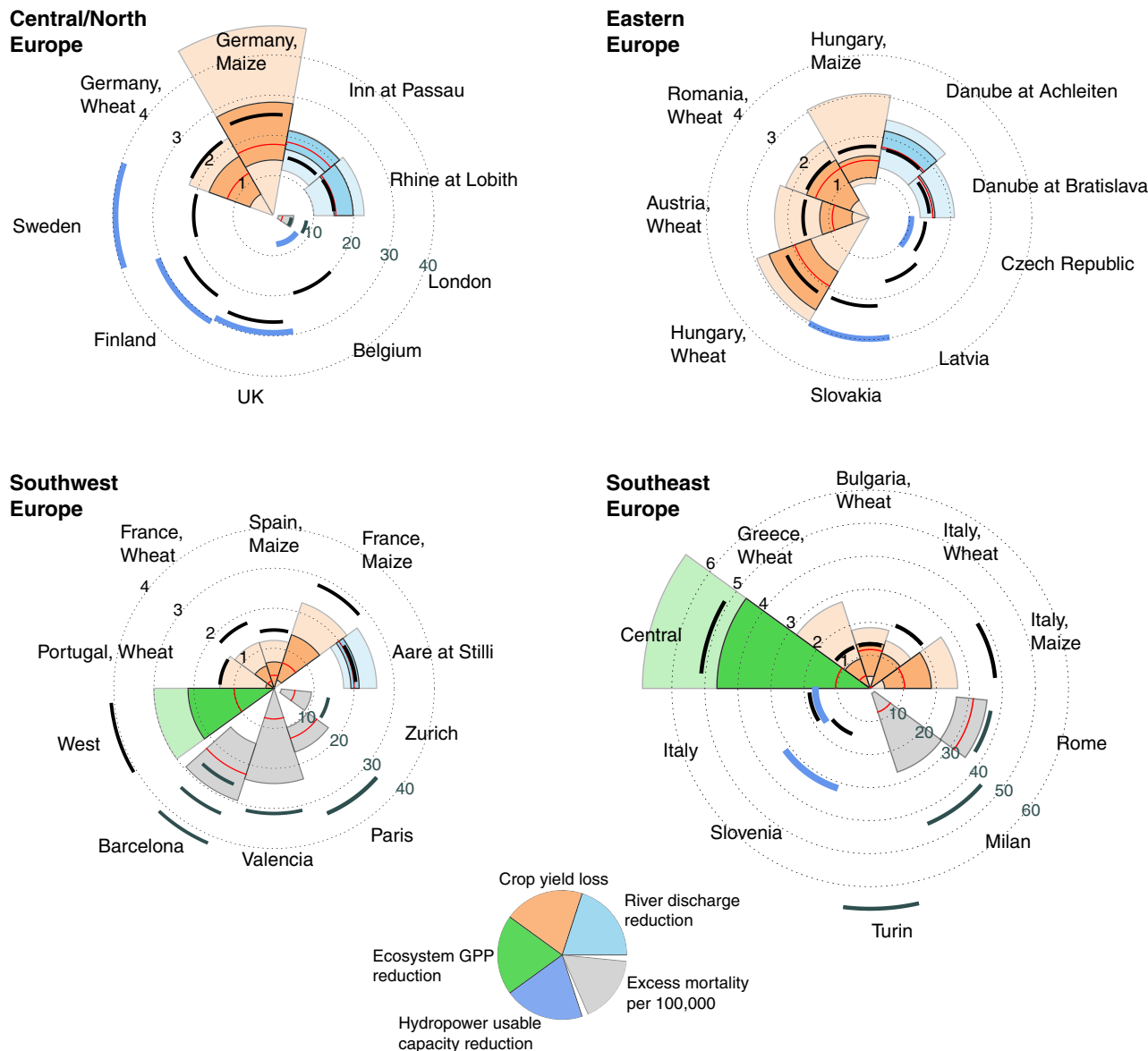
**The 2003 European heat wave and drought.** The 2003 EHWD stretched over the entire summer, with large and persistent hot anomalies especially during June and August (Supplementary Fig. 1), and it extended across much of Western and Central Europe (Supplementary Fig. 2). The June–August average temperature anomalies (relative to 1961–1990) were extreme, reaching 2 °C (2 standard deviations ( $\sigma$ ) above the mean) averaged over Europe and more than 5 °C regionally (more than  $3\sigma$ , and in some locations  $5\sigma$ )<sup>17,18</sup>. Due to both decreased precipitation and increased evapotranspiration, the high temperatures were accompanied by an intense drought, and the dry soils in turn amplified the heat wave<sup>19,20</sup>.

The EHWD had numerous impacts on the environment, economy and human health<sup>18</sup>. In this paper, we focus on large-scale impact indicators which can be compared between observations and global-scale models (Fig. 1). In terms of these indicators, the summer of 2003 is characterised by anomalies of up to  $5\sigma$  in observed data. In particular, southern Europe saw extreme reductions in ecosystem gross primary productivity (GPP) and large excess human mortality rates. Substantial relative reductions in crop yields, river flow and hydropower production were experienced across the different parts of Europe affected by the EHWD. The ability of impact models to capture these large anomalies is summarised in Fig. 1, and discussed in the following sections.

The EHWD was exceptional compared to the historic record, but given continued global warming, comparable events are anticipated to occur about every 10 years by the middle of the 21st century, depending on future greenhouse gas emissions<sup>21–23</sup>. Even when changes in the mean climate state are discounted—tantamount to assuming full adaptation to gradual climate change—relative deviations like the 2003 EHWD could still occur compared to the background climate state in the late 21st century<sup>24</sup>. Furthermore, the impacts of the 2003 event could have been much greater if spring conditions had been drier<sup>19</sup>, e.g. similar to those that prevailed in 2011<sup>25</sup>, which underlines the risk of even more-extreme events in the future. Being able to estimate the damages from such events is therefore crucial for assessments of future climate change impacts.

**River flow and water resources.** Due to the prolonged rainfall deficit (Supplementary Figs. 1–3), the summer of 2003 was anomalously dry in central Europe. This was visible in surface runoff and river discharge levels<sup>26–28</sup>, which are indicators of renewable freshwater availability for ecosystems and human uses, such as irrigation or the cooling of thermal power plants<sup>29</sup>. In August 2003, discharge in both the Rhine and Elbe rivers reached record low levels<sup>30</sup>. Navigation was impeded on the German section of the Rhine on 37 days in 2003 due to low flow<sup>31</sup>. Satellite-based gravity measurements indicate an exceptional depletion of terrestrial water storage compared to 2002<sup>32</sup>.

We examine monthly average river discharge at major gauging stations across Europe where continuous data since 1979 is available from the Global Runoff Data Centre, and compare this to an ensemble of global hydrological models (Methods). The observed data indicate negative anomalies of  $\sim 1.5$ – $2\sigma$  during August 2003 at five of the stations: in the Rhine (Lobith) and the Danube basin (Bratislava, Achleiten, and Inn at Passau-Ingling), as well as in the largest Swiss catchment, Aare (Figs. 1 and 2). These larger anomalies are reproduced closely by most model simulations, with a relatively small spread across the ensemble. The multi-model median is very close to the reported value at the Rhine, Tisza, Danube, and Aare stations. Conversely, many models overestimate the more moderate anomalies observed e.g.



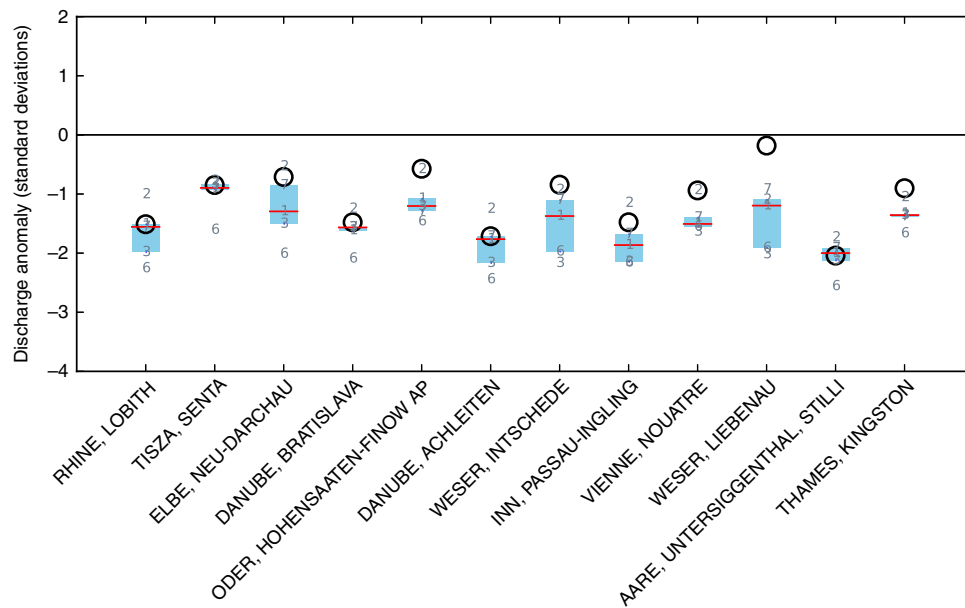
**Fig. 1** Multi-sector impacts of the 2003 EHWD. Black arcs represent observations, and colours represent model results. Units are standard deviations (black axis labels), except for human mortality which is given in excess deaths per 100,000 (grey axis labels). For river discharge, crop yields and ecosystem GPP, the thin red line marks the multi-model median; the dark-coloured segment marks the interquartile range; and the light-coloured segment marks the full range of model results. For hydropower, only one model is available which is marked by the thick blue arcs. For mortality, the red line and grey segments mark the median and the full range, respectively, across three climate forcing data sets and three different heat-mortality relations. Note the larger axis range for Southeast Europe. This figure only includes river discharge, crop yield, GPP and hydropower results for those locations where a negative anomaly larger than 1 standard deviation was observed. Figures 2–6 include further details on the data shown here, as well as additional data for locations with smaller or positive anomalies. The West and Central regions used for ecosystem GPP are defined in Methods

in the Elbe and Oder rivers, and the ensemble spread tends to be larger there (Fig. 2). Out of 12 stations, there are 5 stations where 75% or more of the models simulate a negative discharge anomaly larger than 1 standard deviation even though the observed anomaly is smaller. There is no station where the magnitude of the observed anomaly is underestimated by many models.

Results are similar in simulations that ignore the effects of human land use, dams and reservoirs, and water withdrawals (Supplementary Fig. 4), suggesting that present human modifications of the hydrological system do not substantially change the response of the system to this type of extreme event. This is consistent with the moderate effect of human interventions on

streamflow and drought conditions in central and western Europe reported in earlier studies<sup>33,34</sup>. Indeed, in the hydrological models that include human interventions, irrigation water demand would increase under drought, but this has little effect on river flow when the flow is already low and limits the amount available for withdrawal (Supplementary Fig. 5).

Model performance differs little between June, July and August (Supplementary Fig. 6). While there is some variation in which stations are matched best in each month, in those cases where there is a large anomaly (such as in the Danube in July) the model ensemble tends to reproduce that anomaly closely. Results are also insensitive to the choice of climate forcing data set



**Fig. 2** August average river discharge anomalies in 2003. Black circles are observed (GRDC) data. Grey numbers are the global hydrological models (see Methods); red lines indicate the median, and blue boxes the interquartile range, of the model ensemble. Stations are ordered by catchment size; the smallest catchment (Thames river at Kingston) has an area of about 10,000 km<sup>2</sup>, which corresponds to the size of four model grid cells

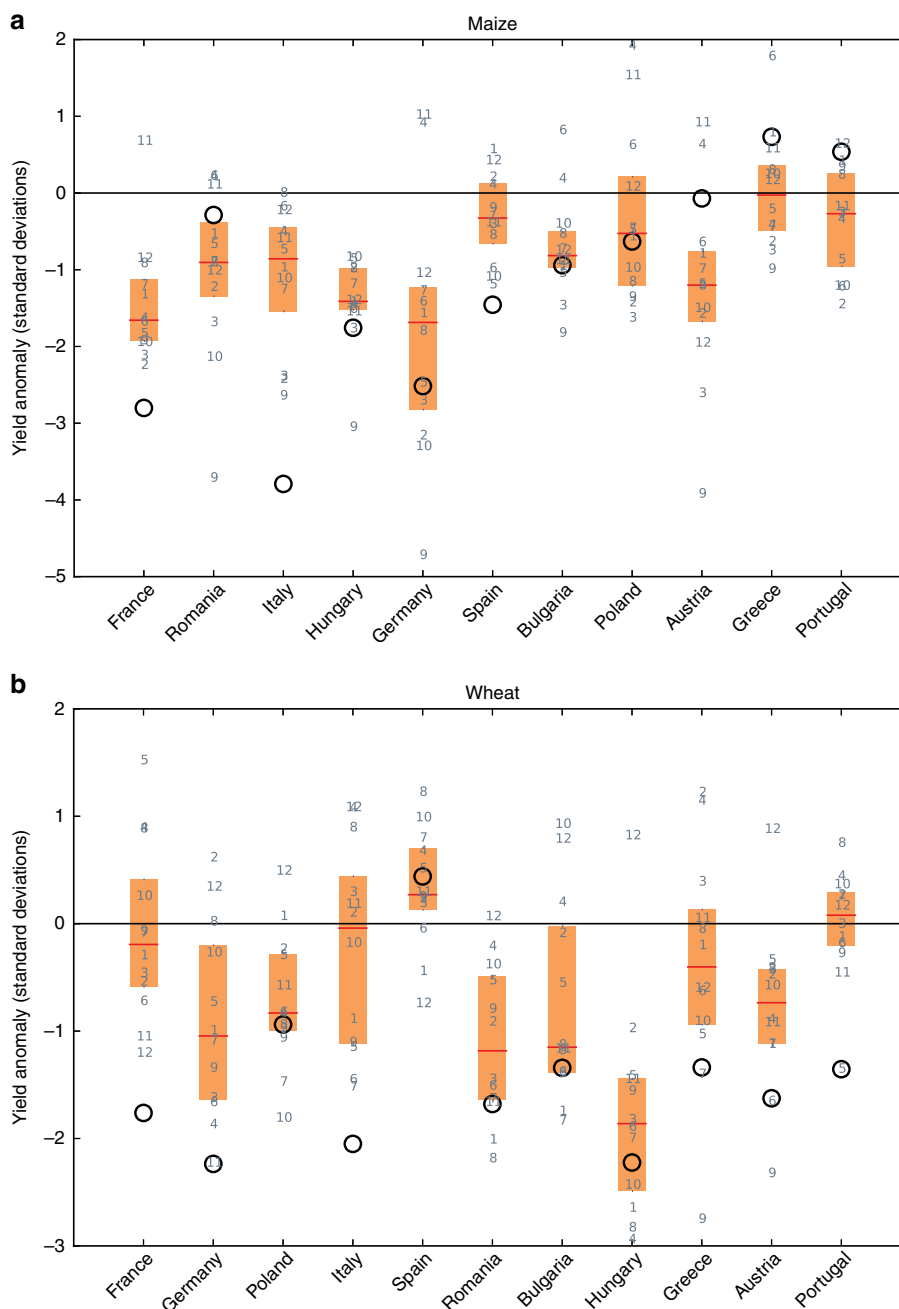
(Supplementary Fig. 7). Our results are consistent with recent studies attesting a relatively good performance of global hydrological models for discharge and runoff variability in European catchments, compared to other regions<sup>12,35</sup>. At the same time, the tendency towards false-alarms is in agreement with a dry bias induced by the models' potential evaporation schemes<sup>36</sup>. The fact that one of our models (number 6) appears consistently at the dry end of the ensemble may be because it applies a temperature-based evaporation scheme (Hamon) that has been shown to induce a particularly large bias when applied outside its calibration range<sup>36</sup>.

**Agriculture.** Agricultural summer crop yields were poor in 2003 due to the combined effect of the drought and the excessive temperatures recorded since June<sup>18,37</sup>. Low harvests incurred an estimated €13 billion of uninsured losses in the European Union (EU)<sup>38</sup>. At the EU level, maize and wheat were the most damaged crops, with production shortfalls of 21 and 11%, respectively<sup>18</sup>. We examine maize and wheat crop yields simulated by an ensemble of 12 global gridded crop models, and compare them with data reported by FAOSTAT (obtained from <http://faostat.fao.org/site/567/default.aspx> on 30 August 2016) for central and southern European countries with continuous data since 1979. Consistent with previous studies<sup>39</sup> and early assessments by the COPA-COGECA agricultural association<sup>40</sup> (Supplementary Table 2), we find the 2003 EHWD was associated with large negative yield anomalies in the FAOSTAT data set in France, Germany and Italy for both maize and wheat; as well as in Spain for maize, and in Austria and Portugal for wheat (Fig. 3, black circles). Additionally, we find substantial yield reductions of one standard deviation or more for both crops in Hungary, and for wheat in Bulgaria, Greece and Romania. Hungary and Romania in particular are among the EU's largest maize producers.

The ability of the crop model ensemble to reproduce these impacts is mixed (Fig. 3, boxes and numbers). Out of 5 countries with an observed negative maize yield anomaly larger than

1 standard deviation, there are 3 (4) countries where all models (more than 75% of models) underestimate that anomaly. For wheat, the numbers are 2 (7) out of 9 countries. The best-matching models differ from country to country, in line with a previous evaluation of these models<sup>11</sup> which also found mixed skill in reproducing overall inter-annual yield variability. There is broad agreement across models on the sign of the anomaly in most of the strongly impacted countries; except for wheat in Italy, France, Portugal and Greece, where nearly half of the models show a positive anomaly. In terms of magnitude, the large anomalies in France and Italy are underestimated by the entire ensemble for both crops. Generally, the agreement is somewhat better in Middle and Eastern European countries, such as Hungary, Bulgaria or Poland, compared to Western Europe. These general results are robust against changing the climate forcing data set, although the yield loss in Italy is better reproduced with an alternative data set (Supplementary Fig. 8). We also note that COPA-COGECA report a decline in maize yields in Austria by about 10% between 2002 and 2003, which is not reported in the FAOSTAT data (Supplementary Table 2); thus, the real observed value in Fig. 3 may be closer to the model ensemble mean than the FAOSTAT value shown by the black circle.

The poorer performance in Southwestern versus Eastern European countries may be due to more widespread irrigation in the former. The crop models assume full irrigation in irrigated areas and do not account for potential limitations in water availability due to drought, which induces an overestimation of irrigated yields, and thus biases total yields in countries with much irrigation, such as France or Italy (Supplementary Fig. 9). This is in line with a recent study showing that extreme heat leads to strong declines in maize yield in both observations and models only under rainfed conditions<sup>15</sup>. Moreover, vernalisation of winter crops (the requirement of cold temperatures for flowering) is a known problem in crop models<sup>41</sup>, and the positive sign of the simulated wheat yield anomalies in southern Europe in many models may be due to winter wheat flowering, maturing and being harvested early and thus escaping the heat wave<sup>42</sup>. Finally, as is the case with the other sectors, the crop model simulations were not specifically designed to



**Fig. 3** Crop yield anomalies in 2003. **a** Maize, **b** wheat. Black circles are observed data from FAOSTAT. Grey numbers are the 12 global gridded crop models; red lines indicate the median, and orange boxes the interquartile range, of the model ensemble. Countries are ordered by their total production (<http://ec.europa.eu/eurostat/web/agriculture/data/database>) in 2010, decreasing from left to right

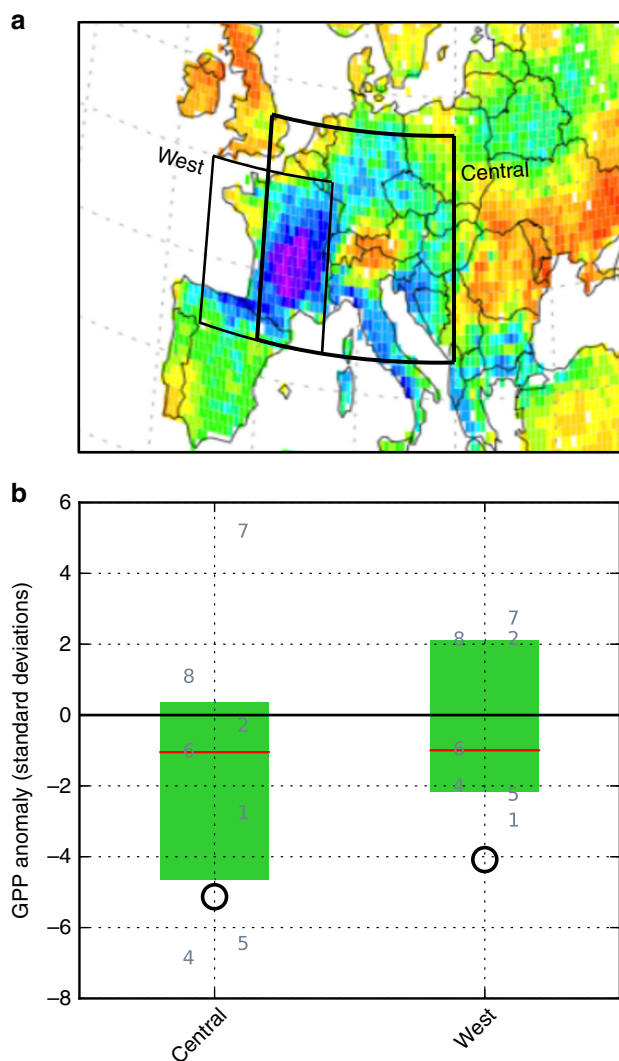
reproduce 2003 yields, and used standard management and phenology settings; for instance, fixed planting dates and a constant number of heat units required for a crop to mature.

**Terrestrial ecosystems.** The summer of 2003 saw a major reduction in overall GPP in parts of Europe, based on vegetation model results<sup>39,43</sup> and remote sensing greenness index data<sup>44,45</sup>. The largest anomaly was situated in France, but negative anomalies were also observed in northern Spain, Italy, and parts of the Balkan and Germany (Supplementary Fig. 10).

We evaluate an ensemble of global vegetation models over two overlapping rectangular regions (Fig. 4a): one (West) comprising France and northern Spain, where the centre of the negative GPP

anomaly was located; and one (Central) including Germany, central Italy, and the western Balkans, which also saw large negative anomalies, but also including the Alps which saw a positive anomaly<sup>46</sup>. As observational benchmark we use MODIS<sup>47,48</sup> GPP estimates derived from remotely sensed absorbed fraction of photosynthetically active radiation (fAPAR; Methods). We find negative anomalies of 4 to 5σ in MODIS-derived GPP during 2003 (Fig. 4b, black circles), consistent with previous analyses<sup>45</sup>.

The magnitude of the observed anomaly is underestimated by 75% of models in the Central region, and by all models in the West region. In both regions the model ensemble spreads over a large range that also includes positive anomalies and near-



**Fig. 4** Summer (June–August) gross primary productivity (GPP) anomaly in 2003. **a** Outlines of the West and Central regions, overlaid on a map of the observed GPP anomaly; see Supplementary Fig. 10 for a more detailed version of this map. **b** Regionally averaged anomaly. Black circles show MODIS remote-sensing-derived estimates. Grey numbers are the global vegetation models; numbers offset to the left represent models that were run without considering any human influence except climate change, while numbers offset to the right represent models that accounted for historical land-use patterns and, in some cases, anthropogenic water withdrawal. Red lines indicate the median, and green boxes the interquartile range, of the model ensemble. One model (no. 4) did not report total GPP but only GPP for individual plant functional types (PFTs); for this model, we show the sum of the four dominant PFTs in the relevant region

zero values (Fig. 4b, numbers). There is no systematic difference between models that include and exclude human-induced changes to the natural vegetation pattern (land use change). The estimates are largely insensitive to the choice of climate forcing data set (Supplementary Fig. 11), despite substantial differences in annual shortwave radiation between data sets<sup>49</sup>.

Our finding of a smaller sensitivity of GPP to drought in models than in observations is consistent with a previous multi-model study which, however, assessed a shorter time span<sup>45</sup>. Inspecting individual models' GPP time series (Supplementary Figs. 12 and 13), we note that those models that most closely reproduce the 2003 relative anomaly also exhibit a pronounced

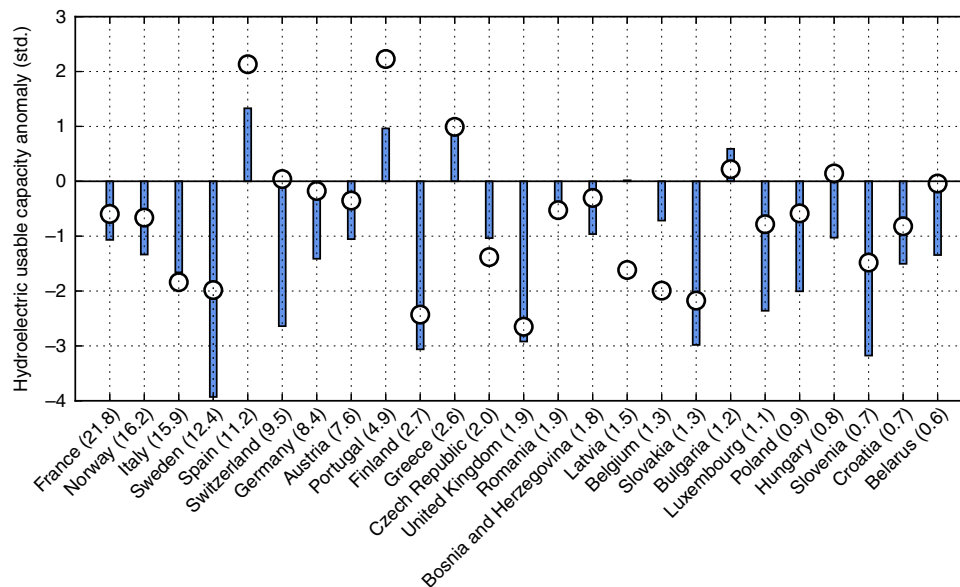
positive anomaly in 2007 and 2008, consistent with the MODIS estimates. The large spread across models for the reduction of GPP in the 2003 EHWD is despite a relatively high spatial correlation with MODIS-derived GPP globally in all these models<sup>49</sup>.

Previous studies<sup>50,51</sup> have suggested that many global vegetation models do not capture the vegetation response to drought, that is, extremely low soil moisture and air humidity, especially for unprecedented or long-lasting cases. While the models adopt mechanistic, or process-based, approaches, most models include empirical parameterisations for temperature and moisture response functions, which may be less reliable outside the historical range of variability for which they have been calibrated. Moreover, not all ecosystem models account for heat stress effects on plant photosynthesis and water stress effects on respiration<sup>52</sup>, which might explain the underestimation of the EHWD's impact by many models. Other possible reasons include the lack of differential species response to drought<sup>53</sup>, the lack of deep rooting access to water for forests, irrigation and more generally a poor description of the phenology of cultivated vegetation<sup>51</sup>.

**Energy.** Low water availability and high water temperatures during the 2003 EHWD affected electricity supply due to declines in the output of hydropower plants, and impaired operation of thermoelectric power plants that require freshwater for cooling. Across Europe, more than 30 nuclear power plant units had to reduce their production in summer 2003 because of constraints for cooling water uses<sup>54,55</sup>. Together with hydroelectric power restrictions, this led to a doubling in electricity spot market prices compared to the previous summer<sup>30</sup>. Even wind power potential was at record low levels in several countries (Methods).

We analyse model simulations of hydropower plants' usable capacity, i.e. the maximum available output at a given point in time (Methods). In the absence of observations of usable capacity, we compare these simulations to reported annual hydropower generation from the global database of the U.S. Energy Information Administration (EIA); while noting that a plant's actual power generation can be lower than its usable capacity depending on management. We exclude countries with less than 20 plants or less than 100 MW total installed capacity. The observed data show substantial negative anomalies of around 1.5–2.5 $\sigma$  in 2003 for a number of countries, among them Italy which is one of the largest hydropower producers in Europe (Fig. 5). Sweden, Finland, and the UK registered large negative anomalies, too; consistent with the observation that northern Scandinavia and the British Isles experienced dry conditions in summer 2003, even though they were not at the centre of the heat wave. At the same time, Spain, Portugal, and Greece registered above-average power generation; consistent with near-average summer precipitation and slightly above-average precipitation during the rest of the year (Supplementary Figs. 2–3).

The simulated anomalies are very close to the observed ones for Italy and the UK, relatively close for Finland and the Czech Republic, and also closely reproduce the positive deviation in Greece. Large observed negative anomalies in Latvia and Belgium are not reproduced. For the remaining countries, the model mostly overestimates the negative anomalies; underestimates the positive anomalies; and simulates below-average usable capacity for some countries that saw near-average power generation, such as Switzerland and Germany. The latter may be because the underlying hydrological model does not capture enhanced glacial melting during the heat wave, which could offset runoff deficits downstream. In total, out of 9 countries with an observed negative hydropower generation anomaly larger than 1 standard deviation, there are 3 countries where the usable capacity model



**Fig. 5** Hydropower anomalies in 2003. Black circles show anomalies in annual hydropower generation reported by EIA (Methods). Blue bars show anomalies in simulated annual hydroelectric usable capacity. Countries are ordered by installed capacity (in GW) as included in the model, indicated in parentheses

underestimates the magnitude of that anomaly. On the other hand, out of 25 countries, there are 10 countries where the model simulates a negative discharge anomaly larger than 1 standard deviation even though the observed anomaly is smaller in magnitude.

A previous evaluation of the hydropower model, using the same climate input, found relatively low root mean squared error in southern European countries, but larger values in some north European countries, with particularly poor performance in Latvia<sup>56</sup>. Our analysis broadly confirms this regional pattern, while focusing explicitly on a drought event. We note the complexity in comparing actual hydropower generation with hydropower usable capacity, which is simulated by the model. In particular, annual power generation is strongly affected by power demand, which was above-average in the winter of 2003 due to low temperatures. The hydropower model focuses on the physical impacts of changes in water resources on usable power plant capacity; economic feedbacks of the assessed water constraints and related management or adaption options (e.g. on energy prices and the supply-demand portfolio) are not modelled. More importantly, no monthly observational data are available, so that we can only compare annual averages between model and observations. Simulated summer-only (June–August) anomalies are generally larger than for the annual average, and show a similar relative distribution across countries, which suggests that the anomalies seen in the annual average are largely due to the EHWD (Supplementary Fig. 14).

**Human health.** The 2003 EHWD caused a high human death toll mainly due to circulatory- and respiratory-related causes, especially among the elderly population<sup>57</sup>. As many as 70,000 excess deaths were estimated for Europe, and 14,800 for France alone<sup>58,59</sup>. We apply a set of city-specific statistical models that describe a linear relationship between maximum daily apparent temperature ( $AT_{\max}$ ) and daily mortality (see Methods). We calculate  $AT_{\max}$  from daily mean and maximum temperatures and relative humidity, from the same observations-based climate data set used in the other sectors above, as well as two additional data sets (GSWP3 and PGFv2).

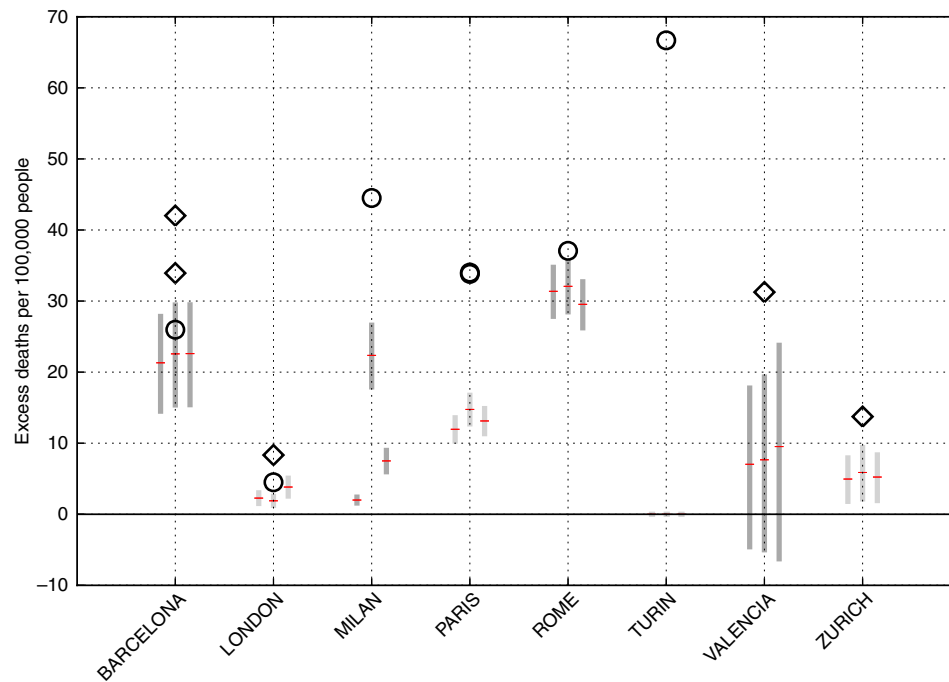
There is no comprehensive and consistent database of city-scale excess mortality to compare our estimates to. Instead, we compile results from earlier studies that specifically estimated the effect of the 2003 EHWD on natural mortality in European cities (Supplementary Table 3). These are based on actual mortality data reported for the 2003 summer e.g. by hospitals and city authorities, and are the closest available analogue to direct observations. In contrast, the models we applied<sup>60</sup> were trained on mortality and climate data from 1990–2000, and our estimates for 2003 are then derived based on the estimated model, reported baseline mortality rates, and the climate forcing.

We find that our statistical estimates are consistent with the range of literature values for Barcelona, London, and Rome, while they are substantially below the observed value(s) for Paris, Milan and Turin (Fig. 6). For Turin, our estimate is zero for all three climate forcing data sets, while an analysis<sup>61,62</sup> based on observed mortality found more than 60 excess deaths per 100,000 people occurred during the 2003 EHWD. For Valencia and Zurich, our models also underestimate observed values, although they come somewhat close when using the upper 95% confidence level for the slope of the temperature-mortality relation. Altogether, our models substantially underestimate the magnitude of the impact in 5 out of 8 cities for which excess mortality was observed. To our knowledge, this is the first independent evaluation of the performance of these models outside their calibration period.

Potential reasons for the discrepancies between our model estimates and observed mortality include compounding effects that the statistical models do not account for, including: the added effect of heat wave duration, the urban heat island effect, and air pollution (both with particulate matter and ozone); as well as the difference in spatial resolution between the climate data used to train the models (local airport weather stations) and the 0.5°-resolution forcing data sets (see Methods for an extended discussion).

Interestingly, while results are similar across the different climate forcing data sets for most of the cities, we find a strong dependence on the climate forcing for Milan. Unlike impacts on water resources or plant growth, which integrate the weather





**Fig. 6** City-specific estimates of the excess mortality attributable to the 2003 EHWD. Black symbols are observed estimates from the literature (Supplementary Table 3; note two very similar estimates for Paris). Circles denote studies that have reported both the number of excess deaths and the baseline population; diamonds denote studies which have only reported the number of excess deaths and where we have used the corresponding city population reported for 2003 in official statistics as a baseline. Grey bars and red lines are results from this study, for three different climate forcing datasets (left: GSWP3, middle: PGFv2, right: WFDEI). The bars indicate the results obtained by using the lower and upper 95% confidence intervals for the linear exposure response function slopes from ref. <sup>60</sup>, and the red line indicates the result obtained using the central estimate for the slope

effect over timespans on the order of weeks to months and over larger areas, excess mortality can result from just a few days of extreme heat in a particular location (city). Thus, small-scale deviations between the different climate forcing datasets or, for that matter, between a gridded reanalysis dataset and the local observations, can lead to large differences in the mortality outcome.

**Marine ecosystems.** The 2003 EHWD also produced anomalously high sea surface temperatures (SST) in the seas around Europe, most pronounced in the north-western Mediterranean Sea (Supplementary Fig. 15). This surface warming affected subsurface temperatures, vertical mixing, and currents, with potential effects on marine ecosystems and, by extension, on fisheries<sup>63</sup>. Indeed, several impacts, positive as well as negative, on local and regional marine ecosystems have been reported in the literature, including a die-off of *Posidonia oceanica* seagrass habitats<sup>64</sup> and rocky benthic communities<sup>65</sup>; and effects on the reproduction of some fish species like Bluefin tuna<sup>66</sup>, anchovy and round sardinella<sup>67</sup>. However, no direct impacts on commercial fisheries are reported in the literature due to the EHWD, despite evidence for longer-term effects of warming on fisheries landings<sup>68</sup>.

We simulated total consumer biomass at a 0.5° horizontal resolution with an ensemble of four marine ecosystem models<sup>69</sup>. As there is no widely accepted long-term database of fish biomass appropriate to the scales relevant to the EHWD, to gauge the impact of the event on marine ecosystems and fisheries, we compare relative change in reconstructed catch<sup>70</sup> to the relative change in simulated consumer biomass, taking the assumption that harvest is a reasonable indicator of biomass. We find no discernible impact of the 2003 EHWD on fish biomass in any of

the models, nor in the reconstructed catch data (Supplementary Fig. 16). In a few cases, models show a drop in 2004, which may be a lagged effect of the EHWD, but this is hard to verify in the absence of a longer time series, and not supported by the reconstructed catch data.

Nonetheless, the ocean reanalysis dataset used to force the marine ecosystem models does capture the 2003 anomalous warming in the Mediterranean and North Sea with a similar magnitude as in the observed data (Supplementary Fig. 17). It also shows a local maximum in plankton net primary productivity (NPP) in 2003 in the Mediterranean and the North Sea, and captures an observed phytoplankton abundance peak in the English Channel<sup>71</sup> (Supplementary Fig. 18). At the same time, simulated NPP exhibits large year-to-year variability, and the 2003 value is generally less exceptional in a long-term context than for temperature (Supplementary Fig. 19).

Thus, literature, data and models appear to agree that the 2003 EHWD affected physical ocean parameters as well as plants and planktonic NPP, but that the signal did not propagate to higher trophic levels on a large scale immediately. This may be because the event was too brief or too weak to cause a significant change in upper trophic levels. Another reason may be that temporary warming can act on fish biomass in two opposite ways: on the one hand, higher temperatures tend to increase the amount of food that fish need to maintain growth rates, as well as their mortality due to predation, disease or senescence. On the other hand, short-term warming increases phytoplankton growth, which has a positive effect on fish biomass. The net effect may thus be small. Further evaluation of marine ecosystem model simulations resolving individual species, and/or forced with higher-resolution climate data e.g. from regional climate models, might illuminate whether the models agree with the data for the right reasons; and longer simulation runs would be needed to account

for any lagged effects beyond 2004. Nonetheless, our results suggest that this modelling framework is applicable to extreme events like the 2003 EHWD in a similar fashion as in the terrestrial impact sectors.

## Discussion

Our study is unique in assessing the ability of several state-of-the-art impact models, for multiple sectors, to replicate the impacts from a recent, high-impact climatic extreme event, the 2003 EHWD. When impacts occur simultaneously in different sectors, it is important that models replicate not only the average variability of sectoral impact indicators, but the actual timing and magnitude of particular events. The rationale of our study was therefore to approach the assessment of model ability differently—not through a classical evaluation of impact statistics over several decades<sup>11,14,35</sup>—but instead to answer a what-if question: what if these models are used to predict the impacts of a single extreme event like the EHWD?

The answer, according to our results, is that the total impacts of such an event would likely be seriously underestimated. This is despite some encouraging results in individual sectors. For instance, the observed low water levels in the Rhine, Danube, and Aare rivers are matched very well by the median across our hydrological models, and the spread across models is moderate. Furthermore, in most countries where a substantial reduction in hydropower generation is observed, our hydropower model also simulates a substantial reduction in usable capacity. With respect to human health, our statistical models yield heat-related excess mortality rates of a similar magnitude seen in some observed estimates for Barcelona, London and Rome. We also find that generally the choice of climate forcing data set is not a major source of uncertainty.

Importantly though, severe impacts on agriculture and ecosystem productivity are underestimated by a large margin by most of the sector models. None or few of the crop models (depending on the climate forcing) reproduce the extremeness of the negative maize yield anomalies, of three and four standard deviations, respectively, in France and Italy, which are both among the largest maize producers in Europe. The crop model ensemble also systematically under-predicts the severity of the EHWD's impacts for wheat in Western Europe, implying that increasing temperatures and more frequent heatwaves due to climate change could have substantially more severe impacts in the region than typically assumed. Similar results are found for natural ecosystem primary productivity: only a few of the global vegetation models come close to the enormous negative anomaly seen in remote sensing-derived GPP in Western Europe, and some models even display a positive anomaly. For both the crop and GPP models, neither the multi-model mean nor any individual model reliably reproduce observations. The mortality models estimate only half of the observed rate of heat-related excess mortality in Paris, which saw by far the most fatalities of all cities.

Taking all sectors together, and across all countries, cities, regions and rivers studied, severe negative impacts (deviation of more than  $1\sigma$ , or more than 10 excess deaths per 100,000 people) were observed in 38 cases (Fig. 1). In 20 of them, the severity of impacts is underestimated by the respective models, counting in those countries where 75% or more, but not all, of the crop models underestimate yield impacts. Leaving out those countries, we are left with 14 cases where the entirety of models underestimates the observed impacts. This summary statistic is necessarily crude, but highlights the magnitude of the problem.

While many of the largest impacts of the EHWD are thus underestimated by the sectoral model ensembles, we also find a number of “false alarms”, mainly for water resources and hydropower. Hydrological models show flow reductions of up to

$2\sigma$  in the Weser, Elbe and Oder rivers, where no substantial anomalies were observed. The hydropower model shows large reductions in usable capacity in Switzerland, Poland and Germany, while reported hydropower generation was near average in all these countries. On a regional level, the underestimation of impacts across sectors is particularly evident in southern Europe; while in northern and eastern Europe, impacts on crop yields and river flow are matched relatively well, and some hydropower impacts are even greatly overestimated (Fig. 1).

Whilst our study is the most comprehensive systematic comparison to date between simulated and observed impacts for an extreme climatic event, it is not exhaustive in the treatment of different sectors, e.g. we did not include livestock production, vector-borne diseases or forest fires<sup>18</sup>. Assembling an even larger multi-disciplinary team could have partly addressed this. For some sectors suitable observational data for model comparison were unavailable (nuclear power generation). In all sectors, we have focused on large-scale impact indicators—e.g. average monthly discharge or country-level crop yield—which are close to what the models are designed to represent, and which we assume to reflect the total societal costs and key environmental impacts of the 2003 EHWD to a large extent. Smaller-scale characteristics of some of the impacts (e.g. minimum daily river flow or maximum daily excess mortality) could have important implications not captured by the more aggregate indicators, for example the overburdening of critical infrastructure for a short time. Representing such processes accurately will be a substantial challenge for the relatively large-scale models applied here.

We have investigated a single climate event, but our results are relevant beyond the case of the 2003 EHWD. In 2015 and 2018, central Europe was again hit by extremely hot and dry summer conditions, and projected trends in temperature and precipitation suggest that such events will likely occur more frequently in the future<sup>72</sup>. While each event is unique, similar events produce similar biophysical impacts: future drought-heat wave events will again lead to low river flows, reduce ecosystem productivity, and damage crop yields. As long as the model shortcomings exposed by our analysis persist, they will likely lead to biased estimates of extreme event impacts.

These shortcomings are related to the representation of both natural processes and human management in the models. For instance, human adaptation (or the lack thereof) can affect the severity of impacts for a given extreme event<sup>73</sup>. Indeed, excess human mortality during the 2006 heatwave in Europe was much lower than 2003, possibly because of better preparedness<sup>57</sup>. This might go some way towards explaining the underestimation of observed impacts by the heat-related mortality models: the 2003 EHWD was outside the climatological range that the models were trained on, and the range that the health system was prepared for. In process-based models, the capacity to respond to extreme climate conditions can be overestimated if resource constraints are not accounted for. An example is the lack of constraints on irrigation water in our crop model simulations. For instance, streamflow in the Vienne basin, which contains large shares of irrigated maize cropland<sup>74</sup>, was anomalously low in summer 2003 (Fig. 2); and indeed, the maize yield anomaly in France is simulated more realistically when irrigation is switched off, than in the default model setup with irrigation (Supplementary Fig. 9). Accounting for such cross-sectoral interactions<sup>75</sup> and reflecting changing levels of societal adaptation in models will be critical for more accurate impact and damage estimates in the future.

Important natural processes such as heat stress effects on ecosystem photosynthesis are also missing in some models, while other processes such as potential evapotranspiration are parameterised through empirical relationships that may not hold outside past ranges of climatic variability<sup>36</sup>. Such shortcomings

will affect impact assessments of other climatic events, too. In addition, there are known issues that have not emerged in our study, but may prove problematic when assessing the impacts of extreme events of different types, or outside Europe. For instance, the global ecosystem models do not consider the potentially amplifying interactions between climate change and forest disturbances, such as wind storms<sup>76</sup>; and some global hydrological models struggle to capture the magnitude and timing of processes such as transmission loss and snowmelt accurately<sup>35</sup>. Finally, our models may perform worse in parts of the developing world where constraints from observational data are more limited.

Our results emphasise the importance of considering the uncertainties of current biophysical impact estimates. Modelling studies of future climate change impacts on mean conditions in different sectors<sup>1,2,77,78</sup> have shown that the uncertainties of biophysical impact estimates are of a similar magnitude as climate modelling uncertainty. A common practice has since been to use multi-model ensembles rather than single impact models, as is standard in climate modelling. We have shown that for an extreme event, even an entire ensemble of impact models can sometimes completely miss observations. This urgently calls for a shift in the agenda of model development and evaluation away from mean conditions towards extremes. In the meantime, a precautionary interpretation of impact estimates is recommended, where the most pessimistic model in an ensemble must be considered just as plausible as those in the centre of the distribution.

For quantifying the ultimate impacts of climate change on different aspects of the economy, or parts of society, different metrics need to be applied beyond the (mainly) biophysical indicators studied here; e.g., macro-economic damages, public health burdens, or effects on societal equality and political stability. However, any such analysis needs to rest on a firm assessment of the biophysical impact, and will be unreliable if the biophysical impact is not captured accurately. Our study thus provides important context to any subsequent analysis of economic and societal impacts.

## Methods

**Climate forcing data.** The primary climate forcing dataset used in this study is the latest WATCH meteorological forcing data set (WFDEI)<sup>79</sup>, which is based on ERA-Interim reanalysis data. Two other data sets are used for comparison, for some sectors: the Global Soil Wetness Project phase 3 forcing data set (GSWP3; <http://hydro.iis.u-tokyo.ac.jp/GSWP3/>)<sup>80</sup>, based on the 20th century reanalysis (20CR); and the Princeton Global Forcing data set version 2 (PGFv2)<sup>81</sup>, based on NCEP/NCAR Reanalysis. The magnitude and spatio-temporal pattern of the 2003 summer temperature and precipitation anomaly is similar in all three data sets (Supplementary Figs. 1 and 2), and in agreement with other reanalysis products<sup>17,18</sup>.

**Model ensemble.** Our multi-sectoral ensemble of impact models (Supplementary Table 1) includes 7 global hydrological models, 12 global gridded crop models, and 8 global vegetation models. In addition, we analysed the impacts on hydropower production as well as heat-related mortality, using single models; and potential impacts on fisheries using four marine ecosystem models. All terrestrial model simulations were driven by WFDEI climate forcing for the period 1979–2010, and followed a common protocol (ISIMIP2a, [www.isimip.org](http://www.isimip.org)), allowing us to compare models within a sector (such as agriculture) and to apply a common method of analysis across sectors. For some sectors, additional simulations were conducted, driven with PGFv2 or GSWP3 climate forcing. In those cases, ensembles are slightly different for the different climate forcings because not all models were run with all climate data sets. In Figs. 2–4, as well as the corresponding supplementary figures, models are uniquely identified by numerals (see Supplementary Table 1).

In contrast to the terrestrial models, the marine ecosystem models need not only atmospheric but also oceanic forcing variables, and are therefore driven by a reanalysis-based climate model run. While thus not completely consistent with the other sectors, these simulations demonstrate how our analysis framework can be expanded to the marine realm.

Importantly, the model simulations were not specifically designed to capture a particular event. Models were set up for best overall performance, and parameters were kept constant throughout the simulation period, with climate as the only source of annual-scale variations (except that the hydrological simulations account

for the construction of dams and reservoirs over time). This is similar to simulations under future climate change scenarios, where e.g. agricultural management choices in any particular year are not known. Therefore, if models are found to accurately reproduce the impacts of the 2003 EHWD using the present set-up, this increases confidence in their ability to quantify the impacts of future extreme events represented in climate projections.

**Impact metric.** In all sectors except human health, we measure impacts as a given variable's deviation in 2003 from the historical average, adjusted for long-term trends, and normalised by the historical standard deviation (see below for more details). That is, we measure the extremeness of the event in a given variable, not the absolute magnitude of change in that variable. If a model correctly simulated the absolute magnitude of change in a variable in 2003, but simulated the same magnitude of change every other year, then the agreement in 2003 may not be for the right reasons, and there would be little confidence in the model's response to extreme events in future climate scenarios. In the paper we use the term "impact" to refer to the relative deviation, or extremeness, of an indicator, unless otherwise stated. An exception is heat-related excess mortality, which is already defined as a deviation from the expected baseline mortality.

**Water resources.** Monthly discharge data was obtained from the Global Runoff Data Centre (GRDC, [http://www.bafg.de/GRDC/EN/02\\_srvcs/21\\_tmrs/riverdischarge\\_node.html](http://www.bafg.de/GRDC/EN/02_srvcs/21_tmrs/riverdischarge_node.html)) for gauging stations along major rivers in the geographic region that was affected by the 2003 EHWD. Only stations with catchments of ~10,000 km<sup>2</sup> (equivalent to ~4 model grid cells) or larger were selected. Stations also had to have continuous monthly data available from GRDC for the whole period 1979–2008. Simulated daily discharge data was first averaged to monthly values. Then, for both simulated and reported discharge, a single month of interest was extracted (e.g., August), resulting in a yearly series of values for that month over 1979–2008. From this series, a linear trend was removed, and the standard deviation calculated. The 2003 relative discharge anomaly was then calculated as the difference between 2003 and the 1979–2008 average, divided by the standard deviation (with 2003 excluded from the calculation of the standard deviation). We note that none of the models consider changes in glacial meltwater runoff, which was an important factor during the 2003 EHWD in Alpine catchments<sup>26</sup>.

**Agriculture.** We analyse crop model simulations covering the period 1980–2009 under the "default" setting of the ISIMIP2a simulation protocol, where modelling teams individually determined how sowing dates, harvesting dates, fertiliser application rates and crop varieties are parameterised in their model, with the objective of best overall performance for the simulation period. Total crop yield in each grid cell was calculated as the weighted sum of irrigated and rainfed yields, with the weights being the respective cell fractions under irrigated and rainfed cultivation around the year 2000, as reported in the MIRCA2000 data set<sup>82</sup>. The 2003 crop yield anomaly was calculated from the simulated and observed series of annual harvests, analogously to the discharge anomaly, as described above; except that rather than a linear trend, a 7-year moving average was removed from the FAOSTAT reported crop yields, in order to account for non-linear changes induced e.g. by technological and management changes. Such factors are absent in the crop model simulations, but we applied the same de-trending method to the simulation data for consistency. The size of the moving average window was decreased at the edges to preserve the length of the time series. Alternative results with PGFv2 climate forcing and with linear de-trending are shown in Supplementary Figs. 8 and 20, respectively. There is some ambiguity in assigning wheat harvests to calendar years, but our results are robust against shifting the relevant time series by one year (Supplementary Methods).

**Terrestrial ecosystems.** The regions used for analysis are West (7°W–7°E, 42°–50°N) and Central (0°–20°E, 42°–53°N). Unlike in some other sectors, the observational time series for GPP is relatively short—the MODIS data set extends from 2000–2011. For a consistent model-data comparison, the models' GPP anomaly was therefore calculated in terms of the standard deviation over 2000–2010, after removing a linear trend over the period 1979–2010. We note that the standard deviation over this short period, and thus the GPP anomaly metric, may be rather sensitive to inclusion of additional years. Calculating the standard deviation of the models' GPP anomaly over the entire simulated period 1971–2010, rather than just 2000–2010, has little effect on the models' relative distribution, but contracts the spread across models, as can be expected (Supplementary Fig. 21). Note that the conversion of the MODIS space-borne measurements of the fraction of absorbed photosynthetically active radiation (FAPAR) into GPP involves additional uncertainties. To obtain GPP from FAPAR, a light use efficiency (LUE) factor is applied. The MODIS LUE model is based on simple functions of climatic data (i.e., temperature and vapour pressure deficit) calibrated only at a few sites<sup>83</sup>. The fact that the MODIS LUE model ignores soil moisture variations and CO<sub>2</sub> physiological effects makes it prone to GPP biases, especially for drought. On the other hand, the negative 2003 summer anomaly found in MODIS GPP is consistent with eddy covariance measurements of carbon fluxes at different sites across Europe<sup>45</sup>.

Potential biases in absolute GPP values are also circumvented by our use of standard deviations as unit of comparison.

**Energy.** Hydropower: The anomalies in observed hydropower generation and simulated usable capacity were calculated analogously to the discharge anomaly, as described above; except that instead of a linear trend, a constant mean value was subtracted, since there were no substantial trends. Annual country-level hydropower generation data was obtained from the EIA database (<https://www.eia.gov/beta/international/data/browser/>). For some countries, EIA data is limited to the time since their independence; the shortest series (starting 1993) is available for Slovakia and the Czech Republic. Hydropower usable capacity was simulated for the period 1981–2010 based on plant level regulated streamflow simulations from the VIC hydrological model forced with WFDEI climate, hydraulic head, total efficiency of the power generating unit, density of freshwater and the gravitational acceleration (see supplementary information of ref. 56). Simulations were performed for hydropower plants in Europe using data from the World Electric Power Plant Database<sup>84</sup> and simulated usable capacity was then aggregated to country level. The hydropower usable capacity model<sup>56</sup> simulates the physical impacts of changes in water resources on usable power plant capacity. Economic feedbacks of the assessed water constraints and related adaptation options (e.g. on energy prices, supply demands portfolio) are not modelled.

Wind and solar power: In addition to modelling hydropower impacts, we assessed the anomalies in wind power potential capacity factors assuming a homogeneous capacity installed. This assumes the use of a single wind turbine and extrapolation of wind at hub height using the methodology of ref. 85. We found a record low capacity factor in several Western European countries in summer (June–August) 2003, relative to the 1979–2014 average, using the WFDEI climate data set. In France the anomaly was –39%. While installed wind power capacity was very low in 2003, these results suggest that comparable future climatic events, in a world with much more installed wind power capacity, would induce large negative anomalies in wind power generation. In contrast, we found a positive short-wave radiation anomaly of 5–20% in Western European countries, which would have been relevant for solar power generation. The consequence of all these anomalies would therefore have been a very unusual pattern of energy systems impacts from weather, a situation that transmission operators have to manage in order to balance supply and demand.

**Human health.** Models: We calculate excess mortality attributable to heat for 8 cities for which comparable estimates of observed mortality were found in the literature. Our calculations are based on previously published models called epidemiological exposure-response functions (ERF). These ERFs were originally derived from a set of non-linear curves developed from a flexible parametric approach<sup>60</sup>, and summarised as the percentage change in daily mortality ( $b$ ) per degree increase in maximum daily apparent temperature ( $AT_{max}$ ) above a threshold. Based on a relative risk framework<sup>86</sup>, daily excess mortality attributable to heat ( $D_{attr}$ ) can then be derived from the ERFs as follows:

$$D_{attr} = D_{obs} (1 - e^{-b\Delta AT_{max}})$$

where  $D_{obs}$  is the observed average daily mortality in the warm season (as reported in ref. 60). We sum  $D_{attr}$  over all days in June to August 2003 on which  $AT_{max}$  was above the threshold ( $\Delta AT_{max} > 0$ ). A similar approach has previously been used in climate change impact assessments by refs. 73 and 87. Here we use the central estimate and the upper and lower 95% confidence intervals of the slope ( $b$ ) reported for each city in ref. 60.

The ERFs used here have been applied for modelling adaptation to heat stress under future climate change<sup>73</sup>, but here we assume no adaptation, i.e. the ERFs remain unchanged between the reference period 1990–2000 and 2003.  $AT_{max}$  is calculated from daily mean and maximum temperatures and relative humidity, following Tetens' equation<sup>88</sup>. As in ref. 60, we use 3-day averages of  $AT_{max}$  in all of our analyses to account for possible lag-effects.

In contrast to the sectors above, excess mortality results are presented not in terms of standard deviations from the long-term average, but in terms of a city's total population. This is because excess mortality is already defined as an excursion from a baseline mortality rate, and is not a continuous variable that could be assumed as approximately normally distributed.

Caveats: We note that the studies we use for evaluating the health models (Supplementary Table 3) differ somewhat in methodology among each other as well as compared to our approach. For instance, while most studies<sup>61,62,87,89–92</sup>—including ours—analyse the whole summer (June–August), two studies<sup>93,94</sup> considered only a shorter period during which the actual heat wave occurred in the respective city. However, since excess mortality occurred mainly during the heat wave days, this difference in approach is expected to have only a minimal effect on the magnitude of the difference between modelled and reported mortality estimates. In addition, our approach only considers the effects of hot temperatures on mortality on each day independent from other days. The models do not account for the potential compounding effects on mortality of heat-wave duration<sup>95,96</sup>, nor do they model mortality displacement<sup>97</sup>.

Another important caveat is that the ERFs were trained on local weather data obtained from airport weather stations<sup>60</sup>, but the outputs from the ERFs are

computed by using data from a spatially corresponding 0.5° climate dataset-grid cell as input. To the extent that extreme values may be attenuated at coarser spatial scales, this may bias our results towards lower mortality. We note that this limitation exists in other sectors, too (e.g. when crop models are calibrated on site-level weather variables but forced with grid-level climate data), and for a given location it may be alleviated through bias-correction techniques. However, spatially distributed projections of future climate are typically available on a grid level only, and thus our methodology reflects how the impacts models are typically applied to project future climate change impacts<sup>98,99</sup>. To reiterate, we do not assess how models perform under ideal conditions for a specific event or location, but how well the model and data framework presently available for climate impacts simulations performs for extreme events like the 2003 EHWD.

The ERFs and input climate data do not account specifically for the urban heat island effect, which has been shown to increase local temperature, and in turn, mortality rates<sup>100</sup>. Including this effect would therefore increase our modelled estimates for mortality. However, we do not attempt to include it because the ERFs were trained using local airport weather data and not city-centre weather data<sup>60</sup>. The use of the former for training the ERFs means that the mortality estimates from the ERFs are actually representative of temperatures outside the urban heat island, so it would be inconsistent to add an urban heat island effect to the temperatures that are input to the ERFs. This limitation might explain why our estimates are much lower than the literature values for some cities, as might the potential attenuation of extreme temperatures in the coarse-resolution climate forcing data, as mentioned above. Another potential reason for the discrepancies is that the heat effects on summer mortality in 2003 were probably compounded by high ozone concentrations and resulting poor air quality<sup>18</sup>, whereas air quality effects have been removed in the derivation of the ERFs<sup>60</sup>.

**Marine ecosystems.** Within ISIMIP2a, an ensemble of global and regional marine ecosystem models was run with forcing from a global coupled ocean circulation and planktonic ecosystem model<sup>101</sup>, which in turn was forced with an atmospheric reanalysis dataset<sup>102,103</sup>. This set-up is used as an analogue to the observational forcing of the terrestrial impacts models, given that three-dimensional oceanic observations or reanalysis data are not available. Marine ecosystem models were run without fishing, in order to capture only climate effects on fish biomass. No ocean acidification was assumed.

**Code availability.** Data processing scripts are available from the corresponding author upon request. Availability of impact model code varies by model and is indicated in Supplementary Table 1.

## Data availability

Simulation data from gridded impact models is available through <https://esg-pik-potsdam.de/projects/isimip2a/> and citable using the following DOIs: <https://doi.org/10.5880/PIK.2017.002> (terrestrial ecosystems); <https://doi.org/10.5880/PIK.2017.006> (agriculture); <https://doi.org/10.5880/PIK.2017.010> (hydrology); <https://doi.org/10.5880/PIK.2018.004> (marine ecosystems, regional); <https://doi.org/10.5880/PIK.2018.005> (marine ecosystems, global). The city-level mortality model data and country-level hydropower model data are available from the authors upon request.

Received: 9 January 2018 Accepted: 28 January 2019

Published online: 01 March 2019

## References

- Rosenzweig, C. et al. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl Acad. Sci. USA* **14**, 1–6 (2014).
- Schewe, J. et al. Multimodel assessment of water scarcity under climate change. *Proc. Natl Acad. Sci. USA* **111**, 3245–3250 (2014).
- van Vliet, M. T. H. et al. Multi-model assessment of global hydropower and cooling water discharge potential under climate change. *Glob. Environ. Chang* **40**, 156–170 (2016).
- Whitmee, S. et al. Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health. *Lancet* **386**, 1973–2028 (2015).
- Pecl, G. T. et al. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science (80-)* **355**, eaai9214 (2017).
- Pindyck, R. S. Climate change policy: what do the models tell us? *J. Econ. Lit.* **51**, 860–872 (2013).
- Hsiang, S. et al. Estimating economic damage from climate change in the United States. *Science (80-)* **356**, 1362–1369 (2017).
- Monier, E. et al. Toward a consistent modeling framework to assess multi-sectoral climate impacts. *Nat. Commun.* **9**, 660 (2018).

9. Harrison, P. A. et al. Differences between low-end and high-end climate change impacts in Europe across multiple sectors. *Reg. Environ. Chang.* <https://doi.org/10.1007/s10113-018-1352-4> (2018).
10. Franzke, C. L. E. Impacts of a changing climate on economic damages and insurance. *Econ. Disasters Clim. Chang.* **1**, 95–110 (2017).
11. Müller, C. et al. Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geosci. Model Dev.* **10**, 1403–1422 (2017).
12. Beck, H. E. et al. Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.* **21**, 2881–2903 (2017).
13. Hattermann, F. F. et al. Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. *Clim. Change* **141**, 561–576 (2017).
14. Traore, A. K. et al. Evaluation of the ORCHIDEE ecosystem model over Africa against 25 years of satellite-based water and carbon measurements. *J. Geophys. Res. Biogeosci.* **119**, 1554–1575 (2014).
15. Schauburger, B. et al. Consistent negative response of US crops to high temperatures in observations and crop models. *Nat. Commun.* **8**, 13931 (2017).
16. Dosio, A., Mentaschi, L., Fischer, E. M. & Wyser, K. Extreme heat waves under 1.5 °C and 2 °C global warming. *Environ. Res. Lett.* **13**, 054006 (2018).
17. Schär, C. et al. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–336 (2004).
18. García-Herrera, R., Díaz, J., Trigo, R. M., Luterbacher, J. & Fischer, E. M. A review of the European summer heat wave of 2003. *Crit. Rev. Environ. Sci. Technol.* **40**, 267–306 (2010).
19. Seneviratne, S. I. et al. Swiss prealpine Rietholzloch research catchment and lysimeter: 32 year time series and 2003 drought event. *Water Resour. Res.* **48**, W06526 (2012).
20. Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D. & Schär, C. Soil moisture–atmosphere interactions during the 2003 European summer heat wave. *J. Clim.* **20**, 5081–5099 (2007).
21. Beniston, M. The 2003 heat wave in Europe: a shape of things to come? An analysis based on Swiss climatological data and model simulations. *Geophys. Res. Lett.* **31**, L02202 (2004).
22. Stott, P. A., Stone, D. A. & Allen, M. R. Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).
23. Zhao, C. et al. Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl Acad. Sci. USA* **114**, 9326–9331 (2017).
24. Bador, M. et al. Future summer mega-heatwave and record-breaking temperatures in a warmer France climate. *Environ. Res. Lett.* **12**, 074025 (2017).
25. Whan, K. et al. Impact of soil moisture on extreme maximum temperatures in Europe. *Weather Clim. Extrem.* **9**, 57–67 (2015).
26. Zappa, M. & Kan, C. Extreme heat and runoff extremes in the Swiss Alps. *Nat. Hazards Earth Syst. Sci.* **7**, 375–389 (2007).
27. Hannaford, J., Lloyd-Hughes, B., Keef, C., Parry, S. & Prudhomme, C. Examining the large-scale spatial coherence of European drought using regional indicators of precipitation and streamflow deficit. *Hydrol. Process.* **25**, 1146–1162 (2011).
28. Gudmundsson, L. & Seneviratne, S. I. Observation-based gridded runoff estimates for Europe (E-RUN version 1.1). *Earth Syst. Sci. Data* **8**, 279–295 (2016).
29. Oki, T. & Kanae, S. Global hydrological cycles and world water resources. *Science* **313**, 1068–1072 (2006).
30. Fink, A. H. et al. The 2003 European summer heatwaves and drought –synoptic diagnosis and impacts. *Weather* **59**, 209–216 (2004).
31. Umweltbundesamt. *Monitoringbericht 2015 zur Deutschen Anpassungsstrategie an den Klimawandel* (Dessau-Roßlau, Germany) <http://www.umweltbundesamt.de/publikationen/monitoringbericht-2015> (2015).
32. Andersen, O. B., Seneviratne, S. I., Hinderer, J. & Viterbo, P. GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophys. Res. Lett.* **32**, L18405 (2005).
33. Haddeland, I. et al. Global water resources affected by human interventions and climate change. *Proc. Natl Acad. Sci. USA* **111**, 3251–3256 (2013).
34. Wada, Y., van Beek, L. P. H., Wanders, N. & Bierkens, M. F. P. Human water consumption intensifies hydrological drought worldwide. *Environ. Res. Lett.* **8**, 034036 (2013).
35. Zaherpour, J. et al. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environ. Res. Lett.* **13**, 065015 (2018).
36. Milly, P. C. D. & Dunne, K. A. A hydrologic drying bias in water-resource impact analyses of anthropogenic climate change. *JAWRA*, **53**, 822–838 (2017).
37. van der Velde, M., Tubiello, F. N., Vrieling, A. & Bouraoui, F. Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. *Clim. Change* **113**, 751–765 (2012).
38. Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E. (Eds) in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. 843–868 (Cambridge University Press, Cambridge, 2007).
39. Ciais, P. et al. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* **437**, 529–533 (2005).
40. Copa-Cogeca. *Assessment of the impact of the heat wave and drought of the summer 2003 on agriculture and forestry* [http://docs.gip-ecofor.org/libre/COPA\\_COGECA\\_2004.pdf](http://docs.gip-ecofor.org/libre/COPA_COGECA_2004.pdf) (2004).
41. van Bussel, L. G. J., Stehfest, E., Siebert, S., Müller, C. & Ewert, F. Simulation of the phenological development of wheat and maize at the global scale. *Glob. Ecol. Biogeogr.* **24**, 1018–1029 (2015).
42. Smith, P. C. et al. European-wide simulations of croplands using an improved terrestrial biosphere model: 2. Interannual yields and anomalous CO<sub>2</sub> fluxes in 2003. *J. Geophys. Res.* **115**(G4), G04028 (2010).
43. Vetter, M. et al. Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using seven models. *Biogeosciences* **5**, 561–583 (2008).
44. Zaitchik, B. F., Macalady, A. K., Bonneau, L. R. & Smith, R. B. Europe’s 2003 heat wave: a satellite view of impacts and land–atmosphere feedbacks. *Int. J. Climatol.* **26**, 743–769 (2006).
45. Reichstein, M. et al. Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Glob. Chang. Biol.* **13**, 634–651 (2007).
46. Jolly, W. M., Dobbertin, M., Zimmermann, N. E. & Reichstein, M. Divergent vegetation growth responses to the 2003 heat wave in the Swiss Alps. *Geophys. Res. Lett.* **32**, L18409 (2005).
47. Running, S. W. et al. A continuous satellite-derived measure of global terrestrial primary production. *Bioscience* **54**, 547–560 (2004).
48. Zhao, M., Heinsch, F. A., Nemani, R. R. & Running, S. W. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.* **95**, 164–176 (2005).
49. Ito, A. et al. Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: benchmarking for impact assessment studies. *Environ. Res. Lett.* **12**, 085001 (2017).
50. Huang, Y., Gerber, S., Huang, T. & Lichstein, J. W. Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data. *Glob. Biogeochem. Cycles* **30**, 1827–1846 (2016).
51. Van Oijen, M. et al. Impact of droughts on the C-cycle in European vegetation: a probabilistic risk analysis using six vegetation models. *Biogeosci. Discuss.* **11**, 8325–8371 (2014).
52. Chen, M. et al. Regional contribution to variability and trends of global gross primary productivity. *Environ. Res. Lett.* **12**, 105005 (2017).
53. De Kauwe, M. G. et al. Do land surface models need to include differential plant species responses to drought? Examining model predictions across a mesic-xeric gradient in Europe. *Biogeosciences* **12**, 7503–7518 (2015).
54. IAEA. *Operating experience with nuclear power stations in Member States in 2003*. [www-pub.iaea.org/MTCD/Publications/PDF/Pub1219\\_web.pdf](http://www-pub.iaea.org/MTCD/Publications/PDF/Pub1219_web.pdf) (IAEA, Vienna, 2004).
55. Rübhelke, D. & Vögele, S. Impacts of climate change on European critical infrastructures: the case of the power sector. *Environ. Sci. Policy*, **14**, 53–63 (2011).
56. van Vliet, M. T. H., Wiberg, D., Leduc, S. & Riahi, K. Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nat. Clim. Chang.* **6**, 375–380 (2016).
57. Murray, V. et al. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) 487–542 (Cambridge University Press, Cambridge and New York, 2012).
58. Robine, J.-M. et al. Death toll exceeded 70,000 in Europe during the summer of 2003. *C. R. Biol.* **331**, 171–178 (2008).
59. Pirard, P. et al. Summary of the mortality impact assessment of the 2003 heat wave in France. *Euro. Surveill.* **10**, 153–156 (2005).
60. Baccini, M. et al. Heat effects on mortality in 15 European cities. *Epidemiology* **19**, 711–719 (2008).
61. Michelozzi, P. et al. The impact of the summer 2003 heat waves on mortality in four Italian cities. *Eur. Surveill.* **10**, 161–165 (2005).
62. Michelozzi, P. et al. Temperature and summer mortality: geographical and temporal variations in four Italian cities. *J. Epidemiol. Community Health* **60**, 417–423 (2006).
63. Olita, A. et al. Effects of the 2003 European heatwave on the Central Mediterranean Sea: surface fluxes and the dynamical response. *Ocean Sci.* **3**, 273–289 (2007).
64. Marbà, N. & Duarte, C. M. Mediterranean warming triggers seagrass (*Posidonia oceanica*) shoot mortality. *Glob. Chang. Biol.* **16**, 2366–2375 (2009).

65. Garrabou, J. et al. Mass mortality in Northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave. *Glob. Chang. Biol.* **15**, 1090–1103 (2009).
66. García, A. et al. Climate-induced environmental conditions influencing interannual variability of Mediterranean bluefin (*Thunnus thynnus*) larval growth. *Fish. Oceanogr.* **22**, 273–287 (2013).
67. Maynou, F., Sabatés, A. & Salat, J. Clues from the recent past to assess recruitment of Mediterranean small pelagic fishes under sea warming scenarios. *Clim. Change* **126**, 175–188 (2014).
68. Tzanatos, E., Raitso, D. E., Triantafyllou, G., Somarakis, S. & Tsonis, A. A. Indications of a climate effect on Mediterranean fisheries. *Clim. Change* **122**, 41–54 (2014).
69. Tittensor, D. P. et al. A protocol for the intercomparison of marine fishery and ecosystem models: Fish-MIP v1.0. *Geosci. Model Dev.* **11**, 1421–1442 (2018).
70. Pauly, D. & Zeller, D. Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nat. Commun.* **7**, 10244 (2016).
71. Gómez, F. & Souissi, S. The impact of the 2003 summer heat wave and the 2005 late cold wave on the phytoplankton in the north-eastern English Channel. *C. R. Biol.* **331**, 678–685 (2008).
72. Orth, R. et al. Record dry summer in 2015 challenges precipitation projections in Central Europe. *Sci. Rep.* **6**, 28334 (2016).
73. Gosling, S. N. et al. Adaptation to climate change: a comparative analysis of modeling methods for heat-related mortality. *Environ. Health Perspect.* **125**, 1–45 (2017).
74. van der Velde, M., Wriedt, G. & Bouraoui, F. Estimating irrigation use and effects on maize yield during the 2003 heatwave in France. *Agric. Ecosyst. Environ.* **135**, 90–97 (2010).
75. Frieler, K. et al. A framework for the cross-sectoral integration of multi-model impact projections: land use decisions under climate impacts uncertainties. *Earth Syst. Dyn.* **6**, 447–460 (2015).
76. Rey, C. P. O. et al. Are forest disturbances amplifying or canceling out climate change-induced productivity changes in European forests? *Environ. Res. Lett.* **12**, 034027 (2017).
77. Caminade, C. et al. Impact of climate change on global malaria distribution. *Proc. Natl Acad. Sci. USA* **111**, 1–6 (2014).
78. Friend, A. D. et al. Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO<sub>2</sub>. *Proc. Natl Acad. Sci. USA* **111**, 3280–3285 (2014).
79. Weedon, G. P. et al. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.* **50**, 7505–7514 (2014).
80. van den Hurk, B. et al. LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome. *Geosci. Model Dev.* **9**, 2809–2832 (2016).
81. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111 (2006).
82. Portmann, F. T., Siebert, S. & Döll, P. MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. *Glob. Biogeochem. Cycles* **24**, GB1011 (2010).
83. Running, S. W. & Zhao, M. *User's Guide: Daily GPP and annual NPP (MOD17A2/A3) products NASA Earth Observing System MODIS land algorithm*. Version 3.0. [https://www.ntsg.umt.edu/files/modis/MOD17UsersGuide2015\\_v3.pdf](https://www.ntsg.umt.edu/files/modis/MOD17UsersGuide2015_v3.pdf) (2015).
84. UDI. *World electric power plants database* <http://www.platts.com/> (2013).
85. Tobin, I. et al. Assessing climate change impacts on European wind energy from ENSEMBLES high-resolution climate projections. *Clim. Change* **128**, 99–112 (2015).
86. Gasparrini, A. & Leone, M. Attributable risk from distributed lag models. *Bmc Med. Res. Methodol.* **14**, 55 (2014).
87. Mitchell, D. et al. Attributing human mortality during extreme heat waves to anthropogenic climate change. *Environ. Res Lett.* **11**, 074006 (2016).
88. Monteith, J. L. & Unsworth, M. *Principles of Environmental Physics*. (Elsevier Science, San Diego, 2007).
89. Borrell, C. et al. Socioeconomic position and excess mortality during the heat wave of 2003 in Barcelona. *Eur. J. Epidemiol.* **21**, 633–640 (2006).
90. Grize, L., Huss, A., Thommen, O., Schindler, C. & Braun-Fahrlander, C. Heat wave 2003 and mortality in Switzerland. *Swiss Med. Wkly* **135**, 200–205 (2005).
91. Martínez Navarro, F., Simón-Soria, F. & López-Abente, G. Valoración del impacto de la ola de calor del verano de 2003 sobre la mortalidad. *Gac. Sanit.* **18**(Supl.1), 250–258 (2004).
92. Tobías, A. et al. Short-term effects of extreme hot summer temperatures on total daily mortality in Barcelona, Spain. *Int. J. Biometeorol.* **54**, 115–117 (2010).
93. Le Tertre, A. et al. Impact of the 2003 heatwave on all-cause mortality in 9 french cities. *Epidemiology* **17**, 75–79 (2006).
94. Johnson, H. et al. The impact of the 2003 heat wave on mortality and hospital admissions in England. *Health Stat. Q* **25**, 6–11 (2005).
95. Hajat, S. et al. Impact of high temperatures on mortality. *Epidemiology* **17**, 632–638 (2006).
96. Gasparrini, A. & Armstrong, B. The impact of heat waves on mortality. *Epidemiology* **22**, 68–73 (2011).
97. Gosling, S. N., Lowe, J. A., McGregor, G. R., Pelling, M. & Malamud, B. D. Associations between elevated atmospheric temperature and human mortality: a critical review of the literature. *Clim. Change* **92**, 299–341 (2009).
98. Gosling, S. N., McGregor, G. R. & Lowe, J. A. The benefits of quantifying climate model uncertainty in climate change impacts assessment: an example with heat-related mortality change estimates. *Clim. Change* **112**, 217–231 (2012).
99. Kingsley, S. L., Eliot, M. N., Gold, J., Vanderslice, R. R. & Wellenius, G. A. Current and projected heat-related morbidity and mortality in Rhode Island. *Environ. Health Perspect.* **124**, 460–467 (2015).
100. Smargiassi, A. et al. Variation of daily warm season mortality as a function of micro-urban heat islands. *J. Epidemiol. Community Health* **63**, 659–664 (2009).
101. Stock, C. A., Dunne, J. P. & John, J. G. Global-scale carbon and energy flows through the marine planktonic food web: an analysis with a coupled physical–biological model. *Prog. Oceanogr.* **120**, 1–28 (2014).
102. Griffies, S. M. et al. Coordinated ocean-ice reference experiments (COREs). *Ocean Model.* **26**, 1–46 (2009).
103. Large, W. G. & Yeager, S. G. *Diurnal to Decadal Global Forcing For Ocean and Sea-Ice Models: The Data Sets and Flux Climatologies*. NCAR Technical Note NCAR/TN-460+STR (National Center for Atmospheric Research, Boulder, 2004).

## Acknowledgements

The work was supported within the framework of the Leibniz Competition (SAW-2013-PIK-5), the EU FP7 project HELIX (grant no. FP7-603864-2), the FP7 project IMPACT2C (grant agreement#282746) and by the German Federal Ministry of Education and Research (BMBF, grant no. 01LS1201A1). H.K. acknowledges support by Grant-in-Aid for specially promoted research 16H06291 from JSPS. Guided access is provided via [www.isimip.org](http://www.isimip.org). We thank the Global Runoff Data Centre (GRDC) for making available their river discharge data.

## Author contributions

J.Sc. designed the research, analysed and interpreted data, and wrote the paper. S.N.G., C.R., F.Z., P.C., J.E., L.F., V.H., H.K.L., S.I.S., M.T.H.v.V., R.V. and Y.W. helped design the research and interpret results. J.Sc., J.E., K.F., L.W., F.Z., C.R., T.D.E., D.P.T., P.C., H.M.S. and S.N.G. coordinated impact model simulations. J.E., L.F., S.N.G., M.T.H.v.V., R.V., Y.W., F.Z., D.A.C., J.C., M.C., D.D., A.d.W., T.D.E., C.F., D.G., N.H., A.I., N.K., P.L., C.Mo., C.M., H.M.S., S.O., Y.P., T.A.M.P., G.S., Y.S., E.S., T.S., J.Stee., J.Stee., Q.T., H. T. and X.W. performed impact model simulations. L.F. and H.K. provided observational data. M.B. and J.V. processed data. L.B., D.A.C., P.C., M.C., T.D.E., J.E., C.F., L.F., A.D.F., D.G., S.N.G., L.G., N.H., V.H., A.I., N.K., H.K.L., C.Mo., C.Mü., H.M.S., R.O., T.A.M.P., C.R., E.S., S.I.S., T.S., J.Stee., J.Stee., Q.T., H.T., D.P.T., M.T.H.v.V., R.V. and Y.W. provided comments and guidance on the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-08745-6>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Jacob Schewe<sup>1</sup>, Simon N. Gosling<sup>2</sup>, Christopher Reyer<sup>1</sup>, Fang Zhao<sup>3</sup>, Philippe Ciais<sup>4</sup>, Joshua Elliott<sup>5</sup>, Louis Francois<sup>6</sup>, Veronika Huber<sup>7</sup>, Heike K. Lotze<sup>8</sup>, Sonia I. Seneviratne<sup>9</sup>, Michelle T.H. van Vliet<sup>10</sup>, Robert Vautard<sup>4</sup>, Yoshihide Wada<sup>11</sup>, Lutz Breuer<sup>12,13</sup>, Matthias Büchner<sup>1</sup>, David A. Carozza<sup>14,43</sup>, Jinfeng Chang<sup>4</sup>, Marta Coll<sup>15</sup>, Delphine Deryng<sup>16,17</sup>, Allard de Wit<sup>18</sup>, Tyler D. Eddy<sup>8,19,20</sup>, Christian Folberth<sup>11</sup>, Katja Frieler<sup>1</sup>, Andrew D. Friend<sup>21</sup>, Dieter Gerten<sup>1,22</sup>, Lukas Gudmundsson<sup>9</sup>, Naota Hanasaki<sup>23</sup>, Akihiko Ito<sup>23</sup>, Nikolay Khabarov<sup>11</sup>, Hyungjun Kim<sup>24</sup>, Peter Lawrence<sup>25</sup>, Catherine Morfopoulos<sup>26</sup>, Christoph Müller<sup>1</sup>, Hannes Müller Schmied<sup>27,28</sup>, René Orth<sup>29,30</sup>, Sebastian Ostberg<sup>1</sup>, Yadu Pokhrel<sup>31</sup>, Thomas A.M. Pugh<sup>32,33</sup>, Gen Sakurai<sup>34</sup>, Yusuke Satoh<sup>10,23</sup>, Erwin Schmid<sup>35</sup>, Tobias Stacke<sup>36</sup>, Jeroen Steenbeek<sup>37</sup>, Jörg Steinkamp<sup>28,44</sup>, Qihong Tang<sup>38</sup>, Hanqin Tian<sup>39</sup>, Derek P. Tittensor<sup>8,40</sup>, Jan Volkholz<sup>1</sup>, Xuhui Wang<sup>4,41,42</sup> & Lila Warszawski<sup>1</sup>

<sup>1</sup>Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, 14473 Potsdam, Germany. <sup>2</sup>School of Geography, University of Nottingham, Nottingham NG7 2RD, UK. <sup>3</sup>School of Geographic Sciences, East China Normal University, Shanghai 200241, China. <sup>4</sup>Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France. <sup>5</sup>University of Chicago and ANL Computation Institute, 5735S. Ellis Ave, Chicago, IL 60637, USA. <sup>6</sup>Institut d'Astrophysique et de Géophysique/U.R. SPHERES, Université de Liège, B-4000 LIEGE, Belgium. <sup>7</sup>Department of Physical, Chemical and Natural Systems, Universidad Pablo de Olavide, Ctra. de Utrera 1, 41013 Sevilla, Spain. <sup>8</sup>Department of Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada. <sup>9</sup>ETH Zurich, Land-Climate Dynamics, Institute for Atmospheric and Climate Science, 8092 Zurich, Switzerland. <sup>10</sup>Water Systems and Global Change group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands. <sup>11</sup>International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria. <sup>12</sup>Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, Heinrich-Buff-Ring 26, 35390 Giessen, Germany. <sup>13</sup>Centre for International Development and Environmental Research (ZEU), Justus Liebig University Giessen, Senckenbergstraße 3, 35392 Giessen, Germany. <sup>14</sup>Department of Earth and Planetary Sciences, McGill University, Montreal H3A 0E8, Canada. <sup>15</sup>Institute of Marine Sciences (ICM - CSIC), Barcelona E-08003, Spain. <sup>16</sup>Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg 15374, Germany. <sup>17</sup>IRI THEsys, Humboldt University of Berlin, 10117 Berlin, Germany. <sup>18</sup>Wageningen Environmental Research, 6700 AA Wageningen, The Netherlands. <sup>19</sup>Nereus Program, Institute for the Oceans and Fisheries, University of British Columbia, Vancouver V6T 1Z4 BC, Canada. <sup>20</sup>Nereus Program, Institute for Marine & Coastal Sciences, School of the Earth, Ocean, and Environment, University of South Carolina, Columbia 29208 SC, USA. <sup>21</sup>Department of Geography, University of Cambridge, Cambridge CB2 3EN, UK. <sup>22</sup>Geography Department, Humboldt-Universität zu Berlin, 10099 Berlin, Germany. <sup>23</sup>National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan. <sup>24</sup>Institute of Industrial Science, the University of Tokyo, Tokyo 153-8505, Japan. <sup>25</sup>Terrestrial Science Section, National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, CO 80305, USA. <sup>26</sup>Imperial College of London, Department of Life Science, Silwood Park Campus, Buckhurst Rd, Beks SL5 7PY, UK. <sup>27</sup>Institute of Physical Geography, Goethe-University Frankfurt, Altenhöferallee 1, 60438 Frankfurt am Main, Germany. <sup>28</sup>Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325 Frankfurt, Germany. <sup>29</sup>Department of Physical Geography, Bolin Centre for Climate Research, Stockholm University, SE-10691 Stockholm, Sweden. <sup>30</sup>Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, D-07745 Jena, Germany. <sup>31</sup>Department of Civil and Environmental Engineering, Michigan State University, MI 48824, USA. <sup>32</sup>School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK. <sup>33</sup>Birmingham Institute of Forest Research, University of Birmingham, Birmingham B15 2TT, UK. <sup>34</sup>Institute for Agro-Environmental Sciences, National Agriculture and Food Research Organization, 3-1-3 Kannondai, Tsukuba, Ibaraki 305-8604, Japan. <sup>35</sup>University of Natural Resources and Life Sciences, Vienna, Feistmantelstrasse 4, 1180 Vienna, Austria. <sup>36</sup>Max Planck Institute for Meteorology, 20146 Hamburg, Germany. <sup>37</sup>Ecopath International Initiative, Bellaterra 08193, Spain. <sup>38</sup>Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 100101 Beijing, China. <sup>39</sup>School of Forestry and Wildlife Sciences, Auburn University, 602 Duncan Drive, Auburn, AL 36849, USA. <sup>40</sup>UN Environment Programme World Conservation Monitoring Centre, 219 Huntingdon Road, Cambridge CB3 0DP, UK. <sup>41</sup>Sino-French Institute of Earth System Sciences, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. <sup>42</sup>Laboratoire de Météorologie Dynamique, Université Pierre et Marie Curie, Paris 75005, France. <sup>43</sup>Present address: Department of Mathematics, Université du Québec à Montréal, Montreal H2X 3Y7, Canada. <sup>44</sup>Present address: Johannes Gutenberg-University, Anselm-Franz-von-Bentzel-Weg 12, 55128 Mainz, Germany