

ENVIRONMENTAL RESEARCH  
LETTERS

## LETTER

## Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios

## OPEN ACCESS

RECEIVED  
1 November 2020REVISED  
14 December 2020ACCEPTED FOR PUBLICATION  
6 January 2021PUBLISHED  
26 February 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Christoph Müller<sup>1,\*</sup> , James Franke<sup>2,3</sup> , Jonas Jägermeyr<sup>1,4,5</sup> , Alex C Ruane<sup>4</sup> , Joshua Elliott<sup>3</sup>, Elisabeth Moyer<sup>2,3</sup>, Jens Heinke<sup>1</sup>, Pete D Falloon<sup>6</sup> , Christian Folberth<sup>7</sup> , Louis Francois<sup>8</sup>, Tobias Hank<sup>9</sup> , R César Izaurralde<sup>10</sup> , Ingrid Jacquemin<sup>8</sup>, Wenfeng Liu<sup>11</sup> , Stefan Olin<sup>12</sup> , Thomas A M Pugh<sup>12,13,14</sup> , Karina Williams<sup>6,15</sup>  and Florian Zabel<sup>9</sup> 

- <sup>1</sup> Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany
  - <sup>2</sup> Department of the Geophysical Sciences, University of Chicago, Chicago, IL, United States of America
  - <sup>3</sup> Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, United States of America
  - <sup>4</sup> NASA Goddard Institute for Space Studies, New York, NY, United States of America
  - <sup>5</sup> Center for Climate Systems Research, Columbia University Earth Institute, New York, United States of America
  - <sup>6</sup> Met Office Hadley Centre, Exeter, United Kingdom
  - <sup>7</sup> Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria
  - <sup>8</sup> Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Liège, Belgium
  - <sup>9</sup> Department of Geography, Ludwig-Maximilians-Universität München (LMU), Munich, Germany
  - <sup>10</sup> Department of Geographical Sciences, University of Maryland, College Park, MD, United States of America
  - <sup>11</sup> College of Water Resources and Civil Engineering, China Agricultural University, Beijing, People's Republic of China
  - <sup>12</sup> Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden
  - <sup>13</sup> School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom
  - <sup>14</sup> Birmingham Institute of Forest Research, University of Birmingham, Birmingham, United Kingdom
  - <sup>15</sup> Global Systems Institute, University of Exeter, Exeter, United Kingdom
- \* Author to whom any correspondence should be addressed.

E-mail: [christoph.mueller@pik-potsdam.de](mailto:christoph.mueller@pik-potsdam.de)

**Keywords:** agriculture, crop modeling, climate change, uncertainty, AgMIP, CMIP

Supplementary material for this article is available [online](#)

**Abstract**

Concerns over climate change are motivated in large part because of their impact on human society. Assessing the effect of that uncertainty on specific potential impacts is demanding, since it requires a systematic survey over both climate and impacts models. We provide a comprehensive evaluation of uncertainty in projected crop yields for maize, spring and winter wheat, rice, and soybean, using a suite of nine crop models and up to 45 CMIP5 and 34 CMIP6 climate projections for three different forcing scenarios. To make this task computationally tractable, we use a new set of statistical crop model emulators. We find that climate and crop models contribute about equally to overall uncertainty. While the ranges of yield uncertainties under CMIP5 and CMIP6 projections are similar, median impact in aggregate total caloric production is typically more negative for the CMIP6 projections (+1% to −19%) than for CMIP5 (+5% to −13%). In the first half of the 21st century and for individual crops is the spread across crop models typically wider than that across climate models, but we find distinct differences between crops: globally, wheat and maize uncertainties are dominated by the crop models, but soybean and rice are more sensitive to the climate projections. Climate models with very similar global mean warming can lead to very different aggregate impacts so that climate model uncertainties remain a significant contributor to agricultural impacts uncertainty. These results show the utility of large-ensemble methods that allow comprehensively evaluating factors affecting crop yields or other impacts under climate change. The crop model ensemble used here is unbalanced and pulls the assumption that all projections are equally plausible into question. Better methods for consistent model testing, also at the level of individual processes, will have to be developed and applied by the crop modeling community.

## 1. Introduction

Climate change impacts on agriculture are subject to large uncertainties from a variety of sources. One of the most important sources of uncertainty is associated with the severity of climate change itself, even for a fixed emission scenario. For example, climate projections in the CMIP5 archive (Coupled Model Inter-comparison Project 5, Taylor *et al* 2012) under the RCP8.5 scenario show a range of 3.2–4.9 K warming in mean growing-season temperatures, and the more recent CMIP6 projections (Eyring *et al* 2016) show a range from 3.6 to 5.9 K. Climate models also differ not only in mean projected changes over large regions but in the spatial patterns of those changes, with precipitation an especial concern (e.g. Akinsanola *et al* 2020, Almazroui *et al* 2020). Recent papers have compared CMIP6 to CMIP5 across a range of impact relevant climate features such as extreme heat, precipitation, ENSO, and the monsoon (e.g. Fan *et al* 2020, Freund *et al* 2020, Jiang *et al* 2020, Xin *et al* 2020, Zhu and Yang 2020) to name a few. In many cases, CMIP6 has improved in skill of representing these climate features, but climate models still show little improvement in some areas. In general CMIP6 is noticeably more sensitive to CO<sub>2</sub> than CMIP5, largely due to the updated representation of aerosols (e.g. Wyser *et al* 2020). Given these wide uncertainties in climate projections, it is important to understand how they translate into uncertainties in potential impacts on crop yields.

Process-based crop models provide a means of understanding the impact of different climate changes on crop yields (Jones *et al* 2017). While these models were first developed for application to individual sites and crop model ensembles were also used at the site level to explore model-induced uncertainty (e.g. Palosuo *et al* 2011, Asseng *et al* 2013), they have been extended to provide global coverage in the global gridded crop model intercomparison (GGCMI, Elliott *et al* 2015) of the Agricultural Model Intercomparison and Improvement Project (AgMIP, Rosenzweig *et al* 2013). Global-scale crop model applications are required for understanding future challenges to agricultural production since production zones may shift under climate change, and individual farms and regions are connected via agricultural markets and technological development and innovation. The combination of global and regional scale analyses has been shown to help in understanding the dynamics of agricultural production systems (Rosenzweig *et al* 2018, Ruane *et al* 2018). Global crop simulations do suffer some uncertainties since many processes cannot be fully calibrated at large scales—suitable reference data and management information is not available for all regions—but global crop simulations have been shown to have skill in reproducing observed historical inter-annual variability and spatial patterns (Müller *et al* 2017).

Global assessments across ensembles of both crop models and climate projections require some means of reducing computational demands. A comprehensive set of climate projections in the CMIP5 or CMIP6 archives would consist of up to 34 and 45 members per radiative forcing scenario, and the GGCMI phase 2 experiment alone involved 12 different global crop models (Franke *et al* 2020a). These numbers are prohibitive for computing a full set of crop yield simulations driven by different climate projections. In practice, studies of future climate impacts on crop yields are often performed using small and sometimes arbitrary selections of climate projections, crop models, or crops. For example, Mcsweeney and Jones (2016) find that considering only five individual climate models in global impact assessments falls short of representing the underlying uncertainty. A larger sample is required to fully characterize the uncertainty range of climate models. Yet, a higher number of climate scenarios often proves unpractical from the perspective of computational resources and climate change impact assessments on agriculture often rely on climate projections from a small set of climate models (e.g. Rosenzweig *et al* 2014).

In this work we avoid these computational bottlenecks and provide a more comprehensive impacts assessment by using statistical emulators of individual crop models. We present results of a global-scale assessment of potential crop yield changes that explores the full range of the CMIP5 and CMIP6 climate projection archive. We use a set of nine global gridded crop model (GGCM) emulators (Franke *et al* 2020b) that were trained on a very large systematic input sensitivity analysis with up to 1404 simulation data sets per crop and crop model, each of 31 years in length and with near-global coverage (Franke *et al* 2020a). The training domain represents an unprecedentedly rich data base for emulator training, with perturbations in atmospheric carbon dioxide (CO<sub>2</sub>) concentrations (four levels from 360 ppm to 810 ppm), air temperature (seven levels from −1 to +6 K), water supply (eight levels from −50% to +30% precipitation and full irrigation), nitrogen (N) fertilization (three levels from 10 to 200 kgN ha<sup>−1</sup>) and adaptation (two levels: none and maintained growing seasons). The emulators themselves are grid cell-specific regression models with 34 coefficients (Franke *et al* 2020b). Emulation allows a computationally light-weight means of assessing crop yield impacts under arbitrary climate and CO<sub>2</sub> scenarios that can be applied to the full CMIP5 and CMIP6 climate archive. This exercise therefore allows us to evaluate the uncertainty in climate model projections through the perspective of its implications for global food production.

In this analysis, we break down the different sources of uncertainty (greenhouse gas concentration pathways, climate model, crop model) assess the role of the modeled response to CO<sub>2</sub>

fertilization and growing season adaptation and identify future directions for crop model development and improvements.

## 2. Methods

In order to assess the current uncertainty in projections of future crop productivity on current cropland, we combine the full GGCM phase 2 crop model emulator ensemble (Franke *et al* 2020b) with the full GCM ensemble of the CMIP5 and CMIP6 archives for three different radiative forcing pathways: the representative concentration pathways (RCP) 2.6, 4.6 and 8.5 (van Vuuren *et al* 2011).

The crop model emulator ensemble consists of 3rd order polynomial regression models for nine different GGCMs for the major staple crops maize, spring wheat, winter wheat, rice and soybean. The emulators can reproduce well the response of the original crop models to changes in CO<sub>2</sub> ( $C$ ), temperatures ( $T$ ), water supply ( $W$ ), N inputs and growing season adaptation ( $A$ ) that the models showed in a large input sensitivity study using systematic parameter sweeps along the CTWN-A dimensions. All crops are simulated separately for purely rainfed and for fully irrigated systems, where irrigation is one element on the water availability dimension ( $W$ ). The CTWN-A experiment of the GGCM phase 2 is described in detail by Franke *et al* (2020a). Emulator design and performance is described in detail by Franke *et al* (2020b). The emulators compute crop yields per crop and geographic location (geographic grid at 0.5° longitude/latitude resolution) from atmospheric CO<sub>2</sub>, changes in growing season temperature ( $\Delta T$ ) and growing season precipitation ( $\Delta P$ ), as well as N fertilizer inputs. Separate emulators exist for purely rainfed and irrigated production systems as well as for the non-adapted setting (same planting dates and variety selection as in the baseline period) and the adapted setting (same planting dates, but new varieties that allow for maintaining the original growing season length under warming). In this analysis, we work explicitly with the crop model emulators, but since these are crop model specific emulators, we refer to the GGCM-specific emulators with the names of the underlying GGCMs (CARAIB (Dury *et al* 2011), EPIC-TAMU (Izaurrealde *et al* 2006), GEPIC (Folberth *et al* 2012), JULES (Williams *et al* 2017), LPJ-GUESS (Olin *et al* 2015), LPJml (von Bloh *et al* 2018), PEPIC (Liu *et al* 2016), PROMET (Hank *et al* 2015), pDSSAT (Elliott *et al* 2014)).

We obtained the largest possible climate model (GCM) ensemble from the CMIP5 and CMIP6 archives that provide data for the historical period and at least one of the three RCPs considered here (RCP2.6, RCP4.5 and RCP8.5 for CMIP5; SSP126, SSP245 and SSP585 from the ScenarioMIP in CMIP6 (O'Neill *et al* 2016)). As we only consider GCMs that

contribute at least one of the considered RCPs and the historical period, our GCM ensemble can differ from other ensembles (e.g. Meehl *et al* 2020). In order to compute future yield projections, we compute average growing season mean temperatures and average total growing season precipitation for a baseline period (1980–2010) for each grid cell that is currently used to produce any of the five crops considered here, following the MIRCA2000 data set (Portmann *et al* 2010) and distinguishing both irrigated and rainfed growing seasons. The 31 year baseline period corresponds to the reference period of the AgMERRA data set (Ruane *et al* 2015) that was used as the basis for the GGCM phase 2 CTWN-A simulations (Franke *et al* 2020a) and previous crop model evaluation (Müller *et al* 2017).

Against this crop- and grid cell-specific baseline conditions, we compute absolute differences in average growing season temperature ( $\Delta T$  in K) and relative differences in total growing season precipitation ( $\Delta P$ , unitless) for all future 31 year moving window periods in the 21st century (2011–2084). As we are only interested in changes in 31 year average  $T$  and  $P$  from the historical simulation of the same GCM, no bias correction is necessary for the computation of  $\Delta T$  and  $\Delta P$ .

Growing season  $T$  is computed as the weighted average of monthly  $T$  data from each GCM, using the days per month within the growing season as weights. Growing season  $P$  is computed in a similar way, but using growing season totals, by adding monthly precipitation sums using days per month within the growing season to compute shares of precipitation that are considered as part of the growing season. Crop- and grid cell-specific growing season start and end dates are taken from the dataset used in the GGCM simulation phases 1 and 2 (Elliott *et al* 2015, Franke *et al* 2020a) so that these are consistent with what is assumed by the emulators. We do not change growing season length with increasing warming for the computation of average growing season conditions.

We consider all climate model projections in the CMIP archive that provide historical and future scenarios in a consistent manner. We used monthly data rather than daily data to increase sample size, which we consider more important than daily resolution. We assume errors induced by this are small, especially since growing season conditions are computed as 31 year moving window averages, which is the time frame on which the emulators have been trained (Franke *et al* 2020b). We accept different parameterization schemes of the same GCM as separate models where available to further increase sample size. We always only consider one ensemble simulation set per GCM, parameterization and RCP, selecting the smallest run number in the archive if several versions are available. Detailed information on the 45 CMIP5 and 34 CMIP6 models considered, including version and

ensemble member numbers, are listed in the supplementary tables S1 and S2 (which are available online at [stacks.iop.org/ERL/16/034040/mmedia](https://stacks.iop.org/ERL/16/034040/mmedia)).

As some GGCMs tend to differ in their simulated baseline crop productivity levels (see e.g. Müller *et al* 2017), we harmonize simulated crop yields ( $Y^*_t$ ) to match observed yield patterns from Mueller *et al* (2012) as in equation (1), where  $Y_t$  is the simulated yield in time step  $t$ ,  $A_c$  is the harvested area in grid cell  $c$ ,  $O_{r,c}$  is the observed yield in the reference period  $r$  and cell  $c$  and  $Y_{r,c}$  is the simulated yield in the reference period in cell  $c$ .

$$Y^*_t = \sum Y_{t,c} \times A_c \times O_{r,c} / Y_{r,c}. \quad (1)$$

This is a simple multiplicative bias adjustment compared to more complex approaches used for the bias adjustment of climate projections. Our analysis is based on 31 year averages so that the focus is not on inter-annual or seasonal variations. Still the adjustment of the productivity levels helps to eliminate increased variance in the crop model ensemble from differences in mean biases as we are interested in projected changes here.

Crop yield data are aggregated to global production ( $P$ ) using crop-specific harvested area data from MIRCA2000 (Portmann *et al* 2010). As winter and spring wheat are not explicitly distinguished in MIRCA2000, we assume that winter wheat is grown in a specific grid cell if the average temperature of the coldest month of the year is between  $-10^\circ\text{C}$  and  $+7^\circ\text{C}$  or if the growing season is longer than 150 d or if the growing season includes December (Northern Hemisphere) or July (Southern Hemisphere). Otherwise we assume that spring wheat is grown (see map in supplementary figure S1). Changes in production are equivalent to changes in productivity (yields) here, as the harvested area data set is static in time (equation (2)).

$$P_t = \sum Y^*_{t,c} \times A_c. \quad (2)$$

For the aggregation of different crops, we compute total calories, assuming net water contents of 12% for maize, spring and winter wheat, 13% for rice and 9% for soybean, according to Wirsenius (2000) and caloric contents of the ‘as purchased’ biomass (i.e. including the water content) of  $3.56 \text{ kcal g}^{-1}$  for maize,  $2.8 \text{ kcal g}^{-1}$  for rice,  $3.35 \text{ kcal g}^{-1}$  for soybean and of  $3.34 \text{ kcal g}^{-1}$  for spring and winter wheat, following (FAO 2001).

As the central metric for uncertainty in crop yield projections, we compute total variance across all GCM  $\times$  GGCM combinations for all crops separately and for total calorie production of all five crops considered here. We assume that the total variance  $\text{var}(\text{total})$  is the sum of the variance across all GCMs  $\text{var}(\text{GCM})$  after averaging across all GGCMs

and of the variance across all GGCMs  $\text{var}(\text{GGCM})$  after averaging across all GCMs, plus a cross term that describes the covariance between GCM and GGCM responses. This cross term cannot be directly computed but we assume it to be the difference to unity (equation (3)).

$$1 = \frac{\text{var}(\text{GCM})}{\text{var}(\text{total})} + \frac{\text{var}(\text{GGCM})}{\text{var}(\text{total})} + \text{cross.terms.} \quad (3)$$

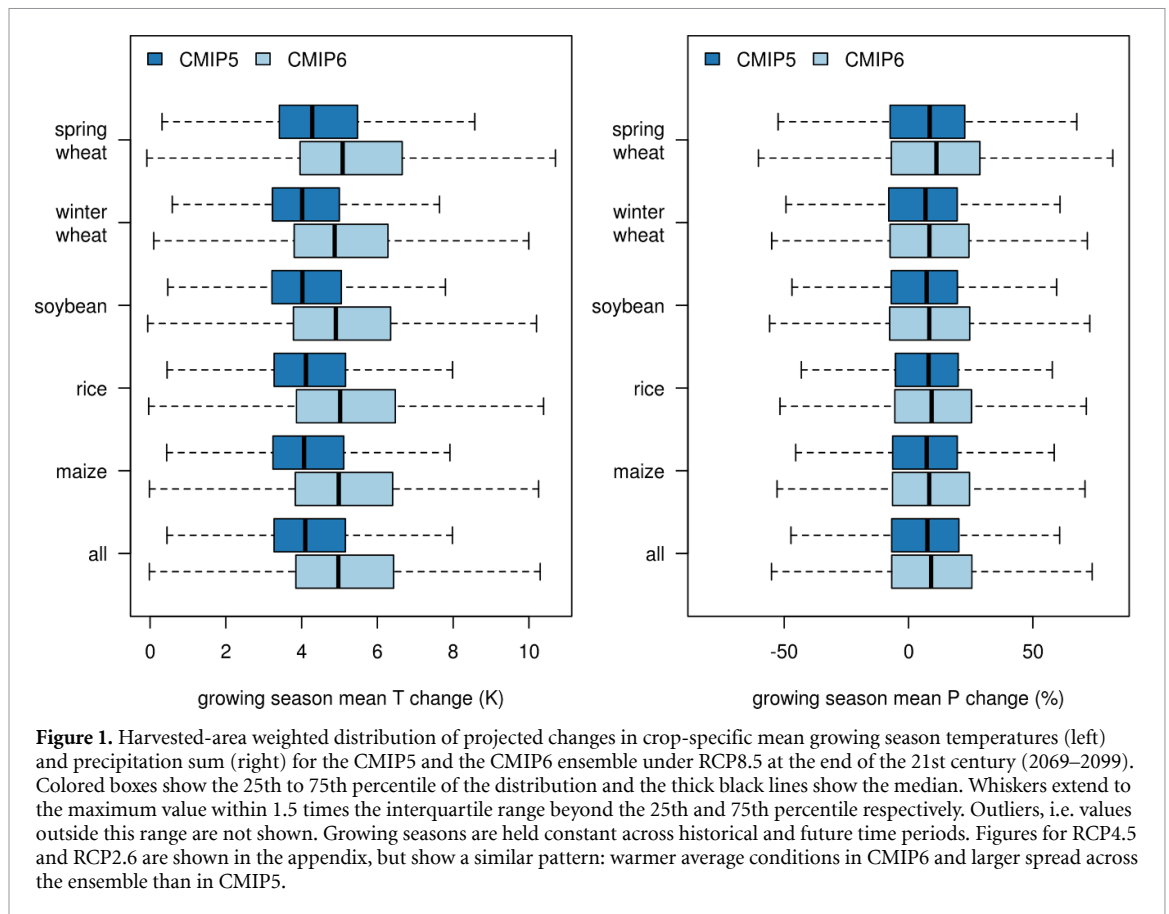
With this assumption, which follows a similar uncertainty decomposition in climate projections by Hawkins and Sutton (2009), shares of total variance can be attributed to differences in GCMs or differences in GGCMs.

To test the robustness of this attribution to the ensemble composition, we compute the variances for all sub-sets, leaving out one GGCM each time (i.e.  $11\% \approx 1/9$ ), testing if variance attribution is sensitive to the ensemble composition.

The GGCM phase 2 input sensitivity CTNW-A experiment tested temperature increases of up to  $+6 \text{ K}$  and precipitation changes between  $-50\%$  and  $+30\%$  (Franke *et al* 2020a). Under RCP8.5 (SSP585 for CMIP6), some GCMs exceed this temperature range for some cropland areas. With the non-linear design of the GGCM crop model emulator ensemble (Franke *et al* 2020b), it is difficult to extrapolate beyond its training domain range, especially in the temperature dimension, which is, together with the ( $\text{CO}_2$ ) dimension, typically the most powerful feature in the models. To avoid overly spurious crop model projections, we capped growing season temperature changes ( $\Delta T$ ) at  $+6 \text{ K}$  and changes in precipitation at  $-50\%$  and  $+30\%$  at the grid cell and crop-specific growing season level. As the emulators rely on the balance of the  $T$  and  $C$  terms, we simultaneously kept ( $\text{CO}_2$ ) constant at the grid cell and crop-specific growing season value at which  $+6 \text{ K}$  for  $\Delta T$  was reached. The majority of GCMs has only small fractions of current cropland that exceed  $\Delta T$  of  $+6 \text{ K}$ , but for some models, this can be substantial. For the CMIP5 ensemble, 7% of all cropland exceeds  $+6 \text{ K}$  (2% for rice to 12% for spring wheat) averaged across all GCMs for RCP8.5, while this is more severe for the CMIP6 ensemble (22% of all cropland, ranging from 9% for rice to 29% for spring wheat; see supplementary figure S2). As we drive the emulators with 31 year moving window average, the last year considered here is 2084 (2069–2099). Therefore, we did not have to generally cap ( $\text{CO}_2$ ) at 810 ppm, as this concentration level is only exceeded after 2086 (Riahi *et al* 2011).

Of the CMIP5 archive, CESM1-CAM5-1-FV2 had to be excluded due to missing precipitation values for Dec 2056 and in the CMIP6 archive, CIESM had to be excluded due to implausible strong decline of temperatures at the end of the 21st century.





### 3. Results

#### 3.1. Changes in T and P projections in CMIP ensembles

Generally, the spread of growing season changes in temperatures and precipitation is larger in the CMIP6 ensemble with 34 members than in the CMIP5 ensemble with 45 members (figure 1). Under the high radiative forcing scenario RCP8.5, the CMIP6 ensemble projects a stronger median warming of about 1 K and similar changes in precipitation as the CMIP5 ensemble. Differences in projected growing season warming are less pronounced in lower radiative forcing cases (RCP2.6 and RCP4.5) and scale with the radiative forcing (figure 1, supplementary figures S3 and S4).

#### 3.2. Projected impacts

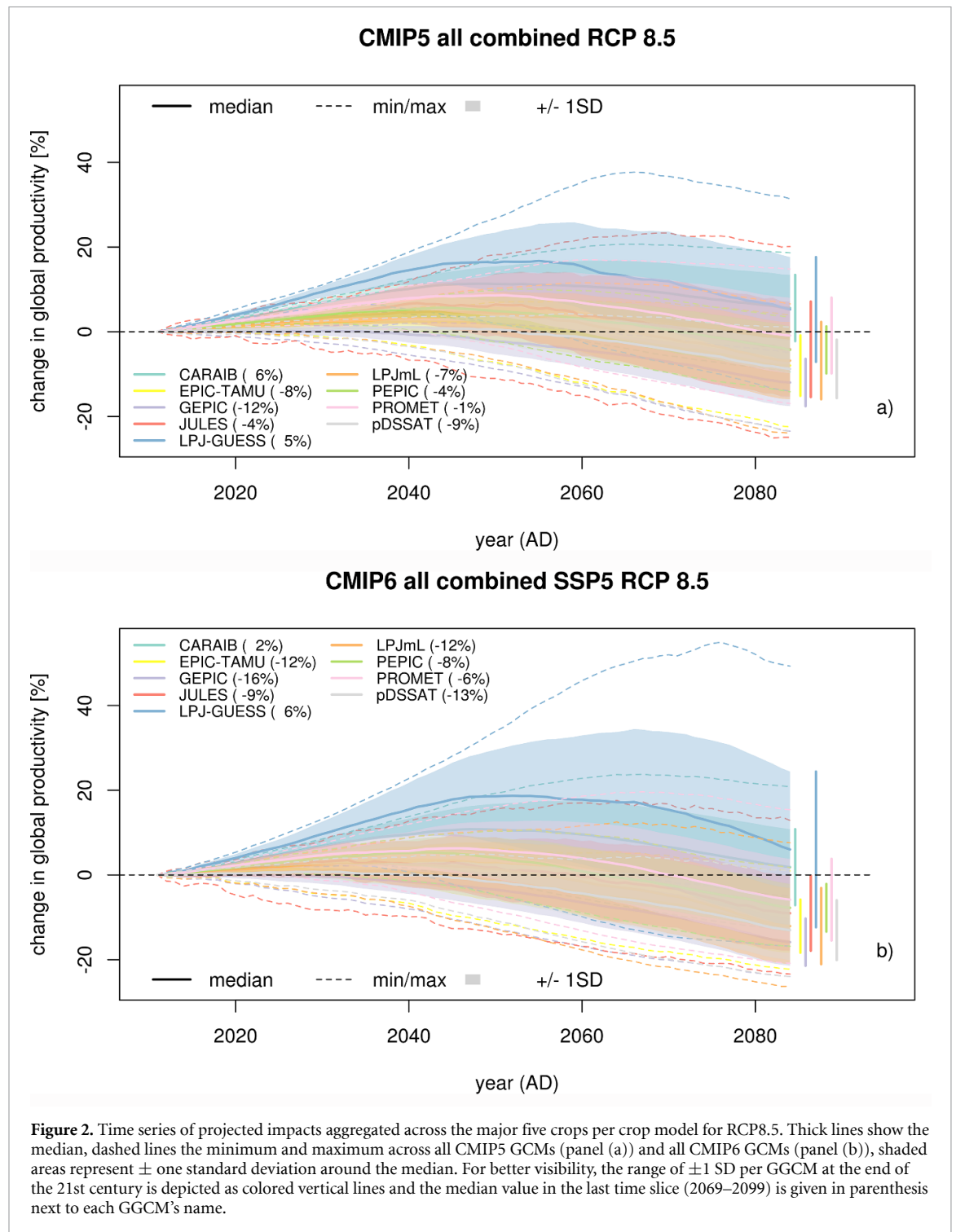
At the most aggregated level (across all crops, globally), the GCM  $\times$  GGCM ensemble projects a broad range of possible climate change impacts on crop productivity on current cropland (figure 2). The ensemble of crop model emulators projects consistently more negative impacts on average (except for LPJ-GUESS where projections increase by 1% point), so that the uncertainty range ( $\pm 1$  standard deviation, colored area in figure 2) of only three GGCMs overlaps the zero line (CARAIB, LPJ-GUESS, PROMET)

for the CMIP6 ensemble, while this is the case for all but three crop models under CMIP5. Still the most extreme projections for the CMIP6 scenario span farther into the positive range than they do under CMIP5 (figure 2).

We observe distinct differences between individual GCMs, with GEPIC and pDSSAT being typically the most pessimistic models and CARIB and LPJ-GUESS the most optimistic ones.

Projected impacts scale with the radiative forcing and with the GCMs' equilibrium climate sensitivity (ECS, taken from Meehl *et al* (2020)), which constitute an important determinant of crop yield projections. Projected impacts are generally less variable at lower radiative forcing (time axes in figure 2 and different RCPs in supplementary figures S5 and S6). Under RCP2.6, all but GEPIC project a positive median change for the CMIP5 ensemble and all but GEPIC and pDSSAT do so for RCP2.6 and CMIP6 (supplementary figure S5) and for RCP4.5 and CMIP5. For RCP4.5 and CMIP6, five of nine GGCMs project negative median impacts by the end of the 21st century (supplementary figure S6).

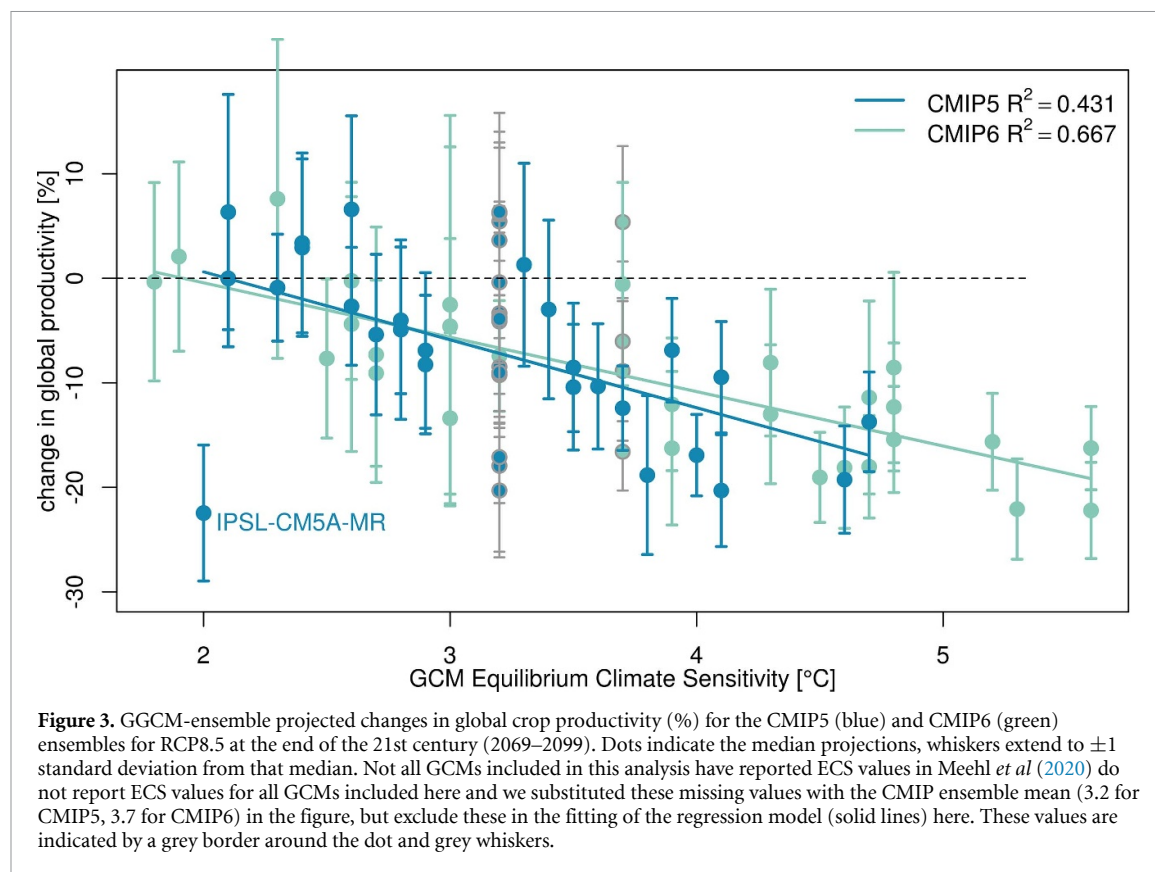
The relationship between ECS and median climate change impact on crop yields is stronger for the CMIP6 ensemble (figure 3). However, the range of projected changes in crop productivity can differ substantially at similar ECS values. The ECS relationship with changes in crop productivity is weaker for the



CMIP5-based ensemble as the GCM with the lowest ECS (IPSL-CM5A-MR) shows the strongest decline in crop productivity (figure 3). The low ECS value reported by Meehl *et al* (2020) is also not reflected in the temperature increase of IPSL-CM5A-MR on current cropland of the five crops considered here, where the mean temperature increase is not exceptionally high in comparison to other GCMs, but certainly not at the low end (supplementary figure S7). This suggests that the IPSL-CM5A-MR model may have a different distribution of warming over oceans

vs land or a much lower warming on non-cultivated land.

At the level of individual crops the GGCM ensemble shows distinct differences, even though GEPIC and pDSSAT generally belong to the more pessimistic models and CARAIB and LPJ-GUESS generally belong to the more optimistic models. For maize, pDSSAT is the most pessimistic model, distinctly more so than the other models, with end-of-the-century median projections of  $-32\%$  ( $-41\%$ ) in comparison to  $-15\%$  ( $-21\%$ ) for



GEPIC, the next most pessimistic GGCM for CMIP5 (CMIP6), see supplementary figure S8), but also the  $\pm 1$  SD range of GEPIC does not overlap with that of CARIB, LPJ-GUESS and PROMET. LPJ-GUESS projections broaden the projection range of the GGCM ensemble substantially to the positive side for spring and winter wheat, but it also covers the very pessimistic projection range for winter wheat. For these crops, LPJ-GUESS is the most sensitive model to different GCMs.

There is no emulator for LPJ-GUESS for rice and soybean, as no simulations were submitted for these crops to the GGCM phase 2 data archive (Franke *et al* 2020a). The  $\pm 1$  SD range of all GGCMs overlap for soybean, whereas those of CARAIB and JULES for rice do not overlap with the  $\pm 1$  SD ranges of EPIC-TAMU and GEPIC and that of JULES does not overlap with PEPIC in both CMIP5 and CMIP6 and with that of pDSSAT only for CMIP5 (supplementary figures S8–S12).

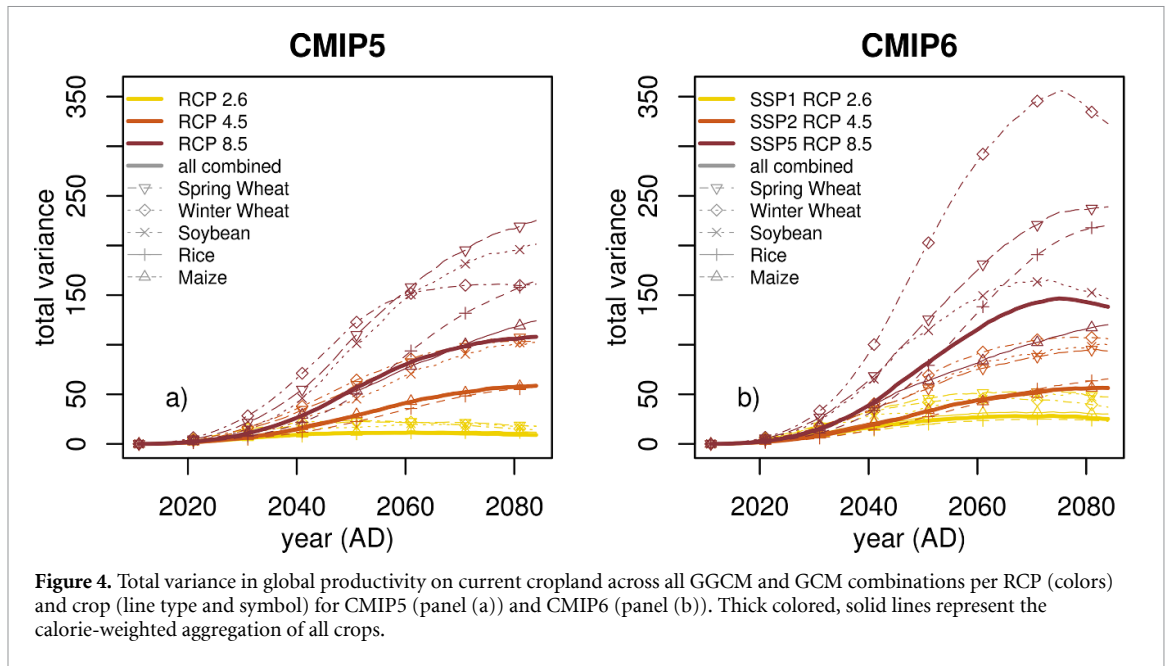
### 3.3. Sources of uncertainty

We find substantial differences in overall variance in projected changes in crop productivity between the CMIP5 and CMIP6 ensembles. Total variance of the full crop model emulator and climate projections ensemble, as a measure for uncertainty, is larger for CMIP6 than for CMIP5 (figure 4) for RCP2.6 and 8.5, but similar for RCP4.5. In the CMIP6 ensemble, the variance of both wheats, but especially winter wheat increases compared to the CMIP5 ensemble under

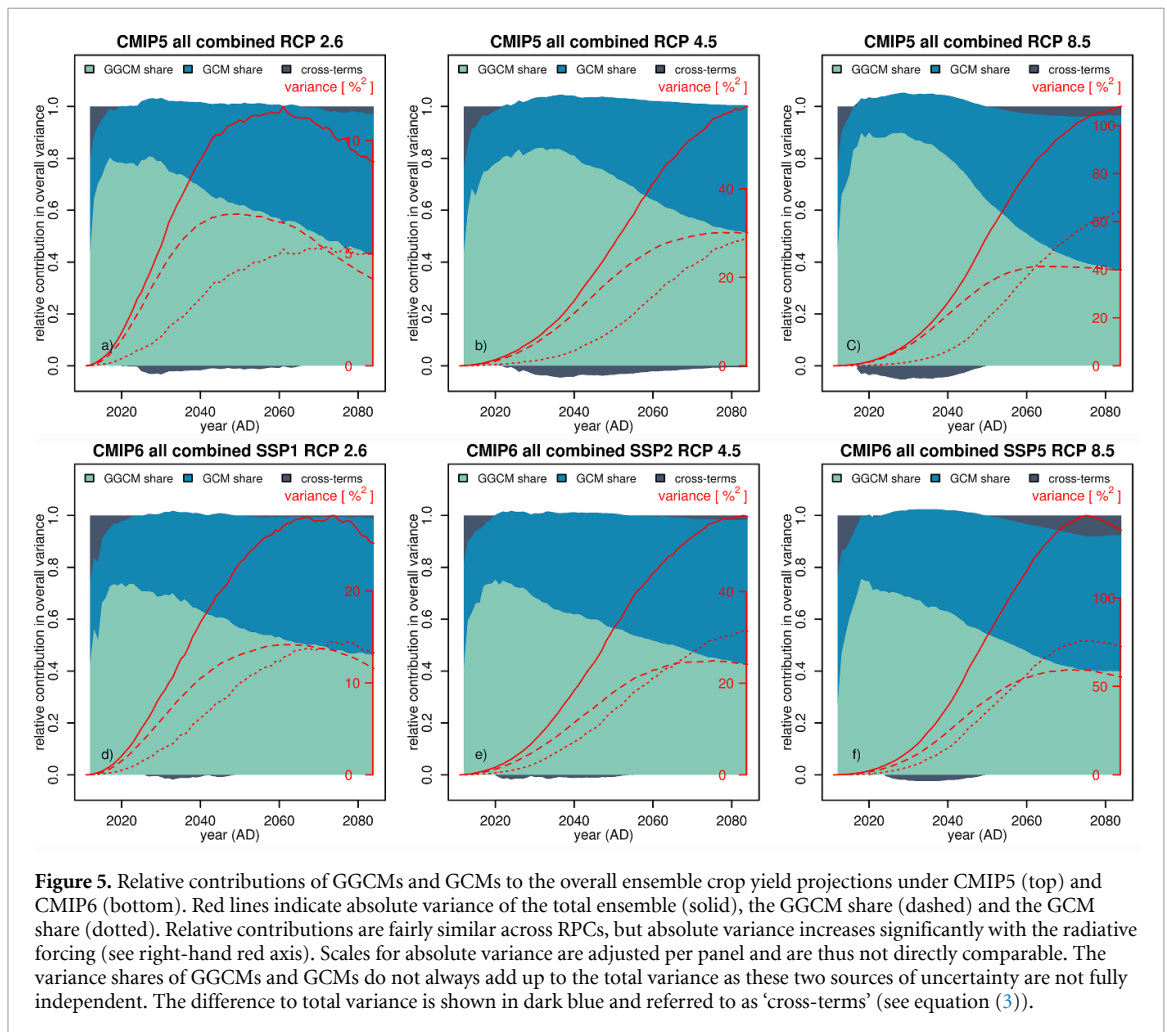
the high radiative forcing pathway RCP8.5, while that of soybean decreases. The overall variance of crop yield projections of the ensemble increases with the radiative forcing (RCP, time) in both the CMIP5 and CMIP6 ensembles (figure 4). This increase is strongest in the middle of the 21st century and levels off towards the end of the 21st century. This leveling-off effect can be observed at all RCPs (figure 5), but is less strong for simulations where the effect of CO<sub>2</sub> fertilization is ignored or where growing season adaptation is considered (figure 6).

Breaking down overall variance in projections into a GGCM and a GCM component, we find that the GGCM component dominates in the first half of the 21st century and the GCM component gradually increases after a peak in GGCM component, typically between 2020 and 2030 (figure 5). The shares of GGCM and GCM-induced variance are largely independent and cross-terms typically account for only a small fraction of the overall variance. The peak in GGCM-induced variance is less pronounced in the CMIP6 ensemble than in CMIP5 ensemble, because the GCM-induced variance increases strongly only in the second half of the 21st century in the CMIP5 ensemble, but increases more steeply (relative to the GGCM-induced variance) from 2020 onwards in the CMIP6 ensemble.

While overall variance can be substantially decreased if the CO<sub>2</sub> fertilization effect is ignored, the share of GGCM-induced variance tends to increase under this setting, especially in the CMIP6 ensemble



**Figure 4.** Total variance in global productivity on current cropland across all GGCM and GCM combinations per RCP (colors) and crop (line type and symbol) for CMIP5 (panel (a)) and CMIP6 (panel (b)). Thick colored, solid lines represent the calorie-weighted aggregation of all crops.

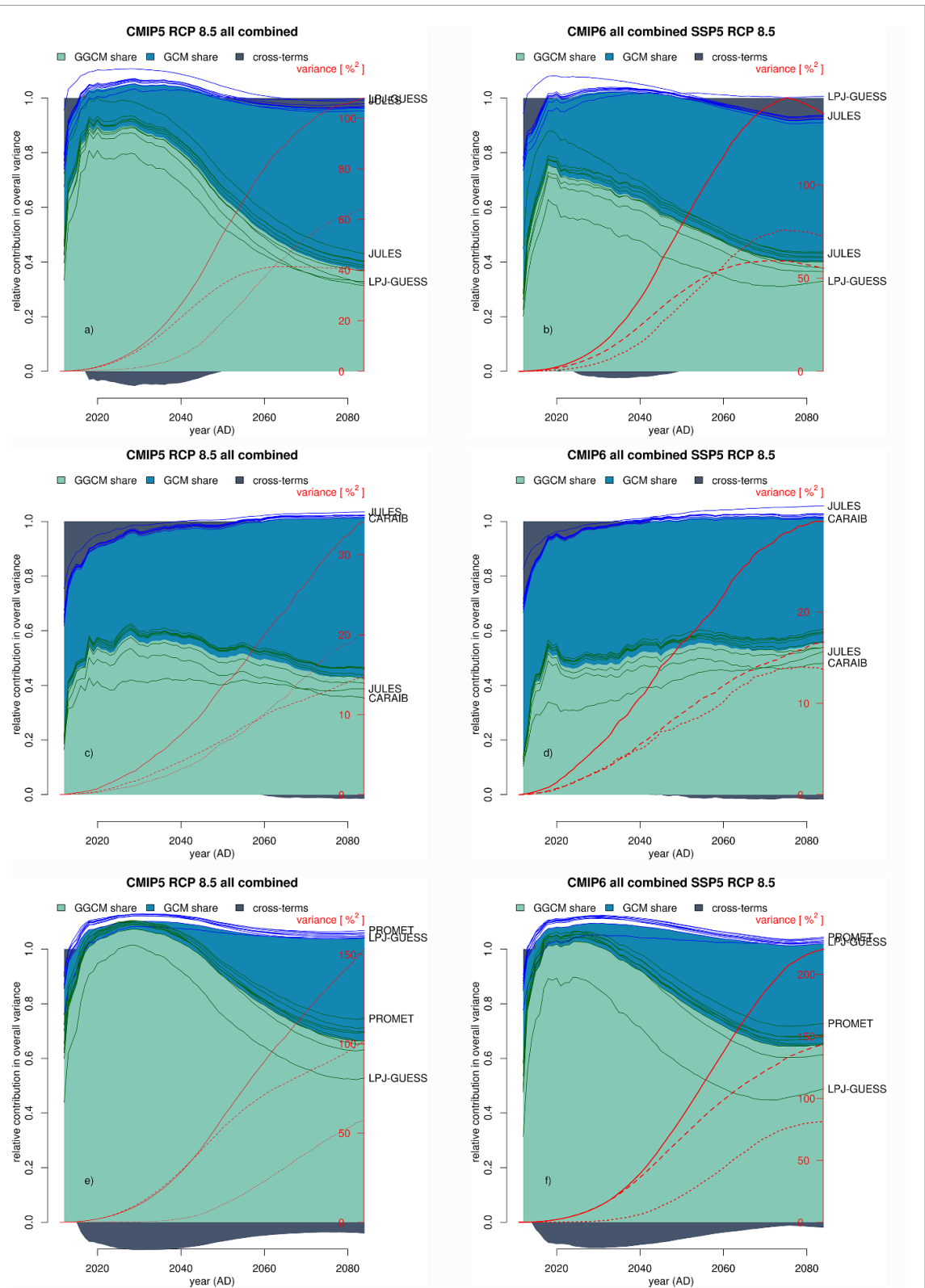


**Figure 5.** Relative contributions of GGCMs and GCMs to the overall ensemble crop yield projections under CMIP5 (top) and CMIP6 (bottom). Red lines indicate absolute variance of the total ensemble (solid), the GGCM share (dashed) and the GCM share (dotted). Relative contributions are fairly similar across RPCs, but absolute variance increases significantly with the radiative forcing (see right-hand red axis). Scales for absolute variance are adjusted per panel and are thus not directly comparable. The variance shares of GGCMs and GCMs do not always add up to the total variance as these two sources of uncertainty are not fully independent. The difference to total variance is shown in dark blue and referred to as ‘cross-terms’ (see equation (3)).

(figure 6). Ignoring the CO<sub>2</sub> fertilization effect does not provide plausible future crop yield projections, but it helps to analyze where the GGCM-induced variance originates from. We find that crop models agree more strongly, if the process of CO<sub>2</sub>

fertilization is ignored. In other words, the simulated effects of CO<sub>2</sub> fertilization on crop yields are an important source of crop model disagreement. Adaptation of cultivars to regain the growing season length that would otherwise be lost due to accelerated





**Figure 6.** As figure 5, but for the standard setting (top row; panels are equivalent to right hand panels in figure 5), the projections ignoring the CO<sub>2</sub> fertilization effect (middle row) and the projections including the variety adaptation to regain the growing season (bottom row) for CMIP5 (left) and CMIP6 (right). Thin lines show how GCM- (blue) and GGCM-induced shares (green) in overall variance would change if one GGCM were excluded from the ensemble. The exclusion of individual GCMs can also affect the contribution of cross-terms, i.e. higher or lower co-variance between the GCM- and GGCM-shares (e.g. thin blue lines above 1.0). Red lines indicate absolute variance of the total ensemble (solid), the GGCM share (dashed) and the GCM share (dotted). Relative contributions are fairly similar across RPCs, but absolute variance increases significantly with the radiative forcing.

phenological development (Minoli *et al* 2019, Franke *et al* 2020a) on the other hand increases the GGCM-induced variance share and overall variance substantially. This is because crop models show very different responses to this adaptation measure so that overall uncertainty is increased if cultivar adaptation (as implemented in the GGCM phase 2 simulations) is considered (Minoli *et al* 2019).

We also find that the ensemble of crop models is very sensitive to the selection of ensemble members. If one of the nine crop models is excluded from the ensemble, the relative contribution to overall variance from crop models can vary strongly (figure 6). Which GGCM has strong effects on the overall variance attribution is crop specific. If random sets of climate models that constitute a similar share of the ensemble size ( $n = 4$  of 34 for CMIP6, roughly equivalent of one in nine crop models), we find that results on the GCM- and GGCM-induced variance shares change less than if individual GGCMs are excluded in the first half of the 21st century, but can be affected similarly strongly at the end of the century (supplementary figure S13), suggesting that the distribution of changes in the GCM ensemble is more balanced in short-term projections than that within the GGCM ensemble.

### 3.4. Crop specific differences

For individual crops, we observe substantial differences in the share of variance that can be attributed to crop models. For maize and spring wheat, the GGCM-induced variance shares clearly dominate the overall variance. GCM-induced variance is clearly the most important contribution to overall variance in soybean yield projections and to lesser extent in rice projections. Winter wheat shows a strong contribution of cross terms to the overall variance, which is also true to some extent for spring wheat. This cross-term contribution can be substantially reduced by excluding LPJ-GUESS from the winter wheat GGCM ensemble. Excluding JULES from the spring wheat GGCM ensemble would increase the GGCM-induced variance share in the first half of the 21st century and would introduce negative cross-terms. Excluding LPJ-GUESS from the spring-wheat ensemble on the other hand would do the opposite and reduce the GGCM-induced variance share throughout most of the 21st century and would introduce larger positive cross-term shares (figure 7).

For most crops, there is a clear outlier model that, if excluded, strongly changes the contribution of GCMs, GGCMs or cross-terms to overall variance. For maize, this is pDSSAT, which projects the most pessimistic yield declines in the GGCM phase 2 emulator ensemble (see appendix figure A1). An exclusion of pDSSAT from the ensemble would reduce overall variance by more than half and substantially reduce the GGCM-induced contribution.

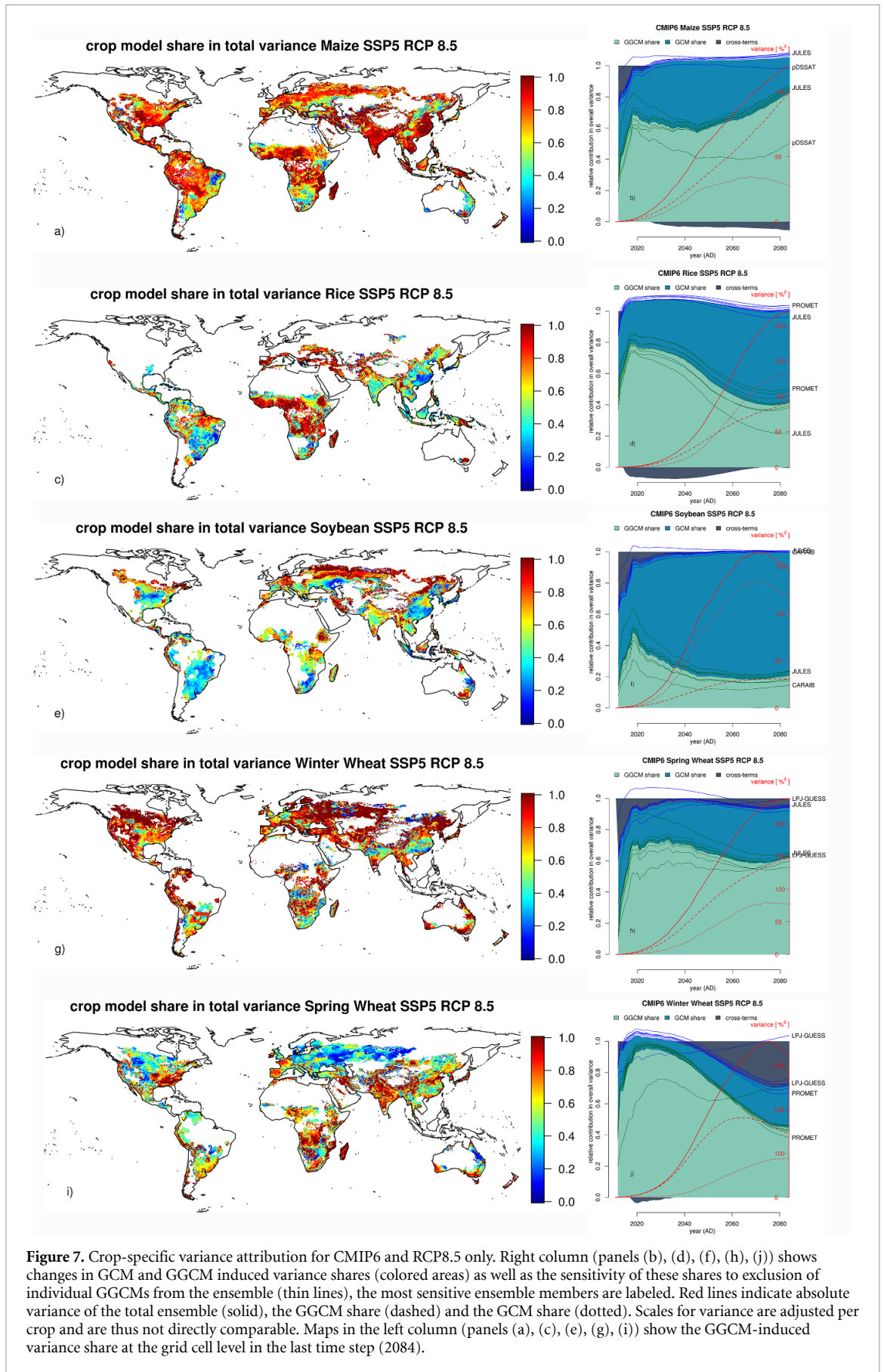
If PROMET, LPJmL or pDSSAT were excluded from the rice model ensemble, the GGCM induced variance would increase, whereas it would substantially decrease if JULES were excluded. The exclusion of JULES would also substantially reduce overall variance of the full GCM  $\times$  GGCM ensemble. Even though there is generally much less GGCM-induced variance in soybean yield projections, the exclusion of CARAIB would lead to a further reduction of overall variance and of the GGCM-induced share.

## 4. Discussion

This unprecedentedly large ensemble of climate projections, crop model (emulators) and crops allows to explore the importance of ensemble composition for climate change impact analyses on crop yields and examine the uncertainty in climate model ensembles through the lens of climate impacts. We find that climate projections can have a substantial influence on crop yield projections, especially in marginal and dry regions, but spatial patterns differ by crop (figure 7).

The use of computationally efficient crop model emulators in place of the process-based crop models is the only option to conduct this large ensemble analysis. While the emulators have very good skill in reproducing the underlying crop models (Franke *et al* 2020b), they are no perfect reproduction of the crop models' dynamics. Our results are thus only indicative of the actual contributions of crop models to overall uncertainty in crop yield projections.

Across the full CMIP5 and CMIP6 archives, there is substantial spread in crop yield projections, independent of the radiative forcing (RCP2.6, 4.5, or 8.5). At the end of the 21st century, climate model-induced variance is often dominant over crop model-induced variance, i.e. the uncertainties in climate projections are more important for projections of changes in crop yields than the uncertainties in crop models—at least at the most aggregate level (combined global productivity of all crops considered here). For individual crops, crop-model induced variance is larger than the climate model-induced variance for maize, spring wheat and winter wheat, which jointly contribute the majority of calories from the five crops considered here. As such, it is surprising to see that climate model-induced variance is dominant over crop model-induced variance when the five crops are aggregated to overall production. This suggests that there is some cancelation of signals when different crops are aggregated. One example of such mutual compensation of variance is the combination of predominantly negative projections for maize productivity (supplementary figure S8) and the predominantly positive projections for spring and especially winter wheat (supplementary figure S12). This may illustrate compensatory responses between crops within



**Figure 7.** Crop-specific variance attribution for CMIP6 and RCP8.5 only. Right column (panels (b), (d), (f), (h), (j)) shows changes in GCM and GGCM induced variance shares (colored areas) as well as the sensitivity of these shares to exclusion of individual GGCMs from the ensemble (thin lines), the most sensitive ensemble members are labeled. Red lines indicate absolute variance of the total ensemble (solid), the GGCM share (dashed) and the GCM share (dotted). Scales for variance are adjusted per crop and are thus not directly comparable. Maps in the left column (panels (a), (c), (e), (g), (i)) show the GGCM-induced variance share at the grid cell level in the last time step (2084).

the crop models and/or changing patterns of warming within the climate models. Similarly, variances in space can cancel out in the aggregation to global productivity if some regions are projected to see positive effects and others to see negative impacts of climate change (e.g. winter wheat in figure 7). Looking at the distribution of projected changes in global crop productivity as done here via the variability metric does thus not represent the full scope of disagreement among simulations. The aggregation of data across space or crops can lead to cancelation of variance at the underlying level of detail that is not visible at the level of analysis here. Still, the analysis provides a unique overview of the breadth of projections of global crop productivity under climate change.

Differences across crops do not necessarily only represent differences in the simulated dynamics and processes of these crops, but can also reflect the differences in the crop model ensemble. LPJ-GUESS for example, which is at the most positive side of the projected yield changes for spring and winter wheat did not supply data for soybean and rice in the CTWN-A experiment (Franke *et al* 2020a) and is thus not included in the emulator ensemble for these crops (Franke *et al* 2020b). However, the exclusion of LPJ-GUESS from the wheat ensembles does not make the uncertainty attribution for spring and winter wheat more similar to that of the other crops.

More crops need to be explicitly considered in climate change impact assessments, as individual crops show distinctly different spatial patterns, uncertainties in crop yield projections and the relative contribution of GCM- vs GGCM-induced variance. As we find that impact projections (supplementary figures S8–S12) as well as drivers of uncertainty (figure 7) differ between different cereal crops, other crops like legumes, tree or other perennial crops must be explicitly analyzed. Therefore, the behavior of crops other than the major five considered here can likely not be well represented by these. Considering the comparative high amount of research attention these five crops have received, uncertainty must be very high for other crops. It is thus of fundamental importance to broaden the range of simulated crops, also because there is the need to represent a much broader set of crops in economic analyses of agricultural markets and land-use dynamics under climate change. The current practice to derive climate change impacts of crops that are not modeled by crop models from a small set of crops that is modeled (Müller and Robertson 2014, Nelson *et al* 2014) thus needs to be challenged, even though there may be little alternative under current constraints on data availability. The next round of AgMIP/ISIMIP future projections (Jägermeyr *et al* in prep) also aims at broadening the scope of simulated crops, but many models are not available for less ubiquitously grown crops.

For short- and mid-term projections, GGCM-induced variance dominates the overall variance

across all scenarios and crops, except for soybean, where crop models generally contribute only a small share to overall variance and where also overall variance is relatively low (2nd after maize). This dominance of the GGCM signal in the first half of the 21st century is likely because of the relatively small differences in radiative forcing in this period, which is also largely independent of the RCP trajectory (van Vuuren *et al* 2011).

Future crop yields are determined by counteracting drivers. Climate change impacts (warming, changes in precipitation) lead to overall negative impacts on crop yields that amplify unequivocally with the radiative forcing at the global aggregation level. However, the main cause of climate change, increasing atmospheric CO<sub>2</sub> concentrations from anthropogenic emissions, also lead to increased crop productivity. There is substantial uncertainty connected to the effects of CO<sub>2</sub> fertilization in models, especially at high concentrations as projected for the end of the 21st century under RCP8.5, where also little experimental evidence can guide model parameterization and development (Toreti *et al* 2020). Nonetheless, the modeled response to elevated atmospheric CO<sub>2</sub> concentrations requires more attention from the modeling community.

In this analysis, we focused on changes in the CTW dimensions of the emulated CTNW-A experiment (Franke *et al* 2020a, 2020b), ignoring the N dimension, which can also contribute to overall uncertainty. We kept N inputs at historical patterns across regions and crops (Elliott *et al* 2015) throughout the simulations. It is plausible to assume that N fertilization would change under changing crop yield potentials, market access and dynamics, or environmental regulation. To our knowledge, there are no such projections available, especially not any that would account for the changes in potential yields under the multitude of climate projections used here. Long-term crop projections here do not account for other technical and management changes in addition to N, which additionally artificially suppresses the crop model-induced component of uncertainty. This is somewhat analogous to the ‘pathway’ uncertainty in the SSP-RCP framework.

Still, we find that the GGCM ensemble contributes relatively little to overall variance in regions with intensive agriculture (supplementary figure S14) as well as for soybean (a N fixing plant) more generally, suggesting that the response to N inputs is also an important driver of uncertainty in crop yield projections. The relationship between nutrient limitations and susceptibility to climate change impacts as well as how nutrient limitation is modeled at different levels of nutrient supply need further scrutiny.

Our results for the end of the 21st century need to be interpreted with some caveats, as we had to cap CTW drivers to the training domain of the CTNW-A experiment (Franke *et al* 2020a), because of the



non-linear functional form of the emulators (Franke *et al* 2020b), which makes extrapolation beyond the training domain volatile and error-prone. For the majority of GCMs and harvested areas, this is not a major caveat as most areas do not exceed +6 K. However, for some GCMs, especially under CMIP6, large fractions of the crops' harvested areas exceed the +6 K warming level (supplementary figure S2). This leads to an artificial reduction of the GCM-induced variance in results. The plausibility of the very high ECS in climate projections has been challenged (Tokarska *et al* 2020) and the ensemble could be pruned on this basis to avoid very warm climate projections. However, the selection of climate scenarios provided to climate impact modeling community in e.g. ISIMIP does not necessarily follow such pre-selection approaches and we thus kept the full CMIP6 archive here. The saturating overall variance that can be observed towards the end of the 21st century could suggest that the capping of the warming at +6 K leads to an artificial reduction of the end-of-the-century variance, however we observe the same general feature (steepest increase in variance in mid-century) also in the other RCPs that are not subject to the capping of temperature signals as warming levels are generally lower (figure 5). The observed saturation of variance towards the end of the 21st century cannot be attributed to a saturation in drivers of climate change as global mean cropland temperatures under RCP8.5 show no sign of levelling off (supplementary figure S7) and also (CO<sub>2</sub>) and radiative forcing do no level off under RCP8.5 (van Vuuren *et al* 2011).

Generally, climate and crop models should be selected on a fit-for-purpose basis. While the climate community has established the standard that the same model versions that provide future projections also provide historical simulations for evaluation purposes, this procedure has not generally been adopted by the crop modeling community. The ISIMIP project is promoting a similar structure in the individual simulation rounds (Frieler *et al* 2017), but crop models need to more rigorously provide meta information on the model version and parameterization, which can greatly affect simulated dynamics (Folberth *et al* 2019). The common practice to reduce the uncertainty space by selecting a small number of climate scenarios by e.g. first availability has already been challenged by Mcsweeney and Jones (2016). We show that, at the global scale, the selection of individual crop models can greatly affect the outcomes and even the exclusion of one out of an ensemble of nine can have substantial effects on results. This pulls the general assumption into question if we can consider all GGCM projections as equally plausible, or if the skewed distribution suggests that some models should indeed be excluded prior to the interpretation of ensemble results. More and also different GGCMs are expected to contribute to the new

round of global crop model simulations of AgMIP and ISIMIP (Jägermeyr *et al* in prep). However, it may not necessarily be desirable to increase the ensemble size to a point where the exclusion of sub-samples no longer affects the overall ensemble response if the unbalanced ensemble may be caused by inclusion of non-plausible projections.

Thus, we call for intensified efforts to understand why crop models differ and to build strategies on how models can be improved—or that lead to a better understanding why it is plausible to have an as broad distribution as our current full ensemble suggests. While better model agreement is not an appropriate aim in itself, model disagreement can be used to identify aspects for coordinated model improvement, as e.g. described by Maiorano *et al* (2017). Also, the assessment of crop models based on their ability to reproduce spatial and temporal patterns of historical crop yields (Müller *et al* 2017) needs to be expanded by plausibility tests in individual model components and processes. Given that crop yields are determined by many interacting processes (Schauberger *et al* 2016), which have not been all implemented or sufficiently tested in crop models (Boote *et al* 2013), we need to do everything possible to minimize the chance of getting the right answer for the wrong reason as shown e.g. by Zhu *et al* (2019) for maize yields in the USA. As such, model performance needs to be also assessed at the level of individual processes before errors in these can mutually cancel out and are not traceable in the yield projections.

Toreti *et al* (2020) call for a set of standard tests on crop models' response to elevated (CO<sub>2</sub>) that should be made accessible as meta-data for each model. Building on this idea, we call for a set of standard tests for crop models across all major drivers of crop yield simulations (CO<sub>2</sub>, temperatures, precipitation, nutrients, management aspects) with respect to single driver effects as well as with respect to their interaction. The CTWN-A experiment (Franke *et al* 2020a) that also covers more crop growth metrics than just yields, provides a suitable basis for such tests, even though the computational requirements are too high to qualify for a standard test.

Protocols for such standard model tests need to be developed in close collaboration with experimentalists as they need to reflect the evolving understanding of physiological processes, and need to include more aspects than just end-of-season yields. Even though global crop model results are difficult to compare to data from experimental sites (Deryng *et al* 2016), global (and field-scale) crop models need to be tested at the site level for plausible response types (e.g. direction of change) and ranges (e.g. size of effects). The comparison of global crop model results with site data has been shown to allow for ex-post corrections of the range of simulated crop yield projections (Wang *et al* 2020).

## 5. Conclusions

We find that future crop yield projections are subject to substantial uncertainties. These increase with the radiative forcing, i.e. over time and also with the emission pathway considered. Crop model-induced uncertainty dominates the overall uncertainty in the first half of the projections for the 21st century and more efforts are needed to improve crop model skill and testing procedures. In the second half of the 21st century, the overall uncertainty surges, mainly driven by a steeper increase of uncertainty from climate models. Long-term projections are thus of mainly academic value that can help to derive insights from comparing scenarios and assumptions but should not be confused with predictions of future developments. This is especially true as modifications in management that can be expected to be implemented by farmers are often ignored due to a lack of data on management systems and missing tools to project these into the future. The unbalanced nature of the crop model ensemble, where often individual models strongly affect the overall ensemble behavior call for intensified research on climate change impact modeling for agriculture. This has been pleaded for by Rötter *et al* (2011) before and the various activities in AgMIP, MAC-SUR, ISIMIP and elsewhere have helped to move in that direction. Still, more efforts are needed, especially with respect to model evaluation standards and testing of other aspects than crop yields, as e.g. by Kimball *et al* (2019).






## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: (<https://doi.org/10.5281/zenodo.4321276>).

## Acknowledgments

CMIP5 model output data provided by the WHOI CMIP5 Community Storage Server, Woods Hole Oceanographic Institution, Woods Hole, MA, USA from their website at <http://cmip5.whoi.edu/>. J A F was supported by the NSF NRT program (Grant No. DGE-1735359) and the NSF Graduate Research Fellowship Program (Grant No. DGE-1746045). RDCEP is funded by NSF through the Decision Making Under Uncertainty program (Grant No. SES-1463644). JJ received support from the Open Philanthropy Project and the NASA Co-op. ACR was supported by NASA GISS Climate Impacts Group funding from the NASA Earth Sciences Directorate. PF and KW were supported by the Newton Fund through the Met Office program Climate Science for Service Partnership Brazil (CSSP Brazil). The publication of this article was partially funded by the Open Access Fund of the Leibniz Association.

## ORCID iDs

Christoph Müller  <https://orcid.org/0000-0002-9491-3550>  
 James Franke  <https://orcid.org/0000-0001-8598-750X>  
 Jonas Jägermeyr  <https://orcid.org/0000-0002-8368-0018>  
 Alex C Ruane  <https://orcid.org/0000-0002-5582-9217>  
 Pete D Falloon  <https://orcid.org/0000-0001-7567-8885>  
 Christian Folberth  <https://orcid.org/0000-0002-6738-5238>  
 Tobias Hank  <https://orcid.org/0000-0001-7491-0291>  
 R César Izaurralde  <https://orcid.org/0000-0002-8797-9500>  
 Wenfeng Liu  <https://orcid.org/0000-0002-8699-3677>  
 Stefan Olin  <https://orcid.org/0000-0002-8621-3300>  
 Thomas A M Pugh  <https://orcid.org/0000-0002-6242-7371>  
 Karina Williams  <https://orcid.org/0000-0002-1185-535X>  
 Florian Zabel  <https://orcid.org/0000-0002-2923-4412>

## References

- Akinsanola A A, Kooperman G J, Reed K A, Pendergrass A G and Hannah W M 2020 Projected changes in seasonal precipitation extremes over the United States in CMIP6 simulations *Environ. Res. Lett.* **15** 104078
- Almazroui M, Saeed F, Saeed S, Nazrul Islam M, Ismail M, Klutse N A B and Siddiqui M H 2020 Projected change in temperature and precipitation over Africa from CMIP6 *Earth Syst. Environ.* **4** 455–75
- Asseng S *et al* 2013 Uncertainty in simulating wheat yields under climate change *Nat. Clim. Change* **3** 827–32
- Boote K J, Jones J W, White J W, Asseng S and Lizaso J I 2013 Putting mechanisms into crop production models *Plant. Cell Environ.* **36** 1658–72
- Deryng D *et al* 2016 Regional disparities in the beneficial effects of rising CO<sub>2</sub> concentrations on crop water productivity *Nat. Clim. Change* **6** 786–90
- Dury M, Hambuckers A, Warnant P, Henrot A, Favre E, Ouberdous M and Francois L 2011 Responses of European forest ecosystems to 21st century climate: assessing changes in interannual variability and fire intensity *iForest Biogeosci. For.* **4** 82–99
- Elliott J *et al* 2015 The global gridded crop model intercomparison: data and modeling protocols for phase 1 (v1.0) *Geosci. Model. Dev.* **8** 261–77
- Elliott J, Kelly D, Chryssanthacopoulos J, Glotter M, Jhunjhunwala K, Best N, Wilde M and Foster I 2014 The parallel system for integrating impact models and sectors (pSIMS) *Environ. Model. Softw.* **62** 509–16
- Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization *Geosci. Model. Dev.* **9** 1937–58
- Fan X, Miao C, Duan Q, Shen C and Wu Y 2020 The performance of CMIP6 versus CMIP5 in simulating temperature

- extremes over the global land surface *J. Geophys. Res. Atmos.* **125** e2020JD033031
- FAO 2001 *Food Balance Sheets: A Handbook* (Rome: FAO)
- Folberth C *et al* 2019 Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble *PLoS One* **14** e0221862
- Folberth C, Gaiser T, Abbaspour K C, Schulin R and Yang H 2012 Regionalization of a large-scale crop growth model for sub-Saharan Africa: model setup, evaluation, and estimation of maize yields *Agric. Ecosyst. Environ.* **151** 21–33
- Franke J A *et al* 2020a The GGCM phase 2 experiment: global gridded crop model simulations under uniform changes in CO<sub>2</sub>, temperature, water, and nitrogen levels (protocol version 1.0) *Geosci. Model. Dev.* **13** 2315–36
- Franke J A *et al* 2020b The GGCM phase 2 emulators: global gridded crop model responses to changes in CO<sub>2</sub>, temperature, water, and nitrogen (version 1.0) *Geosci. Model. Dev.* **13** 3995–4018
- Freund M B, Brown J R, Henley B J, Karoly D J and Brown J N 2020 Warming patterns affect El Niño diversity in CMIP5 and CMIP6 models *J. Clim.* **33** 8237–60
- Frieler K *et al* 2017 Assessing the impacts of 1.5 °C global warming—simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b) *Geosci. Model. Dev.* **10** 4321–45
- Hank T B, Bach H and Mauser W 2015 Using a remote sensing-supported hydro-agroecological model for field-scale simulation of heterogeneous crop growth and yield: application for wheat in central Europe *Remote Sens.* **7** 3934–65
- Hawkins E and Sutton R 2009 The potential to narrow uncertainty in regional climate predictions *Bull. Am. Meteorol. Soc.* **90** 1095–107
- Izaurrealde R C, Williams J R, McGill W B, Rosenberg N J and Jakas M C Q 2006 Simulating soil C dynamics with EPIC: model description and testing against long-term data *Ecol. Modell.* **192** 362–84
- Jägermeyr J *et al* 2017 Climate change signal in agriculture emerges earlier in new generation of projections In preparation
- Jiang D, Hu D, Tian Z and Lang X 2020 Differences between CMIP6 and CMIP5 models in simulating climate over China and the East Asian monsoon *Adv. Atmos. Sci.* **37** 1102–18
- Jones J W *et al* 2017 Brief history of agricultural systems modeling *Agric. Syst.* **155** 240–54
- Kimball B A *et al* 2019 Simulation of maize evapotranspiration: an inter-comparison among 29 maize models *Agric. For. Meteorol.* **271** 264–84
- Liu W, Yang H, Folberth C, Wang X, Luo Q and Schulin R 2016 Global investigation of impacts of PET methods on simulating crop-water relations for maize *Agric. For. Meteorol.* **221** 164–75
- Maïorano A *et al* 2017 Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles *Field Crops Res.* **202** 5–20
- Mcsweeney C F and Jones R G 2016 How representative is the spread of climate projections from the 5 CMIP5 GCMs used in ISI-MIP? *Clim. Serv.* **1** 24–29
- Meehl G A, Senior C A, Eyring V, Flato G, Lamarque J-F, Stouffer R J, Taylor K E and Schlund M 2020 Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models *Sci. Adv.* **6** eaba1981
- Minoli S *et al* 2019 Global response patterns of major rainfed crops to adaptation by maintaining current growing periods and irrigation *Earth's Future* **7** 1464–80
- Mueller N D, Gerber J S, Johnston M, Ray D K, Ramankutty N and Foley J A 2012 Closing yield gaps through nutrient and water management *Nature* **490** 254–7
- Müller C *et al* 2017 Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications *Geosci. Model. Dev.* **10** 1403–22
- Müller C and Robertson R 2014 Projecting future crop productivity for global economic modeling *Agric. Econ.* **45** 37–50
- Nelson G C *et al* 2014 Climate change effects on agriculture: economic responses to biophysical shocks *Proc. Natl Acad. Sci.* **111** 3274–9
- O'Neill B C *et al* 2016 The Scenario Model Intercomparison Project (scenariomip) for CMIP6 *Geosci. Model. Dev.* **9** 3461–82
- Olin S, Schurgers G, Lindeskog M, Wärlind D, Smith B, Bodin P, Holmér J and Arneth A 2015 Modelling the response of yields and tissue C: N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat regions of western Europe *Biogeosciences* **12** 2489–515
- Palosuo T *et al* 2011 Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models *Eur. J. Agron.* **35** 103–14
- Portmann F T, Siebert S and Döll P 2010 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
- Riahi K, Rao S, Krey V, Cho C, Chirkov V, Fischer G, Kindermann G, Nakicenovic N and Rafaj P 2011 RCP 8.5—a scenario of comparatively high greenhouse gas emissions *Clim. Change* **109** 33–57
- Rosenzweig C *et al* 2013 The Agricultural Model Intercomparison and Improvement Project (AgMIP): protocols and pilot studies *Agric. For. Meteorol.* **170** 166–82
- Rosenzweig C *et al* 2014 Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison *Proc. Natl Acad. Sci.* **111** 3268–73
- Rosenzweig C *et al* 2018 Coordinating AgMIP data and models across global and regional scales for 1.5 °C and 2.0 °C assessments *Phil. Trans. R. Soc. A* **376** 20160455
- Rötter R P, Carter T R, Olesen J E and Porter J R 2011 Crop-climate models need an overhaul *Nat. Clim. Change* **1** 175–7
- Ruane A C *et al* 2018 Biophysical and economic implications for agriculture of +1.5° and +2.0 °C global warming using AgMIP coordinated global and regional assessments *Clim. Res.* **76** 17–39
- Ruane A C, Goldberg R and Chryssanthacopoulos J 2015 Climate forcing datasets for agricultural modeling: merged products for gap-filling and historical climate series estimation *Agric. For. Meteorol.* **200** 233–48
- Schauberger B, Rolinski S and Müller C 2016 A network-based approach for semi-quantitative knowledge mining and its application to yield variability *Environ. Res. Lett.* **11** 123001
- Taylor K E, Stouffer R J and Meehl G A 2012 An overview of CMIP5 and the experiment design *Bull. Am. Meteorol. Soc.* **93** 485–98
- Tokarska K B, Stolpe M B, Sippel S, Fischer E M, Smith C J, Lehner F and Knutti R 2020 Past warming trend constrains future warming in CMIP6 models *Sci. Adv.* **6** eaaz9549
- Toreti A *et al* 2020 Narrowing uncertainties in the effects of elevated CO<sub>2</sub> on crops *Nat. Food* **1** 775–82
- van Vuuren D *et al* 2011 The representative concentration pathways: an overview *Clim. Change* **109** 5–31
- von Bloh W, Schaphoff S, Müller C, Rolinski S, Waha K and Zaehle S 2018 Implementing the nitrogen cycle into the dynamic global vegetation, hydrology, and crop growth model LPJml (version 5.0) *Geosci. Model. Dev.* **11** 2789–812
- Wang X *et al* 2020 Emergent constraint on crop yield response to warmer temperature from field experiments *Nat. Sustain.* **908–16**
- Williams K, Gornall J, Harper A, Wiltshire A, Hemming D, Quaife T, Arkebauer T and Scooby D 2017 Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska *Geosci. Model. Dev.* **10** 1291–320

- Wirsenius S 2000 Human use of land and organic materials  
*Department of Physical Resource Theory* (Göteborg:  
Chalmers University of Technology and Göteborg  
University) p 255
- Wyser K, van Noije T, Yang S, von Hardenberg J, O'Donnell D and  
Döscher R 2020 On the increased climate sensitivity in the  
EC-Earth model from CMIP5 to CMIP6 *Geosci. Model. Dev.*  
[13](#) 3465–74
- Xin X, Wu T, Zhang J, Yao J and Fang Y 2020 Comparison of  
CMIP6 and CMIP5 simulations of precipitation in China  
and the East Asian summer monsoon *Int. J. Climatol.*  
[40](#) 6423–40
- Zhu P, Zhuang Q, Archontoulis S V, Bernacchi C and Müller C  
2019 Dissecting the nonlinear response of maize yield to  
high temperature stress with model-data integration *Glob.*  
*Change Biol.* [25](#) 2470–84
- Zhu Y Y and Yang S 2020 Evaluation of CMIP6 for historical  
temperature and precipitation over the Tibetan Plateau and  
its comparison with CMIP5 *Adv. Clim. Change Res.*  
[11](#) 239–51