

# The GGCMI Phase II experiment: global gridded crop model simulations under uniform changes in CO<sub>2</sub>, temperature, water, and nitrogen levels (protocol version 1.0)

James Franke<sup>1,2</sup>, Christoph Müller<sup>3</sup>, Joshua Elliott<sup>2,4</sup>, Alex C. Ruane<sup>5</sup>, Abigail Snyder<sup>6</sup>, Jonas Jägermeyr<sup>3,2,4,5</sup>, Juraj Balkovic<sup>7,8</sup>, Philippe Ciais<sup>9,10</sup>, Marie Dury<sup>11</sup>, Pete Falloon<sup>12</sup>, Christian Folberth<sup>7</sup>, Louis François<sup>11</sup>, Tobias Hank<sup>13</sup>, Munir Hoffmann<sup>14,23</sup>, R. Cesar Izaurralde<sup>15,16</sup>, Ingrid Jacquemin<sup>11</sup>, Curtis Jones<sup>15</sup>, Nikolay Khabarov<sup>7</sup>, Marian Koch<sup>14</sup>, Michelle Li<sup>2,17</sup>, Wenfeng Liu<sup>9,18</sup>, Stefan Olin<sup>19</sup>, Meridell Phillips<sup>5,20</sup>, Thomas A. M. Pugh<sup>21,22</sup>, Ashwan Reddy<sup>15</sup>, Xuhui Wang<sup>9,10</sup>, Karina Williams<sup>12</sup>, Florian Zabel<sup>13</sup>, and Elisabeth Moyer<sup>1,2</sup>

<sup>1</sup>Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

<sup>2</sup>Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

<sup>3</sup>Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany

<sup>4</sup>Department of Computer Science, University of Chicago, Chicago, IL, USA

<sup>5</sup>NASA Goddard Institute for Space Studies, New York, NY, United States

<sup>6</sup>Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

<sup>7</sup>Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

<sup>8</sup>Department of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic

<sup>9</sup>Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

<sup>10</sup>Sino-French Institute of Earth System Sciences, College of Urban and Env. Sciences, Peking University, Beijing, China

<sup>11</sup>Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium

<sup>12</sup>Met Office Hadley Centre, Exeter, United Kingdom

<sup>13</sup>Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany

<sup>14</sup>Georg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modeling, Göttingen, Germany

<sup>15</sup>Department of Geographical Sciences, University of Maryland, College Park, MD, USA

<sup>16</sup>Texas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

<sup>17</sup>Department of Statistics, University of Chicago, Chicago, IL, USA

<sup>18</sup>EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

<sup>19</sup>Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

<sup>20</sup>Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

<sup>21</sup>School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.

<sup>22</sup>Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK.

<sup>23</sup>Leibniz Centre for Agricultural Landscape Research (ZALF), D-15374 Müncheberg, Germany

**Correspondence:** Christoph Müller (cmueller@pik-potsdam.de)

**Abstract.** Concerns about food security under climate change motivate efforts to better understand future changes in crop yields. Process-based crop models, which represent plant physiological processes, are necessary tools for this purpose since they allow representing future climate and management conditions not sampled in the historical record and new locations where cultivation may shift. However, process-based crop models differ in many critical details, and their responses to different interacting factors remain only poorly understood. The Global Gridded Crop Model Intercomparison (GGCMI) Phase II

experiment, an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), is designed to provide a systematic parameter sweep across critical interacting factors, to allow both evaluating model behavior and emulating model responses in impact assessment tools. In this paper we describe the GGCMI Phase II experimental protocol and its simulation data archive. Twelve crop models simulate five crops in simulations with systematic uniform perturbations of historical  
5 climate, varying CO<sub>2</sub>, temperature, precipitation, and applied nitrogen (“CTWN”) for rainfed and irrigated agriculture, and a second set of simulations represents adaptation by allowing adjusted planting dates. We show some crop yield results to illustrate general characteristics of the simulations and potential uses of the GGCMI Phase II archive. For example, modeled yields show robust decreases to warmer temperatures in almost all regions, with a nonlinear dependence that means yields in warmer baseline locations have greater temperature sensitivity. Inter-model uncertainty is qualitatively similar across all the  
10 four input dimensions, but is largest in high-latitude regions where crops may be grown in the future.

## 1 Introduction

Understanding crop yield response to a changing climate is critically important, especially as the global food production system will face pressure from increased demand over the next century (Bodirsky et al., 2015). Climate-related reductions in supply could therefore have severe socioeconomic consequences (e.g. Stevanović et al., 2016; Wiebe et al., 2015). Multiple studies  
15 using different crop or climate models concur in projecting sharp yield reductions on currently cultivated cropland under business-as-usual climate scenarios, although their yield projections show considerable spread (e.g. Rosenzweig et al., 2014; Schauberger et al., 2017; Porter et al. (IPCC), 2014, and references therein). Although forecasts of future yields reductions can be made with simple statistical models based on regressions in historical weather data, process-based models, which simulate the process of photosynthesis and the biology and phenology of individual crops, play a critical role in assessing the impacts  
20 of climate change.

Process-based models are necessary for understanding crop yields in novel conditions not included in historical data, including higher [CO<sub>2</sub>] levels, out-of-sample combinations of rainfall and temperature, cultivation in areas where crops are not currently grown, and differing management practices (e.g. Pugh et al., 2016; Roberts et al., 2017; Minoli et al., 2019). Process-based models have therefore been widely used in studies on future food security (Wheeler and Von Braun, 2013),  
25 options for climate mitigation (Müller et al., 2015) and adaptation (Challinor et al., 2018), and future sustainable development (Humpenöder et al., 2018). Process-based models also allow the globally gridded simulations needed for understanding the global dynamics of agricultural trade, including cultivation area changes and crop selection switching under climate change (Rosenzweig et al., 2018; Ruane et al., 2018), because global market mechanisms may strongly modulate climate change impacts (Stevanović et al., 2016; Hasegawa et al., 2018), global crop model experiments are needed for systematic climate change  
30 assessments (Müller et al., 2017).

Modeling crop responses continues however to be challenging, as crop growth is a function of complex interactions between climate inputs and management practices (J. Boote et al., 2013; Rötter et al., 2011). Models tend to agree broadly in major response patterns, including a reasonable representation of the spatial pattern in historical yields of major crops (e.g. Elliott

et al., 2015; Müller et al., 2017) and projections of shifts in yield under future climate scenarios. But process-based models still struggle with some important details, including reproducing historical year-to-year variability in many regions (e.g. Müller et al., 2017), reproducing historical yields when driven by reanalysis weather (e.g. Glotter et al., 2014), and low sensitivity to extreme events (e.g. Glotter et al., 2015; Jägermeyr and Frieler, 2018; Schewe et al., 2019). Long-term projections therefore 5 retain considerable uncertainty (Wolf and Oijen, 2002; Jagtap and Jones, 2002; Iizumi et al., 2010; Angulo et al., 2013; Asseng et al., 2013, 2015).

Model intercomparison projects such as the Agricultural Model Intercomparison and Improvement Project (AgMIP, Rosenzweig et al., 2013) are crucial in quantifying uncertainties in model projections (Rosenzweig et al., 2014). Intercomparison projects have also been used to develop protocols for evaluating overall model performance (Elliott et al., 2015; Müller et al., 10 2017) and to assess the representation of individual physical mechanisms such as water stress and [CO<sub>2</sub>] fertilization (e.g. Schauberger et al., 2017). However, to date, few such projects have systematically sampled critical factors that may interact strongly in affecting crop yields. A number of modeling exercises in the last five years have begun to use systematic parameter sweeps in crop model evaluation and emulation (e.g. Ruane et al., 2014; Makowski et al., 2015; Pirttioja et al., 2015; Fronzek et al., 2018; Snyder et al., 2018; Ruiz-Ramos et al., 2018), but all involve limited sites and most also limited crops and 15 scenarios.

The Global Gridded Crop Model Intercomparison (GGCMI) Phase II experiment is the first globally gridded crop model intercomparison involving systematic parameter sweep across critical interacting factors. GGCMI Phase II is an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), and a continuation of a multi-model comparison exercise begun in 2014. The initial GGCMI Phase I compared harmonized yield simulations over the historical period, with a 20 primary goals of model evaluation and understanding sources of uncertainty (including model parameterization, weather inputs, and cultivation areas) (Elliott et al., 2015; Müller et al., 2017; Folberth et al., 2016; Porwollik et al., 2017). GGCMI Phase II compares simulations across a set of inputs with uniform perturbations to historical climatology, including CO<sub>2</sub>, temperature, precipitation, and applied nitrogen (collectively referred to as “CTWN”), as well as adaptation to shifting growing seasons. The CTWN experiment is inspired by AgMIP’s Coordinated Climate-Crop Modeling Project (C3MP Ruane et al., 2014; 25 McDermid et al., 2015) and contributes to the AgMIP Coordinated Global and Regional Assessments (CGRA) (Ruane et al., 2018; Rosenzweig et al., 2018).

In this paper, we describe the GGCMI Phase II model experiments and present initial summary results. In the sections that follow, we describe the experimental goals and protocols; the different process-based models included in the intercomparison; the levels of participation by the individual models. We then provide an assessment of model fidelity based on observed yields 30 at the country level, and show some selected examples of the simulation output dataset to illustrate model responses across the input dimensions.

## 2 Simulation objectives and protocol

### 2.1 Goals

The guiding scientific rationale of GGCMI Phase II is to provide a comprehensive, systematic evaluation of the response of process-based crop models to critical interacting factors, including [CO<sub>2</sub>], temperature, water, and applied nitrogen (CTWN).

5 The dataset is designed to allow researchers to:

- Enhance understanding of how models work by characterizing their sensitivity to input climate and nitrogen drivers.
- Study the interactions between climate variables and nitrogen inputs in driving modeled yield impacts.
- Explore differences in crop response to warming across the Earth’s climate regions.
- Provide a dataset that allows statistical emulation of crop model responses for downstream modelers.

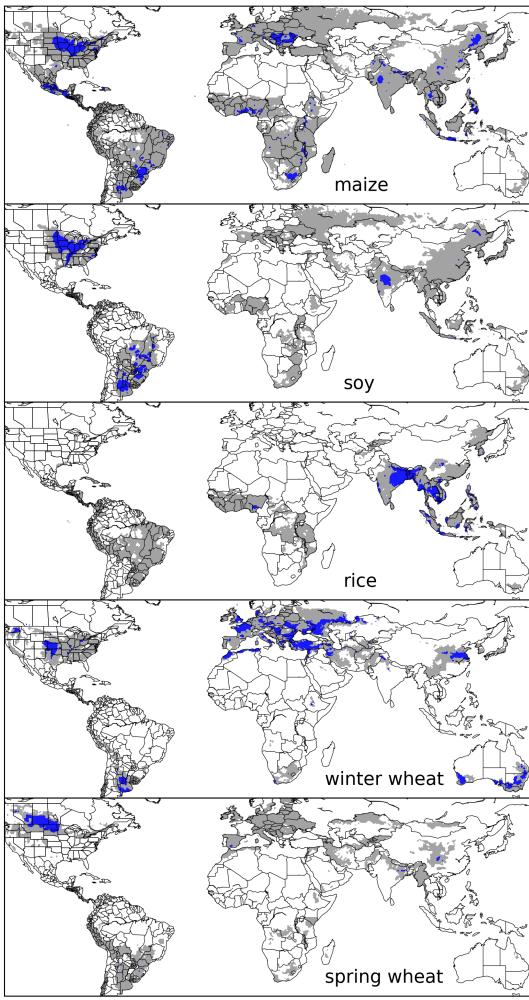
10 – Illustrate differences in potential adaptation via growing season changes.

### 2.2 Modeling protocol

The GGCMI Phase I intercomparison was a relatively limited computational exercise, requiring yield simulations for 19 crops across a total of 310 model-years of historical scenarios, and had the participation of 21 modeling groups. The GGCMI Phase II protocol is substantially larger, involving over 1400 individual 30-year scenarios, or over 42,000 model-years; 12 modeling

15 groups nevertheless participated. To reduce the computational load, the GGCMI Phase II protocol reduced the number crops to 5 (maize, rice, soybean, spring wheat, and winter wheat). The reduced set of crops includes the three major global cereals and the major legume and accounts for over 50% of human calories (in 2016, nearly 3.5 billion tons or 32% of total global crop production by weight) (Food and Agriculture Organization of the United Nations, 2018). This set of major crops has the advantage of historical yield data at sub-national scale across the globe (Ray et al., 2012; Iizumi et al., 2014), and has been  
20 frequently used in agricultural analyses (e.g. Müller et al., 2017; Porwollik et al., 2017).

The Phase II protocol involves a suite of uniform perturbations from a historical weather scenario. The baseline climate scenario for GGCMI Phase II is one of the weather products used in Phase I, daily climate inputs for the years period 1980-2010 from the 0.5 degree NASA AgMERRA (“Agricultural”-modified Modern Era Retrospective analysis for Research and Applications) gridded re-analysis product. AgMERRA is specifically designed for agricultural modeling, with satellite-corrected precipitation (Ruane et al., 2015). The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for [CO<sub>2</sub>], and 3 for applied nitrogen, for a total of 756 simulations, 672 for rainfed agriculture and additional 84 for irrigated (Table 1). For irrigated simulations, models are set to apply near-perfect irrigation to keeps soils wet throughout the entire growing period, with no limitations in water supply. Values of climate variable perturbations are selected to represent reasonable ranges for changes over the medium term (to 2100) under business-as-usual emissions. The resulting GGCMI Phase  
30 II dataset therefore captures the distribution of crop model responses over the space of future potential climate conditions.



**Figure 1.** Presently cultivated area for rainfed crops. Blue indicates grid cells with more than 20,000 hectares ( $\sim 10\%$  of an equatorial grid cell). Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soybean are taken from the MIRCA2000 (“Monthly Irrigated and Rainfed Crop Areas around the year 2000”) dataset (Portmann et al., 2010). Areas for winter and spring wheat areas are adapted from MIRCA2000 and two other sources; see text for details. For analogous figure of irrigated crops, see Figure ??.

While all perturbations are applied uniformly across the historical timeseries, they are applied in different ways. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell. Precipitation perturbations are applied as fractional changes at the grid cell level.  $[CO_2]$  and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Perturbations are applied independently and the protocol samples over all possible permutations. This choice means that  $[CO_2]$  changes are applied independently of changes in climate variables, so that higher  $[CO_2]$  is not associated with particular climate changes, e.g. higher temperatures.

**Table 1.** GGCMI Phase II input parameter levels for each dimension. Temperature and precipitation values indicate the perturbations from the historical climatology. W-percentage does not apply to the irrigated ( $W_{inf}$ ) simulations, which are simulated at the maximum beneficial levels of water. Bold font indicates the ‘baseline’ or historical level for each dimension. Adaptation dimension (A1) repeats all (A0) simulations except the growing season in held fixed. One model provided simulations at the T + 5 level. See Figure S3 in the supplement for number of simulations associated with each combination of input levels.

Input variable	Simulation input values	Unit
[CO <sub>2</sub> ] (C)	<b>360</b> , 510, 660, 810	ppm
Temperature (T)	-1, <b>0</b> , 1, 2, 3, 4, 6	°C
Precipitation (W)	-50, -30, -20, -10, <b>0</b> , 10, 20, 30, (and $W_{inf}$ )	%
Applied nitrogen (N)	10, 60, <b>200</b>	kg ha <sup>-1</sup>
Adaptation (A)	<b>A0: none</b> , A1: new cultivar to maintain original growing season length	-

Each model is run at 0.5 degree spatial resolution and covers all currently cultivated areas and much of the uncultivated land area. See Figure 1 for the present-day cultivated area of rainfed crops, provided by the MIRCA2000 (Monthly Irrigated and Rainfed Crop Area) data product (Portmann et al., 2010), and Supplementary Figure S1 for irrigated crops. Coverage extends considerably outside currently cultivated areas because cultivation will likely shift under climate change. To reduce the computational burden, however, the protocol requires simulation only over **Christoph – XX%** of Earth surface area. Areas are not simulated if they are assumed to remain non-arable even under an extreme climate change; these regions include Greenland, far-northern Canada, Siberia, Antarctica, the Gobi and Sahara Deserts, and central Australia. The protocol also eliminates regions judged unsuitable for cropland for non-climatic reasons. Selection criterion involve a combination of soil suitability indices at 10 arc-minute resolution and excludes those 0.5 degree grid cells in which at least 90% of the area is masked as unsuitable according to any single index, and which do not contain any currently cultivated cropland. Soil suitability indices measure excess salt, oxygen availability, rooting conditions, toxicities, and workability, and are provided by the IIASA (International Institute for Applied Systems Analysis) Global Agro-Ecological Zone model (GAEZ, FAO/IIASA, 2011). The procedure follows that proposed by Pugh et al. (2016). All modeling groups simulate the minimum required coverage, but some provide simulations that extend into masked zones, including e.g. the Sahara Desert and Central Australia. See Figure SX in the supplemental for spatial coverage of simulations.

### 2.3 Harmonization between models

The 12 models included in GGCMI Phase II are all process-based crop models that are widely used in impacts assessments (Table 3). Although some models share a common base (e.g. the LPJ family or the EPIC family of models), they have sub-

sequently developed independently. Wherever possible, the GGCMI Phase II protocol harmonizes inputs, but differences in model structure mean that several key factors cannot be fully standardized across the experiment. These include soil treatment (which affects soil organic matter and carry-over effects of soil moisture across growing years) and baseline climate inputs.

While 10 of the 12 models participating in GGCMI Phase II use the AgMERRA historical daily climate data product, two models require sub-daily input data and thus use different baseline climate inputs: PROMET uses the ERA-Interim reanalysis (Dee et al., 2011); JULES uses a bias-corrected version of ERA-Interim, the 3-hour WFDEI (WATCH-Forcing-Data-ERA-Interim) (Weedon et al., 2014), selecting the WFDEI version in which precipitation is bias-corrected against the CRU TS3.101/TS3.21 precipitation totals(Harris et al., 2014). The WFDEI reanalysis differs slightly from AgMERRA while the ERA-Interim differs more substantially from AgMERRA. Figures ?? and ?? show temperature and precipitation in the three data products over currently cultivated areas for each crop. At this aggregation level, temperatures are very similar between data products, ERA-Interim being about 1.0 and 0.3°C cooler in rice and maize growing areas, respectively. For precipitation, ERA-Interim is substantially wetter in wheat areas (+60mm year<sup>-1</sup>) and also in some years for rice, maize and soybean areas. These differences are still relatively small compared to the perturbations tested in the protocol.

Planting dates and growing season lengths are standardized across models, following the procedure described in Elliott et al. (2015) for the *fullharm* setting. In contrast to GGCMI Phase I (Elliott et al., 2015), we here assume identical growing seasons for rainfed and irrigated scenarios, to allow for direct comparability of simulations along the W dimension, in which irrigation ( $W_{inf}$ ) is one element (see Table 1). While sowing dates are prescribed directly in models, the length of the growing season is a product of crop phenology, which in turn is mostly driven by phenological parameters and temperature. Modelers are asked to adjust the phenological parameters so that growing season length on average matches the harmonization target. Given that temperature varies between years, individual years can vary from the harmonization target. Harmonization of growing seasons is crop- and location-specific, i.e. they vary in space and per crop but not across models. For example, at present maize is sown in March in Spain, in July in Indonesia, and in December in Namibia (Portmann et al., 2010). One exception is CARAIB, which did not harmonize against provided growing season data (Elliott et al., 2015), but kept their own growing seasons.

To roughly account for the importance of adaptation in agricultural production, the GGCMI Phase II protocol includes two sets of experiments that sample adaptation. As simulated growing seasons respond to temperatures, these are sensitive to changes along the T dimension in the CTWN experiment. For adaptation, a fifth dimension “A” was added, for adaptation in phenological traits. The first set (“A0”, where 0 denotes “no adaptation”) involves growing seasons with unmodified, harmonized phenological parameters, which generally result in shorter growing seasons in warmer scenarios. The second set, (“A1”, with adaptation) holds the length of the growing season fixed at that of the baseline climate scenario. For this, modelers had to repeat the baseline calibration of growing seasons length at all temperature levels along the T dimension (Table 1). Even though CARAIB did not harmonize the growing seasons to GGCMI targets, their “A1” simulations follow the same principle, so that phenological parameters are modified to keep growing seasons roughly constant across different warming scenarios. These “A1” simulations roughly capture the case in which adaptive crop cultivar choice maintain the growing season length so that crops reach maturity at roughly the same time as do current varieties under the current temperature regime. This assumption is

simplistic, and does not reflect realistic opportunities and limitations to adaptation (Vadez et al., 2012; Challinor et al., 2018), but provides some insight into how crop modifications could alter projected impacts on yields.

Growing seasons for maize, rice, and soybean are taken from the SAGE (Center for Sustainability and the Global Environment, University of Wisconsin) crop calendar (Sacks et al., 2010) and are identical to those used in GGCMI Phase I (Elliott et al., 2015). In GGCMI Phase II, we separately treat spring and winter wheat and so must define different growing seasons for each. We use the SAGE crop calendar, which separately specifies spring and winter wheat, as the primary source for **JIM – XX% of grid cells**. In the remaining areas where no SAGE information is available, we turn to, in order of preference, the MIRCA2000 crop calendar (Portmann et al., 2010) and to simulated LPJmL growing seasons (Waha et al., 2012). These datasets each provide several options for wheat growing season for each grid cell, but do not label them as spring or winter wheat. We assign a growing season to each wheat type for each location based on its baseline climate conditions. A growing seasons is assigned to winter wheat if all of the following hold:

- the monthly mean temperature is below freezing point ( $<0^{\circ}\text{C}$ ) at most for 5 months per year (i.e. winter is not too long)
- the coldest 3 months of a year are below  $10^{\circ}\text{C}$  (i.e. there is a winter)
- the season start date fits the criteria that
  - if in the N. hemisphere, it is after the warmest *or* before the coldest month of the year (as winter is around the end/beginning of the calendar year)
  - if in the S. hemisphere, it is after the warmest *and* before the coldest month of the year (as winter is in the middle of the calendar year)

and to spring wheat otherwise.

Nitrogen application is also standardized in timing across models. N fertilizer is applied in two doses, as is often the norm in actual practice, to reduce losses to the environment. In the GGCMI Phase II protocol, half of the total fertilizer input is applied at sowing and the other half on day 40 after sowing, for all crops except for winter wheat. For winter wheat, in practice the application date for the second N fertilizer application varies according to local temperature, because the length of winter dormancy can vary strongly. In the GGCMI Phase II protocol, the second fertilization date for winter wheat is set to the middle day of the first month after sowing that has average temperatures above  $5^{\circ}\text{C}$ , with hard limits between 40 days from planting and 50 from maturity.

All stresses are disabled in models other than those related to nitrogen, temperature, and water. For example, model responses to alkalinity, salinity, and non-nitrogen nutrients are all disabled. For a better controlled experiment, no other external N inputs are permitted – that is, there is no atmospheric deposition of nitrogen – but some models allow additional release of plant-available nitrogen through mineralization in soils. Soil mineralization is a part of model treatments of soil organic matter and cannot be disabled in some models (e.g. LPJmL, LPJ-GUESS). Some additional differences in model structure mean that several key factors are not standardized across the experiment. For example, carry-over effects across growing years including residue management and soil moisture are treated differently across models.

## 2.4 Output data products

All models in GGCMI Phase II provide 7 mandatory output variables (Table 2, bold), if available. For each scenario, 0.5 degree grid cell and crop, models provide 30-year timeseries of annual crop yields in units of tons ha<sup>-1</sup> year<sup>-1</sup>, as well as total aboveground biomass yield; the dates of planting, anthesis, and maturity; applied irrigation water in irrigated scenarios;

5 and total evapotranspiration. (Note that several of the EPIC-family models do not output the anthesis date.) Besides these mandatory 7 data products, the protocol requests any or all of 18 optional additional output variables (Table 2, plain text). Participating modeling groups provided between 3 (PEPIC) and 18 (APSIM-UGOE) of these optional variables.

All output data is supplied as netCDF version 4 files, each containing values for one variable in a 30-year timeseries associated with a single scenario, for all grid cells. Filenames follow the naming conventions of GGCMI Phase I (Elliott et al., 2015),

10 which themselves are taken from those of ISIMIP (Frieler et al., 2017). File names are specified (in small caps) as

[model]\_[climate]\_hist\_fullharm\_[irrig.scenario]\_[variable]\_[crop]\_global\_annual\_[start-year]\_[end-year].nc4

Here [model] is the crop model name; [climate] is the original climate input data set (typically AgMERRA); [irrig.scenario] is the irrigation setting ("fир" for fully irrigated and "noirr" for fully rainfed); [variable] is the output variable (of those in Table 2); [crop] is the crop abbreviation ("mai" for maize, "ric" for rice, "soy" for soybean, "swh" for spring wheat, and "wwh"

15 for winter wheat); and [start - year] and \_[end - year] specify the first and last years recorded on file. All filenames include the identifier *global* to distinguish them as global model output.

Output data is provided on a regular geographic grid, identical for all models. Grid cell centers span latitudes -89.75 to 89.75° and longitudes from -179.75 to 179.75°. Missing values where no crop growth has been simulated are distinguished from crop failures: a crop failure is reported as zero yield but non-simulated areas (including ocean grid cells) have yields 20 reported as 1.e+20. Following NetCDF standards, latitude, longitude and time are included as separate variables in ascending order, with units "degrees north", "degrees east", and "growing seasons since 1980-01-01 00:00:00".

Following GGCMI Phase I standards, the first entry in the file is the result of the first simulated cropping cycle that is entirely within the given climate input. For AgMERRA, where the first year provided is 1980, the first harvest record is thus of 1980,

when the prescribed sowing and harvest dates are in 1980 (e.g. sowing in March and harvest in September 1980) but is of 25 1981 if sowing is later in a calendar year than harvest (e.g. sowing in September 1980 and harvest in March 1981). To avoid distortions in harvest events, output files report the sequence of growing periods rather than calendar years. In most cases,

this is equivalent, as there is always only one sowing event per calendar year. As harvest events are internally determined as a function of mostly temperature, these can vary between individual years. If harvest events are around the end of the calendar year (Dec. 31), reported values could contain 2 (one in early January and one in late December), 1 (normal) or none (last was

30 in December of previous year and next is in January of the following year) harvest event if reported per calendar year. As such, the time dimension in the netCDF files is not on calendar years, but on "growing seasons" (séé above).

**Table 2.** Output variables requested per crop in the GGCMI Phase II protocol. Items in **bold** are mandatory, if available in that model.

Variable	variable name	units
<b>Yield</b>	<b>yield_&lt;crop&gt;</b>	<b>t ha<sup>-1</sup> yr<sup>-1</sup> (dry matter)</b>
<b>Total above ground biomass yield</b>	<b>biom_&lt;crop&gt;</b>	<b>t ha<sup>-1</sup> yr<sup>-1</sup> (dry matter)</b>
<b>Actual planting date</b>	<b>plant-day_&lt;crop&gt;</b>	<b>day of year</b>
<b>Anthesis date</b>	<b>anth-day_&lt;crop&gt;</b>	<b>days from planting</b>
<b>Maturity date</b>	<b>maty-day_&lt;crop&gt;</b>	<b>days from planting</b>
<b>Applied irrigation water</b>	<b>pirrww_&lt;crop&gt;</b>	<b>mm yr<sup>-1</sup></b>
<b>Evapotranspiration (growing season sum)</b>	<b>etransp_&lt;crop&gt;</b>	<b>mm yr<sup>-1</sup> (firr scenarios only)</b>
Transpiration (growing season sum)	transp_<crop>	mm yr <sup>-1</sup>
Evaporation (growing season sum)	evap_<crop>	mm yr <sup>-1</sup>
Runoff (total growing season sum, subsurface + surface)	runoff_<crop>	mm yr <sup>-1</sup>
Total available soil moisture in root zone *	trzpah2o_<crop>	mm yr <sup>-1</sup>
Total root biomass	rootm_<crop>	t ha <sup>-1</sup> yr <sup>-1</sup> (dry matter)
Total Nr uptake (total growing season sum)	tnrup_<crop>	kg ha <sup>-1</sup> yr <sup>-1</sup>
Total Nr inputs (total growing season sum)	t nrin_<crop>	kg ha <sup>-1</sup> yr <sup>-1</sup>
Total Nr losses (total growing season sum)	t nrloss_<crop>	kg ha <sup>-1</sup> yr <sup>-1</sup>
Gross primary production (GPP)	gpp_<crop>	gC m <sup>-2</sup> yr <sup>-1</sup>
Net primary production (NPP)	npp_<crop>	gC m <sup>-2</sup> yr <sup>-1</sup>
CO <sub>2</sub> response scaler on NPP	co2npp_<crop>	- {0..inf}
water response scaler on NPP	h2onpp_<crop>	- {0..1}
temperature response scaler on NPP	tnpp_<crop>	- {0..1}
Nr response scaler on NPP	n rnpp_<crop>	- {0..1}
Other nutrient response scaler on NPP	or npp_<crop>	- {0..1}
CO <sub>2</sub> response scaler on transpiration	co2trans_<crop>	- {0..1}
maximum stress response scaler	maxstress_<crop>	- {0..1}
Maximum LAI	laimax_<crop>	m <sup>2</sup> m <sup>-2</sup>

\* growing season sum, basis for computing average soil moisture

**Table 3.** Models included in GGCMI Phase II and the number of C, T, W, and N simulations that each performs, with 756 as the maximum for A0 (no adaptation) and 648 as the maximum for A1 (maintaining growing season adaptation; in A1, T0 is skipped as there is no adaptation to temperature-driven shortening of growing seasons there). “N-Dim.” indicates whether the simulations include varying nitrogen levels. Two models provide only one nitrogen level. All models provide the same set of simulations across all modeled crops, but some omit individual crops. (For example, APSIM does not simulate winter wheat.)

Model (Key Citations)	Maize	Soybean	Rice	Winter wheat	Spring wheat	N dim.	Sims per crop (A0 / A1)
<b>APSIM-UGOE</b> , Keating et al. (2003); Holzworth et al. (2014)	X	X	X	–	X	X	44 / 36
<b>CARAIB</b> , Dury et al. (2011); Pirttioja et al. (2015)	X	X	X	X	X	–	<b>252 / 216</b>
<b>EPIC-IIASA</b> , Balkovič et al. (2014)	X	X	X	X	X	X	39 / 0
<b>EPIC-TAMU</b> , Izaurralde et al. (2006)	X	X	X	X	X	X	<b>756 / 648</b>
<b>JULES</b> , Osborne et al. (2015); Williams and Falloon (2015); Williams et al. (2017)	X	X	X	–	X	–	<b>252 / 0</b>
<b>GEPIC</b> , Liu et al. (2007); Folberth et al. (2012)	X	X	X	X	X	X	430 / 181
<b>LPJ-GUESS</b> , Lindeskog et al. (2013); Olin et al. (2015)	X	–	–	X	X	X	<b>756 / 648</b>
<b>LPJmL</b> , von Bloh et al. (2018)	X	X	X	X	X	X	<b>756 / 648</b>
<b>ORCHIDEE-crop</b> , Wu et al. (2016)	X	–	X	X	–	X	33 / 0
<b>pDSSAT</b> , Elliott et al. (2014); Jones et al. (2003)	X	X	X	X	X	X	<b>756 / 756</b>
<b>PEPIC</b> , Liu et al. (2016a, b)	X	X	X	X	X	X	149 / 121
<b>PROMET</b> , Hank et al. (2015); Mauser et al. (2015)	X	X	X	X	X	X	261 / 232
Totals	12	10	11	11	11	10	5240   3378

### 3 Models contributing

The contributions of the 12 crop models supplying data to the GGCMI Phase II archive are described in Table 3. Modeling groups did not all implement the full protocol described above. Given the substantial computational requirements, different participation tiers were specified to allow submission of smaller sub-sets of the full protocol. These subsets were designed as

alternate samples across the 4 dimensions of the CTWN space, with *full* (12) and *low* (4) options for the C × N variables, and *full* (63), *reduced* (31), and *minimum* (9) options for T × W variables (described below). All participating modeling groups provided identical coverage of the CTWN parameter space for different crops, but some provided no or a more limited sample for scenarios with adaptation (A1). This is, as the adaptation dimension is deemed a side-aspect of the GGCMI Phase II experiment that is only analyzed in specific applications. The different participation levels are defined by combining the CxN sets with the TxW sets:

- **full**: all 756 A0 simulations (all 12 CxN \* all 63 TxW)
- **high**: 362 simulations (all 12 CxN combinations \* *reduced* TxW set of 31 combinations)
- **mid**: 124 simulations (*low* 4 CxN combinations \* *reduced* TxW set of 31 combinations)
- **low**: 36 simulations (*low* 4 CxN combinations \* *minimum* TxW set of 9 combinations)

10

Of the 12 models submitting data, 6 followed the *full* protocol; these are marked with italic text in Table 3. However, note that two of these models (CARAIB and JULES) intrinsically cannot represent nitrogen effects and so do not sample over the the nitrogen dimension at all. 2 models followed *high* with minor modifications (GEPIC adding an additional T level and PROMET omitting the intermediate N level). 1 model (PEPIC) followed *mid* but included an additional C level. 3 models approximately 15 followed *low* with APSIM-UGOE and EPIC-IIASA providing some additional levels and ORCHIDEE-crop omitting a TxW combination.

The combinations of perturbation values in the CxN and TxW parameter spaces used in the various participation levels are chosen to provide maximum coverage over plausible future values. For the CxN space, we specify:

- *full* as 12 pairs, with 4 C values (360, 660, 810 ppm) and 3 N (10, 60, 200 kg ha<sup>-1</sup> yr<sup>-1</sup> ))
- *low* as only 4 pairs: C360\_N10, C360\_N200, C660\_N60, C810\_N200

20

For the TxW space we specify:

- *full* as all 7 T levels and 9W levels.
- *reduced* as 31 alternating combinations, with different Ws for even Ts than for odd Ts. For even Ts (i.e. T0,T2,T4,T6), we use W = -50,-20,0,+30 = 4·4 = 16 pairs. For odd Ts (i.e. T-1,T1,T3) , we use W = -30, -10, +10, +30, inf = 3·5 = 15 pairs.
- *minimum* as 9 combinations: T-1W-10, T0W10, T1W-30, T2W-50, T2W20, T3W30, T4W0, T4Winf, T6W-20

25

## 4 Results

To illustrate the properties of the GGCMI Phase II model simulations, we provide both an assessment of model performance by comparing to observed yields, and selected case studies showing the spread of model responses to climate and management 30 inputs.

#### 4.1 Assessment of model performance

Evaluating the performance of crop models in the GGCMI Phase II archive is complicated by the artificial nature of the protocol: the settings in the CTWN-A experiment design do not reflect actual conditions in the real world. The protocol includes one scenario of near-historical climate inputs ( $T_0$ ,  $W_0$ ,  $C_{360}$ ), but the prescribed uniform nitrogen application levels

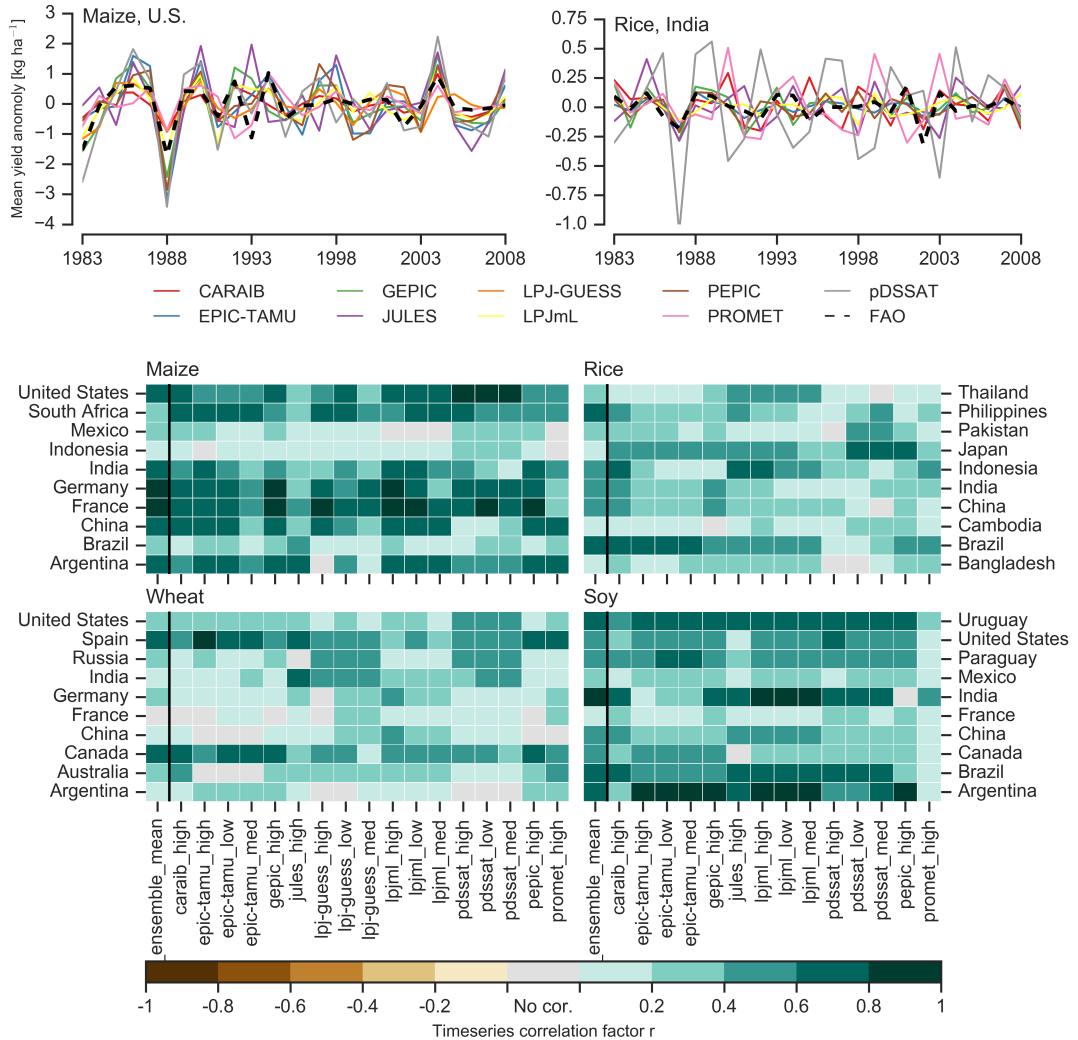
5 do not reflect real-world fertilizer practices. Models also omit detailed calibrations to reflect the performance of historical cultivars.

We provide a partial evaluation of the models' skill in reproducing crop yield characteristics using the methodology of Müller et al. (2017), who described model characteristics in GGCMI Phase I. Müller et al. (2017) evaluate how well model crop yield responses in historical run capture real-world yield variations driven by year-to-year temperature and precipitation  
10 variations. Following this approach, we compare yields in the GGCMI Phase II baseline run with detrended historical yields from the UN Food and Agriculture Organization (FAO) (Food and Agriculture Organization of the United Nations, 2018) by calculating the Pearson product moment correlation coefficient. The procedure is sensitive to the detrending method and the area mask used to aggregate yields; we use a 5-year running mean removal and the MIRCA2000 cultivation area mask for aggregation. In some cases the model time series are shifted by one year to account for errors in FAO or model year reporting.  
15 Because the GGCMI Phase II protocol imposes fixed, uniform nitrogen application levels that are not realistic for individual countries, we evaluate control runs for each model and at multiple N levels in some cases. Nine of the GGCMI Phase II models provide historical runs for all three nitrogen levels (10, 60, and. 200 kg ha<sup>-1</sup> yr<sup>-1</sup>).

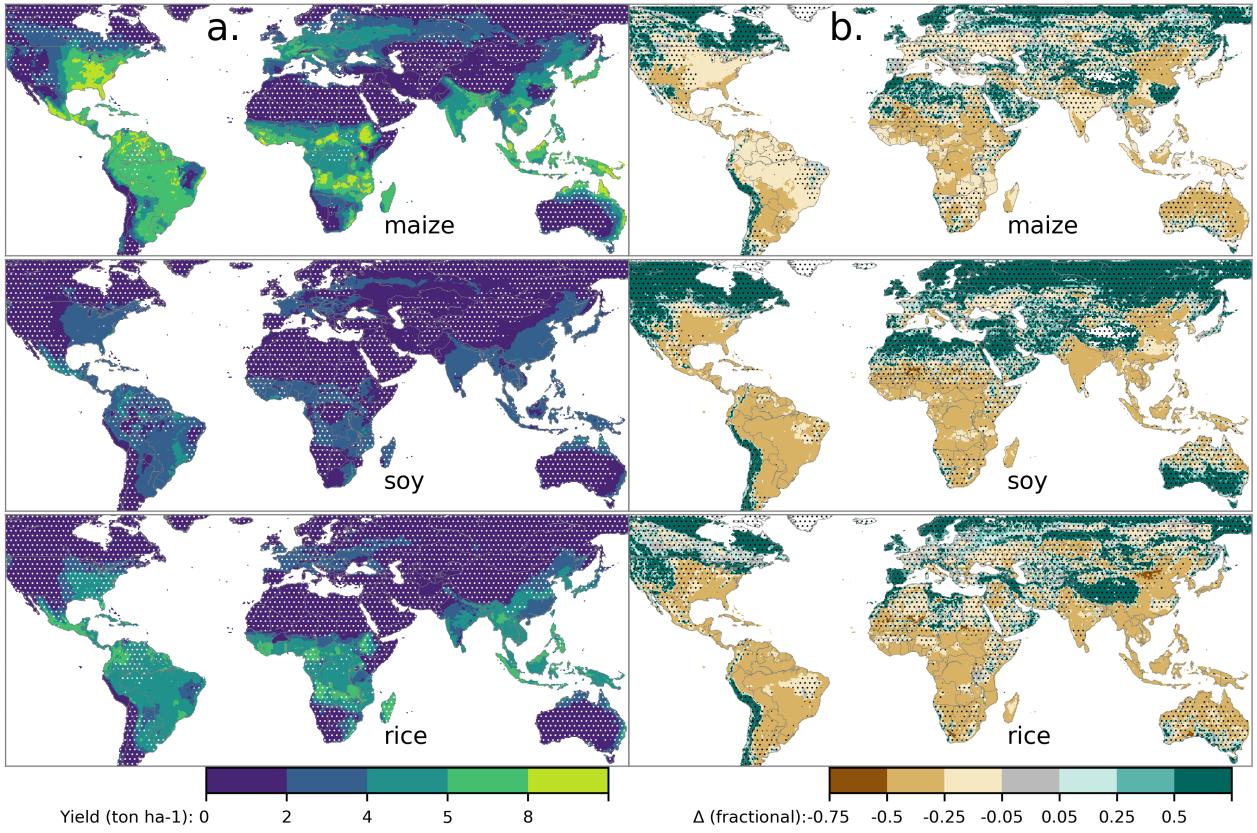
As expected, correlation coefficients are slightly lower than those found in the GGCMI Phase I evaluation, but models show reasonable fidelity at capturing year-over-year variation (Figure 2). [JIM – give a few numbers here](#) (Compare to Müller  
20 et al. (2017) Figures 1–4 and 6.) In GGCMI Phase II, as in Phase I, differences in fidelity between regions and crops exceed differences between models: that is, Figure 2 shows more color similarity in horizontal than vertical bars. For example, maize in the United States is consistently well-simulated while maize in Indonesia is problematic (mean Pearson correlation coefficients of 0.68 and 0.18, respectively). Note that in this methodology, simulations of crops with low year-to-year variability such as irrigated rice and wheat will tend to score more poorly than those with higher variability. In some cases, especially in the  
25 developing world, low correlation coefficients may indicate not model failure but problems in FAO yield data (Ray et al., 2012; Müller et al., 2017). No single model consistently exhibits greater fidelity than others. Instead, each model shows near best-in-class performance for at least one location-crop combination. For example for maize pDSSAT is the best model in the US, LPJmL and GEPIC are the best in Germany, PROMET is best in Argentina, PEPIC and LPJ-GUESS are the best in France, and CARAIB is near best in class in South Africa.

#### 30 4.2 Model crop yield responses under CTWN forcing

Crop models in the GGCMI Phase II ensemble show broadly consistent responses to climate and management perturbations in most regions, with a strong negative impact of increased temperature in all but the coldest regions. Mapping the distribution of baseline yields and yield changes shows the geographic dependencies that underlie these results. Absolute yield potentials

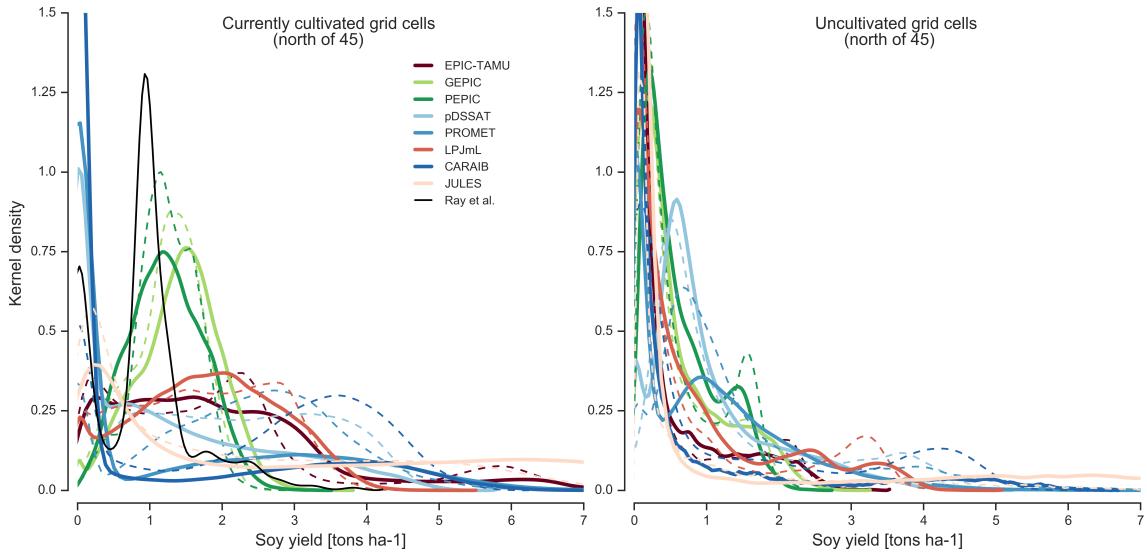


**Figure 2.** Assessment of crop model performance in GGCMI Phase II, following protocol of GGCMI Phase I (Müller et al., 2017). **Top:** example time series comparison between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level for two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application level. **Bottom:** the heatmaps illustrate the Pearson  $r$  correlation coefficient between the detrended simulation mean yield at the country level compared to the detrended FAO yield data for the top producing countries for each crop with continuous FAO data over the 1981–2010 period. Models that provided different nitrogen application levels are shown with low ( $10 \text{ kg N ha}^{-1}$ ), med ( $60 \text{ kg N ha}^{-1}$ ), and high ( $200 \text{ kg N ha}^{-1}$ ) label (models that did not simulate different nitrogen levels are analogous to a high nitrogen application level). The ensemble mean yield, show in the leftmost column, is also correlated with the FAO data (not the mean of the correlations). Wheat contains both spring wheat and winter wheat simulations where supplied, else one or the other (see Table 3). Differences by region and crop are stronger than difference between models, e.g. horizontal bars are more similar in color than vertical bars.



**Figure 3.** Illustration of the spatial pattern of potential yields (left) and yield changes (right) in the GGCMI Phase II ensemble, for three major crops: maize, soybean and rice. Left column (a) shows multi-model mean climatological potential yields for the baseline scenario for (top-bottom) rainfed maize, soybean, and rice. Wheat shows a qualitatively similar response, see Figure SXX in the supplemental material. White stippling on the left panels indicates areas where these crops are not currently cultivated. Absence of cultivation aligns well with the lowest yield contour ( $0\text{--}2 \text{ ton ha}^{-1}$ ). Right column (b) shows the multi-model mean fractional yield change in the  $T+4 \text{ }^{\circ}\text{C}$  scenario. Because yields vary geographically, we show change from a baseline, which we take as the scenario with historical climatology (i.e. T and P changes of 0), C of 360 ppm, and applied N at  $200 \text{ kg N ha}^{-1}$ . In  $T+4$  only temperature is changed and other inputs remain fixed at baseline values. Stippling indicates areas of low confidence ( $\Delta < 1\sigma$ ) and areas without stippling are those where confidence in projections is higher: the multi-model mean fractional change exceeds the standard deviations of the ensemble change ( $\Delta > \sigma$ ). Crop model simulation results in cold areas, where yield impacts are on average positive, also have the highest uncertainty. Spatial coverage is not consistent in number of models and explains some of the patterns in regions where crops are not currently cultivated. See Figure SXX in the supplemental for spatial coverage of simulations for the  $T+4$  case.

show strong spatial variation, with much of the Earth's surface area unsuitable for any of these crops ((Figure 3), left). Crop yield changes under climate perturbations also show distinct geographic pattern (Figure 3, right, which shows fractional yield differences between the  $T+4$  scenario and the baseline scenario with historical climatology). In general, models agree most

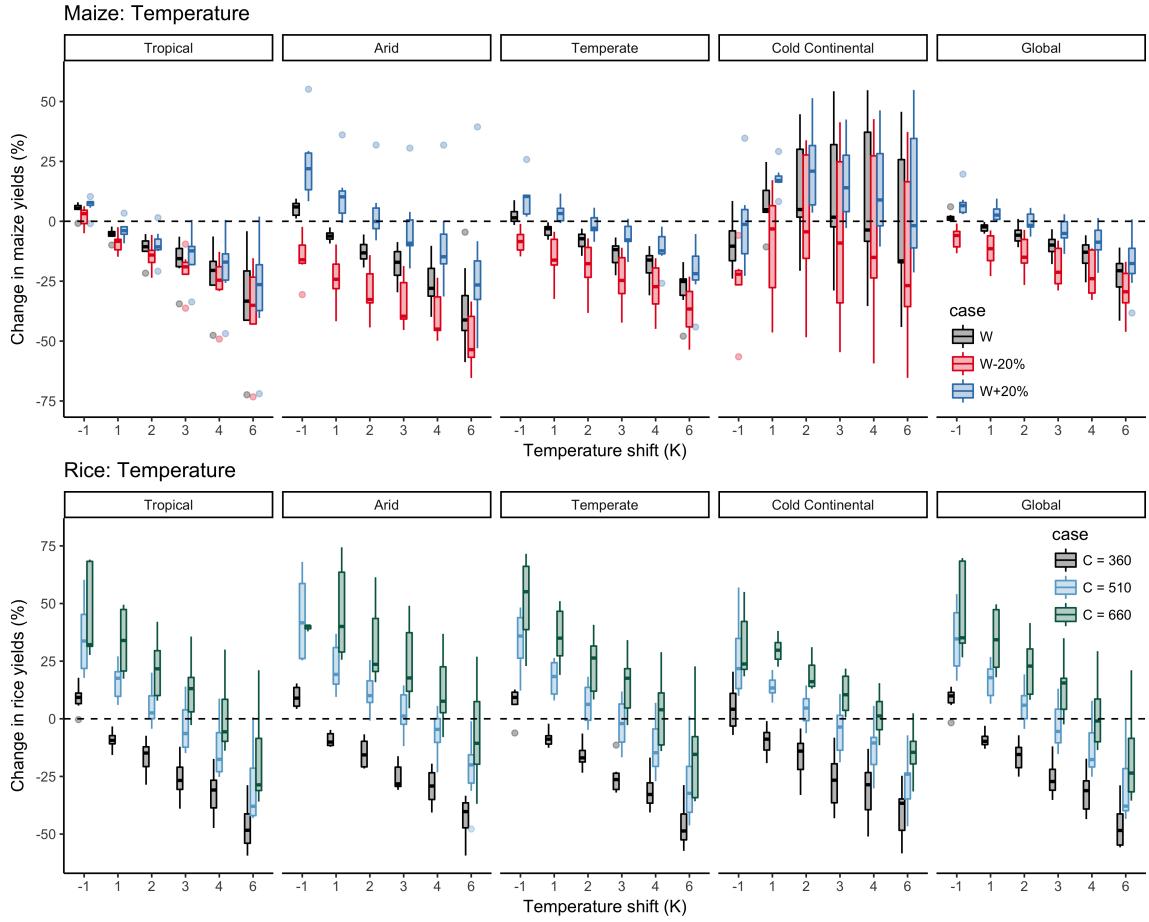


**Figure 4.** Probability density estimate of soybean yields at high latitude for selected models. Probability density functions are estimated across all grid cells north of  $45^{\circ}$  latitude for currently cultivated areas (left panel, approximately 2500 grid cells) and across uncultivated grid cells that were simulated by all 9 models shown here (right panel, approximately 1500 grid cells). The solid lines show the historical(1980-2010) climatology and dashed lines show the T+4 (K) case. Models vary widely on the representation of soy yields at this latitude even for areas where the crop is currently cultivated (left panel). ‘Green’ category models show steep distributions between 1 - 2 (ton ha<sup>-1</sup>) with slight decreases under. ‘Red’ category models show relatively flatter distributions with mixed shifts under warming. ‘Blue’ category models show mostly 0 (ton ha<sup>-1</sup>) yields for the historical climatology and large increases under a 4 degree warming. The solid black line indicates the probability density function for the mean 1997-2003 yield values from the data from Ray et al. (2012). Yield values greater than 6 (ton ha<sup>-1</sup>) are not actually achieved in reality. Models agree to a greater extent about the unsuitability for soy where it is not currently cultivated (right panel) and less so about the change under warming.

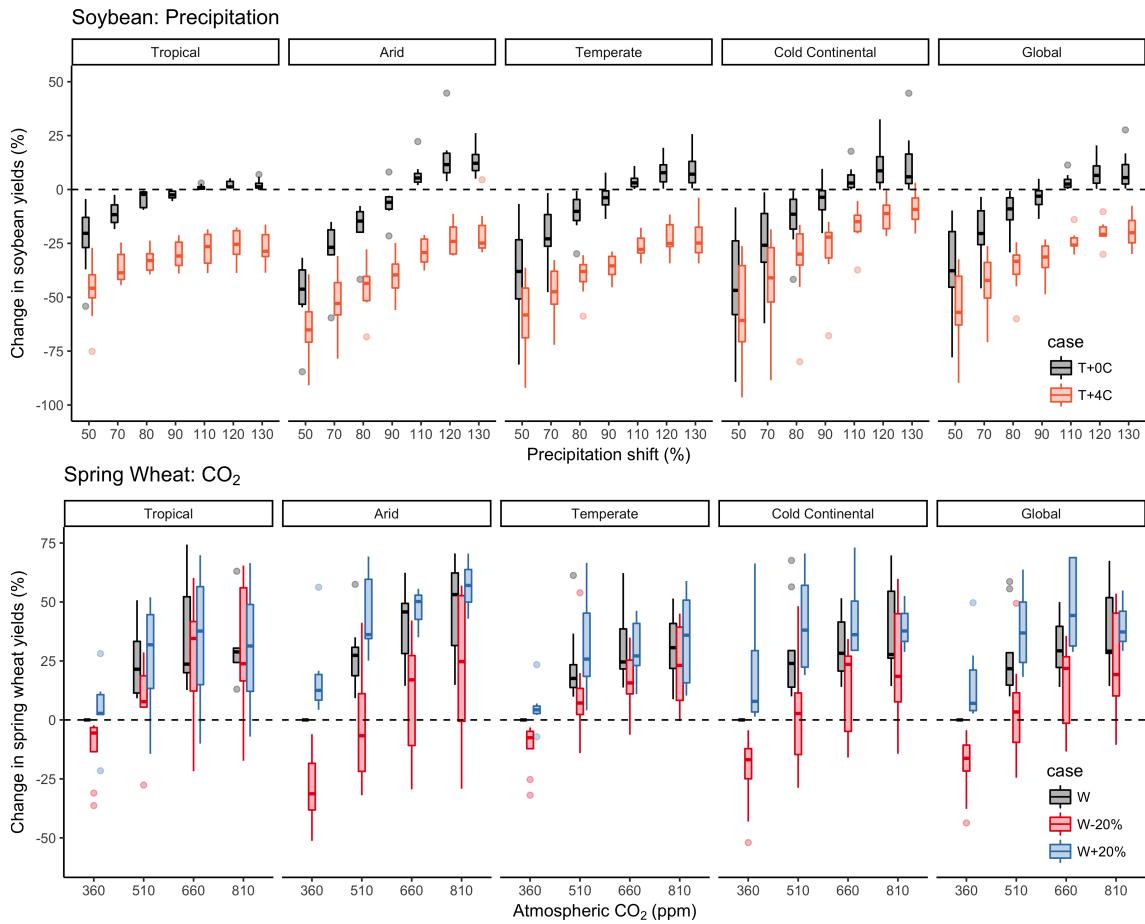
on yield response in regions where yield potentials are currently high and therefore where crops are currently grown. Models show robust decreases in yields at low latitudes, and highly uncertain median increases at most high latitudes.

The larger inter-model spread in yield projections at higher latitudes evident in Figure 3 is due in part to how crop suitability or crop failures differ across models. In many cases the spread results from differences in the present-day scenario rather than the future one. For example, three models in Phase II show almost all 0 (ton ha<sup>-1</sup>) yields north of  $45^{\circ}$  for the historical period and steep increases in yields under a 4 degree warming even for areas currently under cultivation (4, left panel). Other models (PEPIC and GEPIC) show yields between 1-2 (ton ha<sup>-1</sup>) in the historical climate and a slight decrease in yield under a warming scenario. PEPIC and GEPIC reproduce actual yields (Ray et al., 2012) most faithfully. Models agree more about the unsuitability of soy in locations north of  $45^{\circ}$  where it is not currently cultivated, but the low baseline yields result in large and unstable fractional changes under warming.

NOT SURE WHERE TO PUT THESE sentences, all kind of tacked on at moment –XXX The density functions shown in Figure 4 are not weighted by cultivation area. This makes direct comparison to national level statistics difficult because



**Figure 5.** Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel and Kottek, 2010). Y-axis is the fractional change in the regional average climatological (30-year mean) potential yield relative to the baseline. Box-and-whiskers plots show distribution across models, with median marked; edges are first and third quartiles; whiskers extend to 1.5-IQR. Figure shows all modeled land area within each model; see Figure SXX-SXX in the supplemental material for only currently-cultivated land. The right panel (Global) shows yield responses to a globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios. Note that Rubel and Kottek (2010) use the name ‘Snow’ rather than ‘Cold continental’. **Top:** responses for rainfed maize to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (-20%, 0%, and +20%), with  $[CO_2]$  and nitrogen held constant at baseline values (360 ppm and  $200 \text{ kg ha}^{-1} \text{ yr}^{-1}$ ). Outliers in the tropics (strong negative impact of temperature increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Outside high-latitude regions for maize, models generally agree, with projected declines under increasing temperatures larger than inter-model variance. **Bottom:** Temperature response for irrigated rice for three discrete  $[CO_2]$  levels, with nitrogen and precipitation held constant.  $[CO_2]$  does not change the nature of temperature response respective to baseline as the slopes at each  $[CO_2]$  level are relatively constant. There is very little difference across Köppen-Geiger regions for irrigated rice.



**Figure 6.** Illustration of the distribution of regional yield changes across the multi-model ensemble, here for soybeans and spring wheat. Conventions as in Figure 5. **Top:** response for rainfed soy to applied uniform precipitation perturbations, for two discrete temperature levels, with [CO<sub>2</sub>] and nitrogen held constant at baseline levels. Inter model spread increases for reduced precipitation cases more so than increased precipitation. A leveling out on the increased precipitation side illustrates the saturation effect. Precipitation changes are more important in dryer Köppen-Geiger regions, as expected. Increased temperature tends to flatten the precipitation response in a relative sense because of some prioritizing of the two stresses. **Bottom:** response for rainfed spring wheat to atmospheric [CO<sub>2</sub>], for three discrete precipitation perturbation levels with temperature and nitrogen held constant at baseline values. The [CO<sub>2</sub>] Reduced precipitation tends to steepen the [CO<sub>2</sub>] response and increased precipitation tends to flatten it, as expected. Reduced precipitation tends to increase the inter model spread, especially at the highest [CO<sub>2</sub>] levels.

cultivation tends to weight towards the more favorable conditions. Mean yield weighted by growing area in Canada/Northern US in the Ray et al. (2012) data is 2.2 ton ha<sup>-1</sup> for the period 1997-2003 and 1.8 in Russia/ Eastern Europe. No equivalent real world data exist for uncultivated grid cells. Inter-model spreads are largest in wheat projections (Figure SXX), possibly because calibration is most important for wheat (e.g. Asseng et al., 2013).

We illustrate the across-model spread for selected crops in Figures 5 - 6, which shows changes in yields across all simulated grid cells for the primary Köppen-Geiger climate regions (Rubel and Kottek, 2010) for various input dimensions. In warming scenarios with precipitation held constant, all models show decreases in maize yield in the ‘warm temperate’, ‘equatorial’, and ‘arid’ regions that account for nearly three-quarters of global maize production. These impacts are robust for even moderate 5 climate perturbations. In the ‘warm temperate’ zone, even a 1 degree temperature rise with other variables held fixed leads to a median yield reduction that exceeds the variance across models. A 6 degree temperature rise results in median loss of  $\sim 25\%$  of yields with a signal to noise ratio of nearly three to one. A notable exception is the ‘cold continental’ region, where models disagree strongly, extending even to the sign of impacts. The temperature response is qualitatively similar across all 10 crops included in this study (Figures SX - X). Increased [CO<sub>2</sub>] boosts yields overall but does little to change the nature of the temperature response for e.g. irrigated rice (Figure 5, bottom panel). Other crops show similar responses to warming, with robust yield losses in warmer locations and high inter-model variance in the ‘cold continental’ regions (Figure SX).

The effects of rainfall changes on yields are also relatively consistent across models (maize: Figure 5, soybeans: Figure 6). Increased rainfall mitigates the negative effect of higher temperatures by counteracting the increased evapo-transpiration to some degree, most strongly in arid regions. Decreased rainfall amplifies yield losses and also increases inter-model variance; 15 e.g. models agree that the response to decreased water availability is negative in sign but disagree on its magnitude. Increased temperature results in a relative flattening of the precipitation change response (Figure 6, top panel). We show only rainfed maize and soy here; see Figure SX for comparison between rainfed and irrigated maize. As expected, irrigated crops are more resilient to temperature increases in all regions, especially so where water is already limiting. The other crops in this study show a qualitatively similar response to changes in precipitation (Figures SX - X).

20 Mean climatological yield response to the other two GGCMI Phase II dimensions (C and N) are qualitatively similar across crops. The yield response to increased [CO<sub>2</sub>] for spring wheat is a robust increase across models in all climate regions (Figure 6, bottom panel). Increased precipitation allows crops to capture additional yield boost from elevated [CO<sub>2</sub>] levels in the mid range in most climate regions, but this effect saturates at the highest [CO<sub>2</sub>] levels. Increased [CO<sub>2</sub>] outweighs the damages caused by 20% reduced precipitation in all climate regions in the multi-model median, but not all models agree upon the 25 positive sign. Maize has a comparatively muted response to increased [CO<sub>2</sub>] levels.

## 5 Discussion and Conclusions

The GGCMI Phase II experiment provides a database designed to allow detailed study of crop yields from process-based models under climate change. The use of systematic input parameter variations facilitates not only comparing the sensitivities of process-based crop yield models to changing climate and management inputs but also evaluating the complex interactions 30 between driving factors ([CO<sub>2</sub>], temperature, precipitation, and applied nitrogen). Its global extent also allows identifying geographic shifts in high yield potential locations. We expect that the simulations will yield multiple insights in future studies, and show a selection of preliminary results. We discuss below the implications from the experimental design, but refrain from

analyzing simulation results in detail. Data analyses will be conducted in subsequent analyses making use of the GGCMI Phase II data archive.

First, the GGCMI Phase II simulations allow identifying major areas of uncertainty. Inter-model uncertainty is qualitatively similar across all four inputs tested at the globally aggregate level with some notable exceptions. For example, soybean, a 5 nitrogen-fixing legume, is insensitive to nitrogen addition, while wheat is particularly uncertain in its response to [CO<sub>2</sub>] levels and water availability (Figure SXX). Across geographic regions, projections are most robust in the low latitudes where yield impacts are largest, and most uncertain in the high latitudes where yields may increase under warming. Model differences in projected high-latitude yield changes appear driven more by differences in baseline than in responses to CTWN perturbations. PROMET, for example, involves a stronger response to cold than does LPJmL, with frost below -8 °C irreversibly killing 10 non-winter crops and prolonged periods of below-optimum temperatures also leading to complete crop failure. Over the high-latitudes regions simulated by both models, 52% of grid cells in PROMET report 0 yield in the present climate vs. 11% of cells in the T+4 scenario, leading to a strong yield gain in warmer future climates. In LPJmL outputs, the same high-latitude area is deemed suitable for cultivation even in baseline climate, with crop failure rates of 4% and 5% in present and T+4 cases, so that projected yield changes are modest (Figure 4.)

15 Second, the GGCMI Phase II simulations demonstrate the sensitivity of climate-driven yield impacts to the locations of cultivated land. One counterintuitive result apparent in the simulations is that warmer temperatures drive steeper yield reductions in irrigated than rainfed maize when considered only over currently cultivated land, even though water availability increases crop resiliency to temperature increases at any given location (compare Figure 5 and Figures SX to SX). The effect results from geographic differences in cultivation: irrigated maize is grown in warmer locations where the impacts of warming are more severe 20 (See Figures SX-SX for other crops.) Geographic effects also mean that nitrogen fertilization produces stronger responses in irrigated than non-irrigated wheat and maize, presumably because those rainfed crops are limited by water availability (Figure SX).

Some limitations with the model experiment described here are as follows. The phase II experiment must not be confused with a correct historic setup, as there is no trend in the [CO<sub>2</sub>] and the uniform application of 200 kg N ha<sup>-1</sup> is not representative 25 for many world regions (Elliott et al., 2015). Some issues with process based models are driven in part by the diversity of new cultivars and genetic variants, which outstrips the ability of academic modeling groups to capture them (e.g. Jones et al., 2017). Models also do not simulate many additional factors affecting production, including but not limited to: pests, diseases, and weeds. For these reasons, individual studies must generally re-calibrate models to ensure that short-term predictions reflect current cultivars and management levels. The GGCMI Phase II simulations are designed for evaluating changes in yield across 30 the CTNW-A space but not necessarily absolute yields, since they omit detailed calibrations.

In general, the development of multi-model ensembles involving systematic parameters sweeps has large promise for increasing understanding of potential future crop responses and for improving process-based crop models. The data set is unprecedentedly large, being global in extent, covering 31-simulation years per pixel and up to 756 scenarios for 12 GGCMs. We expect that the GGCMI Phase II data archive will be used to analyze the different GGCMs' sensitivity to changes in the

CTWN-A space, including the interaction between drivers. The authors are working on some analyses but many more are facilitated by the GGCMI Phase II data archive.

*Code and data availability.* The simulation outputs of the mandatory 7 output variables (Table 2) are available on zenodo.org. See Appendix A1 for data DOIs. All other simulation output variables are available upon request to the corresponding author. The scripts for

- 5 generating the spring wheat and winter wheat growing seasons and second fertilizer dates and the quality screening script is available at <https://github.com/RDCEP/ggcmi/blob/phase2/>. All input data are available via globus.org (registration required, free of charge): Minimum cropland mask is available at [https://app.globus.org/file-manager?origin\\_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin\\_path=%2FAGMIP.input%2Fother.inputs%2Fphase2.masks%2F](https://app.globus.org/file-manager?origin_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin_path=%2FAGMIP.input%2Fother.inputs%2Fphase2.masks%2F) choose the file boolean\_cropmask\_ggcmi\_phase2.nc4 Growing period data for wheat is now divided up into winter and spring wheat, available at [https://app.globus.org/file-manager?origin\\_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin\\_path=%2FAGMIP.input%2Fother.inputs%2FAGMIP\\_GROWING\\_SEASON.HARM.version2.0%2F](https://app.globus.org/file-manager?origin_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin_path=%2FAGMIP.input%2Fother.inputs%2FAGMIP_GROWING_SEASON.HARM.version2.0%2F) whereas all other growing season data (maize, rice, soybean) are the same as in Phase I (version 1.25), available at [https://app.globus.org/file-manager?origin\\_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin\\_path=%2FAGMIP.input%2Fother.inputs%2FAGMIP\\_GROWING\\_SEASON.HARM.version1.25%2F](https://app.globus.org/file-manager?origin_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin_path=%2FAGMIP.input%2Fother.inputs%2FAGMIP_GROWING_SEASON.HARM.version1.25%2F)
- 10

## Appendix A

### 15 A1 Data Access

Simulation yield output datasets can be found at the DOIs located in table A1. Data are published in crop- and GGCM-specific packages, in order to break down the overall data amount into manageable packages (<50GB per archive).

*Author contributions.* J.E., C.M, and A.R. designed the research. C.M., J.J., J.B., P.C., M.D., P.F., C.F., L.F., M.H., C.I., I.J., C.J., N.K., M.K., W.L., S.O., M.P., T.P., A.R., X.W., K.W., and F.Z. performed the simulations. J.F., J.J., A.S., M.L., C.M., and E.M. performed the analysis and J.F., E.M., and C.M. prepared the manuscript.

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

*Acknowledgements.* This research was performed as part of the Center for Robust Decision-Making on Climate and Energy Policy (RDCEP) at the University of Chicago, and was supported through a variety of sources. RDCEP is funded by NSF grant #SES-1463644 through the Decision Making Under Uncertainty program. J.F. was supported by the NSF NRT program, grant #DGE-1735359. C.M. was supported by the MACMIT project (01LN1317A) funded through the German Federal Ministry of Education and Research (BMBF). C.F. was supported by the European Research Council Synergy grant #ERC-2013-SyG-610028 Imbalance-P. P.F. and K.W. were supported by the Newton Fund

**Table A1.** DOI's for model yield data outputs. All yield output data can be found at <https://doi.org/10.5281/zenodo/XX>. Where XX is the value found in the table.

Model	Maize	Soybean	Rice	Winter wheat	Spring wheat
<b>APSIM-UGOE</b>	2582531	2582535	2582533	2582537	2582539
<b>CARAIB</b>	2582522	2582508	2582504	2582516	2582499
<b>EPIC-IIASA</b>	2582453	2582461	2582457	2582463	2582465
<b>EPIC-TAMU</b>	2582349	2582367	2582352	2582392	2582418
<b>JULES</b>	2582543	2582547	2582545	–	2582551
<b>GEPIC</b>	2582247	2582258	2582251	2582260	2582263
<b>LPJ-GUESS</b>	2581625	–	–	2581638	2581640
<b>LPJmL</b>	2581356	2581498	2581436	2581565	2581606
<b>ORCHIDEE-crop</b>	2582441	–	2582445	2582449	–
<b>pDSSAT</b>	2582111	2582147	2582127	2582163	2582178
<b>PEPIC</b>	2582341	2582433	2582343	2582439	2582455
<b>PROMET</b>	2582467	2582488	2582479	2582490	2582492

through the Met Office Climate Science for Service Partnership Brazil (CSSP Brazil). K.W. was supported by the IMPREX research project supported by the European Commission under the Horizon 2020 Framework programme, grant #641811. A.S. was supported by the Office of Science of the U.S. Department of Energy as part of the Multi-sector Dynamics Research Program Area. S.O. acknowledges support from the Swedish strong research areas BECC and MERGE together with support from LUCCI (Lund University Centre for studies of Carbon Cycle and Climate Interactions). R.C.I. acknowledges support from the Texas Agrilife Research and 634 Extension, Texas A & M University. This is paper number 35 of the Birmingham Institute of Forest Research. Computing resources were provided by the University of Chicago Research Computing Center (RCC).

## References

- Angulo, C., Rötter, R., Lock, R., Enders, A., Fronzek, S., and Ewert, F.: Implication of crop model calibration strategies for assessing regional impacts of climate change in Europe, *Agric. For. Meteorol.*, 170, 32 – 46, <https://doi.org/10.1016/j.agrformet.2012.11.017>, 2013.
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., J. Boote, K., Thorburn, P., Rötter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Müller, C., Kumar, S. N., Nendel, C., Leary, G. O., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Riponche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stöckle, C., Strattonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., Williams, J. R., and Wolf, J.: Uncertainty in simulating wheat yields under climate change, *Nature Climate Change*, 3, 827–832, <https://doi.org/10.1038/nclimate1916>, 2013.
- Asseng, S., Ewert, F., Martre, P., Rötter, R. P., B. Lobell, D., Cammarano, D., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Biernath, C., Challinor, A., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Muller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stockle, C., Strattonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P. J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., and Zhu, Y.: Rising temperatures reduce global wheat production, *Nature Climate Change*, 5, 143–147, <https://doi.org/10.1038/nclimate2470>, 2015.
- Balkovič, J., van der Velde, M., Skalský, R., Xiong, W., Folberth, C., Khabarov, N., Smirnov, A., Mueller, N. D., and Obersteiner, M.: Global wheat production potentials and management flexibility under the representative concentration pathways, *Global and Planetary Change*, 122, 107 – 121, <https://doi.org/10.1016/j.gloplacha.2014.08.010>, 2014.
- Bodirsky, B. L., Rolinski, S., Biewald, A., Weindl, I., Popp, A., and Lotze-Campen, H.: Global Food Demand Scenarios for the 21st Century, *PLOS ONE*, 10, 1–27, <https://doi.org/10.1371/journal.pone.0139201>, <https://doi.org/10.1371/journal.pone.0139201>, 2015.
- Challinor, A. J., Müller, C., Asseng, S., Deva, C., Nicklin, K. J., Wallach, D., Vanuytrecht, E., Whitfield, S., Ramirez-Villegas, J., and Koehler, A.-K.: Improving the use of crop models for risk assessment and climate change adaptation, *Agricultural systems*, 159, 296–306, 2018.
- Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society*, 137, 553–597, 2011.
- Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoos, M., and François, L.: Responses of European forest ecosystems to 21st century climate: assessing changes in interannual variability and fire intensity, *iForest - Biogeosciences and Forestry*, pp. 82–99, <https://doi.org/10.3832/ifor0572-004>, 2011.
- Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., and Foster, I.: The parallel system for integrating impact models and sectors (pSIMS), *Environmental Modelling and Software*, 62, 509–516, <https://doi.org/10.1016/j.envsoft.2014.04.008>, 2014.
- Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., and Sheffield, J.: The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0), *Geoscientific Model Development*, 8, 261–277, <https://doi.org/10.5194/gmd-2016-207>, 2015.
- FAO/IIASA: Global Agro-ecological Zones and FAO-GAEZ Data Portal(GAEZ v3.0), <http://www.gaez.iiasa.ac.at/>, 2011.

- Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., and Yang, H.: Regionalization of a large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation, and estimation of maize yields, *Agriculture, Ecosystems & Environment*, 151, 21 – 33, <https://doi.org/10.1016/j.agee.2012.01.026>, 2012.
- Folberth, C., Elliott, J., Müller, C., Balkovic, J., Chryssanthacopoulos, J., Izaurrealde, R. C., Jones, C. D., Khabarov, N., Liu, W., Reddy, A.,  
 5 Schmid, E., Skalský, R., Yang, H., Arneth, A., Ciais, P., Deryng, D., Lawrence, P. J., Olin, S., Pugh, T. A. M., Ruane, A. C., and Wang, X.: Uncertainties in global crop model frameworks: effects of cultivar distribution, crop management and soil handling on crop yield estimates, *Biogeosciences Discussions*, 2016, 1–30, <https://doi.org/10.5194/bg-2016-527>, <https://www.biogeosciences-discuss.net/bg-2016-527/>, 2016.
- Food and Agriculture Organization of the United Nations: FAOSTAT Database, <http://www.fao.org/faostat/en/home>, 2018.
- 10 Frieler, K., Lange, S., Piontek, F., Reyer, C. P., Schewe, J., Warszawski, L., Zhao, F., Chini, L., Denvil, S., Emanuel, K., et al.: Assessing the impacts of 1.5 C global warming—simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b), *Geoscientific Model Development*, 2017.
- Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F., Dumont, B., Ewert, F., Ferrise, R., François,  
 15 L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., and Rötter, R. P.: Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change, *Agricultural Systems*, 159, 209–224, <https://doi.org/10.1016/j.aggsy.2017.08.004>, 2018.
- 20 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., and Moyer, E. J.: Evaluating the utility of dynamical downscaling in agricultural impacts projections, *Proceedings of the National Academy of Sciences*, 111, 8776–8781, <https://doi.org/10.1073/pnas.1314787111>, 2014.
- Glotter, M., Moyer, E., Ruane, A., and Elliott, J.: Evaluating the Sensitivity of Agricultural Model Performance to Different Climate Inputs, *Journal of Applied Meteorology and Climatology*, 55, 151113145618 001, <https://doi.org/10.1175/JAMC-D-15-0120.1>, 2015.
- Hank, T., Bach, H., and Mauser, W.: 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 Sensing*, 7, 3934–3965, <https://doi.org/10.3390/rs70403934>, 2015.
- 25 Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, *International Journal of Climatology*, 34, 623–642, <https://doi.org/10.1002/joc.3711>, <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3711>, 2014.
- 30 Hasegawa, T., Fujimori, S., Havlík, P., Valin, H., Bodirsky, B. L., Doelman, J. C., Fellmann, T., Kyle, P., Koopman, J. F., Lotze-Campen, H., et al.: Risk of increased food insecurity under stringent global climate change mitigation policy, *Nature Climate Change*, 8, 699, 2018.
- Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., and Keating, B. A.: AP-SIM – Evolution towards a new generation of agricultural systems simulation, *Environmental Modelling and Software*, 62, 327 – 350, <https://doi.org/doi.org/10.1016/j.envsoft.2014.07.009>, 2014.

- Humpenöder, F., Popp, A., Bodirsky, B. L., Weindl, I., Biewald, A., Lotze-Campen, H., Dietrich, J. P., Klein, D., Kreidenweis, U., Müller, C., et al.: Large-scale bioenergy production: how to resolve sustainability trade-offs?, *Environmental Research Letters*, 13, 024 011, 2018.
- Iizumi, T., Nishimori, M., and Yokozawa, M.: Diagnostics of Climate Model Biases in Summer Temperature and Warm-Season Insolation for the Simulation of Regional Paddy Rice Yield in Japan, *Journal of Applied Meteorology and Climatology*, 49, 574–591, 5 <https://doi.org/10.1175/2009JAMC2225.1>, 2010.
- Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanenkov, V., Oetli, P., Newby, T., Ishigooka, Y., and Furuya, J.: Historical changes in global yields: major cereal and legume crops from 1982 to 2006, *Global Ecology and Biogeography*, 23, 346–357, <https://doi.org/10.1111/geb.12120>, 2014.
- Izaurrealde, R., Williams, J., McGill, W., Rosenberg, N., and Quiroga Jakas, M.: Simulating soil C dynamics with EPIC: Model description 10 and testing against long-term data, *Ecological Modelling*, 192, 362–384, <https://doi.org/10.1016/j.ecolmodel.2005.07.010>, 2006.
- J. Boote, K., Jones, J., White, J., Asseng, S., and Lizaso, J.: Putting Mechanisms into Crop Production Models., *Plant, cell & environment*, 36, <https://doi.org/10.1111/pce.12119>, 2013.
- Jägermeyr, J. and Frieler, K.: Spatial variations in crop growing seasons pivotal to reproduce global fluctuations in maize and wheat yields, *Science Advances*, 4, 4517, <https://doi.org/10.1126/sciadv.aat4517>, 2018.
- Jagtap, S. S. and Jones, J. W.: Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production, *Agriculture, Ecosystems & Environment*, 93, 73 – 85, [https://doi.org/10.1016/S0167-8809\(01\)00358-9](https://doi.org/10.1016/S0167-8809(01)00358-9), 2002.
- Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., and Ritchie, J.: The DSSAT cropping system model, *European Journal of Agronomy*, 18, 235 – 265, [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7), 2003.
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, 20 S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., and Wheeler, T. R.: Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science, *Agricultural Systems*, 155, 269 – 288, <https://doi.org/10.1016/j.agsy.2016.09.021>, 2017.
- Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., 25 Freebairn, D., and Smith, C.: An overview of APSIM, a model designed for farming systems simulation, *European Journal of Agronomy*, 18, 267 – 288, [https://doi.org/10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9), 2003.
- Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., and Smith, B.: Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa, *Earth System Dynamics*, 4, 385–407, <https://doi.org/10.5194/esd-4-385-2013>, 2013.
- Liu, J., Williams, J. R., Zehnder, A. J., and Yang, H.: GEPIC - modelling wheat yield and crop water productivity with high resolution on a 30 global scale, *Agricultural Systems*, 94, 478 – 493, <https://doi.org/10.1016/j.agsy.2006.11.019>, 2007.
- Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., and Schulin, R.: Global investigation of impacts of PET methods on simulating crop-water relations for maize, *Agricultural and Forest Meteorology*, 221, 164 – 175, <https://doi.org/10.1016/j.agrformet.2016.02.017>, 2016a.
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., and Schulin, R.: Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations, *Science of The Total Environment*, 572, 526 – 537, 35 <https://doi.org/10.1016/j.scitotenv.2016.08.093>, 2016b.
- Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, D., Challinor, A., Conijn, J., and Wolf, J.: Statistical Analysis of Large

Simulated Yield Datasets for Studying Climate Effects, p. 1100, World Scientific Publishing, <https://doi.org/10.13140/RG.2.1.5173.8328>, 2015.

- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., and Calzadilla, A.: Global biomass production potentials exceed expected future demand without the need for cropland expansion, *Nature Communications*, 6, <https://doi.org/10.1038/ncomms9946>, 2015.
- 5 McDermid, S. P., Ruane, A. C., Rosenzweig, C., Hudson, N. I., Morales, M. D., Agalawatte, P., Ahmad, S., Ahuja, L., Amien, I., Anapalli, S. S., et al.: The AgMIP coordinated climate-crop modeling project (C3MP): methods and protocols, in: *HANDBOOK OF CLIMATE CHANGE AND AGROECOSYSTEMS: The Agricultural Model Intercomparison and Improvement Project Integrated Crop and Economic Assessments*, Part 1, pp. 191–220, World Scientific, 2015.
- Minoli, S., Egli, D. B., Rolinski, S., and Müller, C.: Modelling cropping periods of grain crops at the global scale, *Global and Planetary Change*, 174, 35–46, 2019.
- 10 Müller, C., Elliott, J., Chryssanthacopoulos, J., Deryng, D., Folberth, C., Pugh, T. A., and Schmid, E.: Implications of climate mitigation for future agricultural production, *Environmental Research Letters*, 10, 125 004, 2015.
- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurrealde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., and Yang, H.: Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications, *Geoscientific Model Development*, 10, 1403–1422, <https://doi.org/10.5194/gmd-10-1403-2017>, 2017.
- Olin, S., Schurges, G., Lindeskog, M., Wårlind, D., Smith, B., Bodin, P., Holmér, J., and Arneth, A.: 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, 2489–2515, <https://doi.org/10.5194/bg-12-2489-2015>, 2015.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., and Wheeler, T.: JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator, *Geoscientific Model Development*, 8, 1139–1155, <https://doi.org/10.5194/gmd-8-1139-2015>, 2015.
- Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., and Rötter, R.: Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces, *Climate Research*, 65, 87–105, <https://doi.org/10.3354/cr01322>, 2015.
- Porter et al. (IPCC): Food security and food production systems. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.*, in: *IPCC Fifth Assessment Report*, edited by et al., C. F., pp. 485–533, Cambridge University Press, Cambridge, UK, 2014.
- Portmann, F., Siebert, S., and Doell, P.: MIRCA2000 - Global Monthly Irrigated and Rainfed Crop Areas around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling, *Global Biogeochemical Cycles*, 24, GB1011, <https://doi.org/10.1029/2008GB003435>, 2010.
- Porwollik, V., Müller, C., Elliott, J., Chryssanthacopoulos, J., Iizumi, T., Ray, D. K., Ruane, A. C., Arneth, A., Balkovič, J., Ciais, P., Deryng, D., Folberth, C., Izaurrealde, R. C., Jones, C. D., Khabarov, N., Lawrence, P. J., Liu, W., Pugh, T. A., Reddy, A., Sakurai, G., Schmid,

E., Wang, X., de Wit, A., and Wu, X.: Spatial and temporal uncertainty of crop yield aggregations, European Journal of Agronomy, 88, 10–21, <https://doi.org/10.1016/j.eja.2016.08.006>, 2017.

Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E., and Arneth, A.: Climate analogues suggest limited potential for intensification of production on current croplands under climate change, Nature Communications, 7, 12608, 5 <https://doi.org/10.1038/ncomms12608>, 2016.

Ray, D., Ramankutty, N., Mueller, N., West, P., and A Foley, J.: Recent patterns of crop yield growth and stagnation, Nature communications, 3, 1293, <https://doi.org/10.1038/ncomms2296>, 2012.

Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., and Schlenker, W.: Comparing and combining process-based crop models and statistical models with some implications for climate change, Environmental Research Letters, 12, <https://doi.org/10.1088/1748-9326/aa7f33>., 2017.

10 Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., and Winter, J.: The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies, Agricultural and Forest Meteorology, 170, 166 – 182, <https://doi.org/10.1016/j.agrformet.2012.09.011>, 2013.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., and Jones, J. W.: Assessing agricultural risks of climate change in 15 the 21st century in a global gridded crop model intercomparison, Proceedings of the National Academy of Sciences, 111, 3268–3273, <https://doi.org/10.1073/pnas.1222463110>, 2014.

Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta, A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G., Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M., Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M., Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., and Wiebe, K.: Coordinating AgMIP data and models across 20 global and regional scales for 1.5°C and 2.0°C assessments, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 376, <https://doi.org/10.1098/rsta.2016.0455>, 2018.

Rötter, R. P., Carter, T. R., Olesen, J. E., and Porter, J. R.: Crop–climate models need an overhaul, Nature climate change, 1, 175, 2011.

Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W., Romero, C. C., and Cecil, L. D.: Carbon-temperature-water 25 change analysis for peanut production under climate change: A prototype for the AgMIP Coordinated Climate-Crop Modeling Project (C3MP), Glob. Change Biology, 20, 394–407, <https://doi.org/10.1111/gcb.12412>, 2014.

Ruane, A. C., Goldberg, R., and Chryssanthacopoulos, J.: Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation, Agric. Forest Meteorol., 200, 233–248, <https://doi.org/10.1016/j.agrformet.2014.09.016>, 2015.

Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M., Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., and Rosenzweig, C.: Biophysical and economic implications for agriculture of +1.5° and +2.0°C global warming using AgMIP Coordinated Global and Regional Assessments, Climate Research, 76, 17–39, 30 <https://doi.org/10.3354/cr01520>, 2018.

Rubel, F. and Kottek, M.: Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification, Meteorologische Zeitschrift, 19, 135–141, <https://doi.org/10.1127/0941-2948/2010/0430>, 2010.

Ruiz-Ramos, M., Ferrise, R., Rodríguez, A., Lorite, I., Bindl, M., Carter, T., Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, S., 35 Cammarano, D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P., Hoffmann, H., Höhn, J., Jurecka, F., Kersebaum, K., Krzyszczak, J., Lana, M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J., Ruget, F., Semenov, M., Steinmetz, Z., Stratonovitch, P., Supit, I., Tao, F., Trnka, M., de Wit, A., and Rötter, R.: Adaptation response surfaces for managing wheat under perturbed climate and CO2 in a Mediterranean environment, Agricultural Systems, 159, 260 – 274, <https://doi.org/10.1016/j.agsy.2017.01.009>, 2018.

- Sacks, W. J., Deryng, D., Foley, J. A., and Ramankutty, N.: Crop planting dates: an analysis of global patterns, *Global Ecology and Biogeography*, 19, 607–620, <https://doi.org/10.1029/2009GB003765>, 2010.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., and Frieler, K.: Consistent negative response of US crops to high temperatures in observations and crop models, *Nature Communications*, 8, 13 931, <https://doi.org/10.1038/ncomms13931>, 2017.
- Schewe, J., Gosling, S. N., Reyer, C., Zhao, F., Ciais, P., Elliott, J., Francois, L., Huber, V., Lotze, H. K., Seneviratne, S. I., van Vliet, M. T. H., Vautard, R., Wada, Y., Breuer, L., Büchner, M., Carozza, D. A., Chang, J., Coll, M., Deryng, D., de Wit, A., Eddy, T. D., Folberth, C., Frieler, K., Friend, A. D., Gerten, D., Gudmundsson, L., Hanasaki, N., Ito, A., Khabarov, N., Kim, H., Lawrence, P., Morfopoulos, C., Müller, C., Müller Schmied, H., Orth, R., Ostberg, S., Pokhrel, Y., Pugh, T. A. M., Sakurai, G., Satoh, Y., Schmid, E., Stacke, T., Steenbeek, J., Steinkamp, J., Tang, Q., Tian, H., Tittensor, D. P., Volkholz, J., Wang, X., and Warszawski, L.: State-of-the-art global models underestimate impacts from climate extremes, *Nature Communications*, 10, 1005–, <https://doi.org/10.1038/s41467-019-08745-6>, 2019.
- Snyder, A., Calvin, K. V., Phillips, M., and Ruane, A. C.: A crop yield change emulator for use in GCAM and similar models: Persephone v1.0, Accepted for publication in *Geoscientific Model Development*, pp. 1–42, <https://doi.org/10.5194/gmd-2018-195>, in open review, 2018.
- Stevanović, M., Popp, A., Lotze-Campen, H., Dietrich, J. P., Müller, C., Bonsch, M., Schmitz, C., Bodirsky, B. L., Humpenöder, F., and Weindl, I.: The impact of high-end climate change on agricultural welfare, *Science Advances*, 2, <https://doi.org/10.1126/sciadv.1501452>, <https://advances.sciencemag.org/content/2/8/e1501452>, 2016.
- Vadez, V., Berger, J. D., Warkentin, T., Asseng, S., Ratnakumar, P., Rao, K. P. C., Gaur, P. M., Munier-Jolain, N., Larmure, A., Voisin, A.-S., et al.: Adaptation of grain legumes to climate change: a review, *Agronomy for Sustainable Development*, 32, 31–44, 2012.
- von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., and Zaehle, S.: Implementing the Nitrogen cycle into the dynamic global vegetation, hydrology and crop growth model LPJmL (version 5.0), *Geoscientific Model Development*, 11, 2789–2812, <https://doi.org/10.5194/gmd-11-2789-2018>, 2018.
- Waha, K., Van Bussel, L., Müller, C., and Bondeau, A.: Climate-driven simulation of global crop sowing dates, *Global Ecology and Biogeography*, 21, 247–259, 2012.
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resources Research*, 50, 7505–7514, 2014.
- Wheeler, T. and Von Braun, J.: Climate change impacts on global food security, *Science*, 341, 508–513, 2013.
- Wiebe, K., Lotze-Campen, H., Sands, R., Tabeau, A., van der Mensbrugghe, D., Biewald, A., Bodirsky, B., Islam, S., Kavallari, A., Mason-D'Croz, D., Müller, C., Popp, A., Robertson, R., Robinson, S., van Meijl, H., and Willenbockel, D.: Climate change impacts on agriculture in 2050 under a range of plausible socioeconomic and emissions scenarios, *Environmental Research Letters*, 10, 085 010, [https://doi.org/10.1088%2F1748-9326%2F10%2F8%2F085010](https://doi.org/10.1088/1748-9326/10/8/085010), 2015.
- Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., Arkebauer, T., and Scoby, D.: Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska, *Geoscientific Model Development*, 10, 1291–1320, <https://doi.org/10.5194/gmd-10-1291-2017>, 2017.
- Williams, K. E. and Falloon, P. D.: Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts, *Geoscientific Model Development*, 8, 3987–3997, <https://doi.org/10.5194/gmd-8-3987-2015>, 2015.

- Wolf, J. and Oijen, M.: Modelling the dependence of European potato yields on changes in climate and CO<sub>2</sub>, Agricultural and Forest Meteorology, 112, 217 – 231, [https://doi.org/10.1016/S0168-1923\(02\)00061-8](https://doi.org/10.1016/S0168-1923(02)00061-8), 2002.
- Wu, X., Vuichard, N., Ciais, P., Viovy, N., de Noblet-Ducoudré, N., Wang, X., Magliulo, V., Wattenbach, M., Vitale, L., Di Tommasi, P.,  
Moors, E. J., Jans, W., Elbers, J., Ceschia, E., Tallec, T., Bernhofer, C., Grünwald, T., Moureaux, C., Manise, T., Ligne, A., Cellier, P.,  
5 Loubet, B., Larmanou, E., and Ripoche, D.: ORCHIDEE-CROP (v0), a new process-based agro-land surface model: model description  
and evaluation over Europe, Geoscientific Model Development, 9, 857–873, <https://doi.org/10.5194/gmd-9-857-2016>, 2016.