Author resonse to referee comments on "The GGCMI Phase 2 experiment: global gridded crop model simulations under uniform changes in CO_2, temperature, water, and nitrogen levels (protocol version 1.0)"

James Franke et al.[1,2]

[1]Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA
[2]Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

Correspondence: James Franke (jfranke@uchicago.edu)

1 Cover letter to the Editor

March 13, 2020

Dear Editor,

Attached is the author response to referee comments and the modified manuscript. The largest change involves Subsection 2.2, which has been substantially modified to address reviewer comments. The section now more clearly outlines the rationale for the uniform offsets used in the experiment design. This was the main concern raised by both reviewers. Other significant modifications to address reviewer comments include: a table detailing model differences has been added to the supplement (Table S1), and three figures have been added to the supplement that show model responses to other dimensions as requested by a reviewer. Because the original manuscript led both reviewers to miss key elements of the underlying motivation for the experiment, we have also made numerous minor adjustments throughout the paper to clarify the writing. Please note that we changed "Phase II" throughout the paper to "Phase 2" to match previous publication convention. This includes in the title of the paper. No figures have been modified. Attached are:

- The response to the referee comments, with explanation of line-by-line changes to the manuscript. *Original comments in gray* and author responses in black.
- The pdf with all the text modifications highlighted (using difflatex). Red text has been removed and blue text has been added.

Thank you,
James Franke (and coauthors)
University of Chicago

2 Anonymous Referee 1

COMMENT: The manuscript by Franke et al., details the experimental design for the Phase II GGCMI crop model comparison. The goal is to provide a set of simulations to synchronize a variety of crop models and compare the responses from perturbations of temperature, precipitation, CO2, and nitrogen fertilizer. The result is a dataset of thousands of simulations that can be used to emulate statistical crop model response under varying inputs of climate change. The authors provide some analysis of the dataset, providing examples of non-linear behavior under multiple variable perturbations be-tween temperature, precipitation, and CO2. Furthermore, the authors provide access to other users for additional analysis. The manuscript is well written, the message is clearly defined, with a logical flow throughout, and void of technical errors. The authors did a good job detailing some of the more complex features of their study.

RESPONSE: Thank you for the overall assessment.

COMMENT: My main concern with this manuscript is I find the approach toward the perturbation experiments somewhat unrealistic. I understand the difficulty in generating simulations across different models in a way that is uniform, and I find the large number of simulations included in the dataset impressive but having such a large set of parameters for the simulations makes interpreting the output difficult and negates the heterogeneous(in space and time) behavior of climate. Wouldn't it be easier to use CMIP output to drive simulations which could reduce the number of model runs? Perhaps the authors could provide more discussion on this choice. Also, since these are offline runs, they don't include feedbacks between the atmosphere and land (e.g. irrigation feedbacks to temperature), which are important.

RESPONSE: Yes, the approach of using uniformly perturbed climate inputs does not reflect realistic climate scenarios. If the goal of GGCMI Phase 2 were to use these simulations for climate change impact assessments, this experimental design would be the wrong choice. However, the goals of GGCMI-2 are to (i) scrutinize model response in response to individual and combined drivers and to (ii) develop crop model emulators on these experiments. Both of those goals require sampling across the space of potential perturbations that allows untangling the contributions of individual factors that are highly correlated in realistic future scenarios (e.g. CO2 and temperature). That is, meeting our goals requires a suite of unrealistic inputs.

We do believe that GGCMI-2 can also serve the needs of impacts assessments through the development of emulators. That is, the responses to the CTWN-A factors diagnosed from GGCMI-2 can be used to build up emulations of what the crop models would produce under a realistic climate scenario, including all the heterogeneous aspects of true climate change. (In this exercise, a crop model emulator for each individual grid cell is driven with the timeseries of projected climate changes for that particular location.)

This use is not demonstrated in the manuscript here, which is the "experiment description" paper; instead it is shown in a companion "model description" paper that describes the CTWN-A emulators, and is now available as a GMD discussion paper (https://www.geosci-model-dev-discuss.net/gmd-2019-365/). GMD had requested that we split our discussion of GGCMI-2 into these two components, to clearly distinguish the experimental description from the emulator development. In the companion paper we show that realistic CMIP-based simulations can be reproduced extremely well by emulators built from the CTWN-A experiment. This is true despite the fact that the uniform offset experiments omit some aspects of climate change that could be important to crops - the distribution of weather conditions within growing season, e.g. stronger warming in spring vs. summer. Such effects do not appear large enough to compromise the GGCMI-2 emulators.

We have revised the text of the manuscript under review here to better explain the rationale of GGCMI-2 and to point to the companion emulator paper as a justification of its utility for impacts assessment. This point is extremely important and we thank the reviewer for pointing out that our explanation was insufficiently clear in the submitted manuscript.

It is true that these experiments do not include feedbacks between irrigation and temperature. This is non-ideal, but in practice, the vast majority of crop yield projections under climate change also omit this feedback, and instead simply feed a climate projection to a process-based or statistical crop model. We have added some discussion of this point in the manuscript and suggested the need for future studies.

CHANGES:

- Extensive justification for the experiment set-up has been added to pages 6 and 7.
- The companion paper which details emulation for impact assessment is now cited on page 7, line 19.
- Feedbacks between atmospheres and temperatures are now addressed in new sentences on page 24, lines 1 through 3.

COMMENT: I did not find the A1 simulations discussed anywhere. They seem to be included in the methods section but are not included in the analysis. Perhaps they should be omitted. Similarly, the nitrogen simulations are also missing from the analysis (except for the correlation with observations).

RESPONSE: The A dimension (adaptation in growing season length) is an integral part of the protocol and should be described fully in the experiment description paper. We have now given it more attention in the overall paper.

Note that the adaptation dimension will always be treated somewhat separately in discussion as it is not directly comparable to the other four dimensions (CTWN): it does not address inputs but the parameterization of crop varieties.

We have tried to limit the amount of analysis of results that are shown as experiment description papers are supposed to focus on experimental design, with a few results only as illustrations. However, we have added some additional material on N and A so that these dimensions do not seem less important.

CHANGES:

- A figure has been added to the supplement (S14) that shows the yield response to increased temperature in the A1 scenarios.
- Some discussion of the difference between A1 and A0 response has been added to page 18 line 21 - 25. Clarification between which scenario is being referenced in each case has be added throughout.
- New discussion of model calibration procedures for A1 growing seasons has been added to the supplement page 5.
- A figure has been added to the supplement (S15) that shows the yield response to increased temperature across the nitrogen dimension.

General Comments:

COMMENT: P. 7, Section 2.3: The 12 models included in the study are very different types of models. I know this was discussed in the original paper describing protocol I, but it should also be noted here. How did the model differences inform the experimental design (or limit the scope of the study)?

RESPONSE: Yes, the inclusion of different model types in a model intercomparison both complicates and enriches the analysis. One goal of the GGCMI Phase 2 experiment is to analyze model differences in order to better understand skills and deficiencies and to improve models. We have added this point more clearly in the text. Based on reviewer comments, we are also now adding a section describing key differences among models, and including a table in the supplemental material that describes model differences in inputs, structure and setup.

CHANGES:

- A new paragraph describing model differences has been added to page 12, lines 25-33.
- A new table has been added to the supplement (Table S1) detailing model differences.

COMMENT: P. 9, L. 10: If some models don't output the anthesis date, why is it considered mandatory?

RESPONSE: The anthesis date is an important phenological indicator and was considered a standard output also in the previous stages of GGCMI. However, as some models do not explicitly compute anthesis dates, these cannot deliver these outputs. The "mandatory" label means that models that do compute anthesis should report it. We have modified the text and table caption to make this clear.

CHANGES:

- Table caption modified as noted on page 11.
- The "mandatory" designation has been clarified on page 10, line 26.

COMMENT: P. 15, L. 6: Is the negative impact on yield from increasing temperature due to shorter growing seasons or from actual heat damage to the crop?

RESPONSE: Typically, the effect is a combination of the two mechanisms. The use of both A0 and A1 setups was designed to answer exactly this question. We have added text to emphasize this point. Note however that the experiment description paper here is not intended to conduct all these analyses, but rather to describe the protocol and outputs of the experiments that will allow questions to be answered. We are glad that the experiment provokes such useful responses! GGCMI team members are currently preparing a paper describing in detail the effects of adaptation in these experiments.

CHANGES:

- Text added as noted on page 17, line 15.

COMMENT: P. 15, L 11-13: The change in yields at different latitudes is unrealistic because of the design of the experiment. Simply increasing temperature uniformly and not accounting for the seasonal differences in temperature change (i.e., stronger *winter increase in temperature and weak or no summer increase) results in an unrealistic "warming" during the growing season that might not exist. This is also the probable cause of the increase in yield from the least realistic simulations (Pl. 15, L 28-29).*

RESPONSE: As discussed above, the uniform perturbations are not intended to reproduce a realistic scenario and should not be used as such. The reviewer's comment is useful in telling us that we need to make this point more clear in the paper. We have added language to emphasize that the GGCMI-2 output for a given uniform temperature shift should not be taken as a proxy for an actual projection under a realistic climate scenario that produces the equivalent global mean temperature change.

We have tried to clarify two important points brought up by this comment. First, the climate offsets in the GGCMI-2 experiments refer to offsets during the growing season, not to annual means. The strong increases in yield in high-latitude regions in some simulations are therefore the appropriate response for each model given the applied level of warming during the growing season. Models of course disagree on the extent or even the sign of yield changes, especially in high-latitude regions, and their responses may be unrealistic.

Second, a scenario with a uniform offset (across space and time) will not match a scenario with the same mean change but with the spatial patterns of climate change expected under future scenarios. The effects of these spatial patterns are shown explicitly in the companion GMD "model description" paper (see link above). We now refer to that paper explicitly.

CHANGES:

- New paragraph added as noted on page 7, lines 7-20 as noted.

3 Anonymous Referee 2

COMMENT: The manuscript by Franke et al. documented a new AgMIP GGCMI effort on simulating the crop responses to globally uniform environmental perturbations, includingCO2, temperature, precipitation, nitrogen, and adaptation (CTWN-A). The simulation protocols are described in detail and key model outputs are made publically available. The authors made the first cut on data analysis to show the key characteristics of the simulated dataset. Overall, this manuscript is well organized and written. It also fulfills the scope of GMD and should be of great interests to the broader crop modeling and climate change adaptation community.

RESPONSE: Thank you for the overall positive assessment.

COMMENT: I have the following comments for the authors to consider:

COMMENT: Firstly, I see the nitrogen application rates designed in Table 1 are largely not realistic, especially considering how nitrogen application rates differ for different crops. I am not sure if I misunderstood anything there, but please help to clarify this point.

RESPONSE: The idea of the uniform perturbation and input levels in the CTWN-A experiment is not to be fully realistic but to allow for in-depth analyses by providing a structured analysis framework. Fertilizer application rates differ substantially across crops (e.g. maize vs. soybean) but also across the globe where access to fertilizers is often limited. We designed the ranges of CTWN so that low and high-end values are included and model behavior can be understood across these dimensions.

By using a range of nitrogen input levels (as well as inputs of climate variables), we are able to construct "emulators" of the crop model responses to these factors for arbitrary input levels. That is, the GGCMI-2 experiments allow constructing a response surface that would allow reproducing the output crop models would have produced if run with more realistic (or indeed, any) nitrogen inputs. This use is explained in detail in a companion GCM paper now available online at (https://www.geosci-model-dev-discuss.net/gmd-2019-365/). We have added more discussion in this first "experiment description" paper to make this clear, and now point to the companion paper.

CHANGES:

- Text added on page 6, line 11 to clarify the intent of the N application levels.

COMMENT: Secondly, I found some critical information is missing in the current manuscript. For example, the differences among different models (especially those with the same base),the irrigation triggering rules in different models, key model inputs (such as cultivar in-formation), model tuning method and model spin-up design. Please see later detailed comments.

RESPONSE: We agree with this critique and have expanded the discussion of structural differences among models. We have also included a table in the Supplemental Materials showing key model features and structural differences. This addition will make the manuscript significantly more useful for readers.

CHANGES:

- A new paragraph describing model differences has been added to page 12, lines 25-33.

- A new table has been added to the supplement (Table S1) detailing model differences.

COMMENT: Thirdly, there are 7 mandatory variables in Table 2. However, the authors only discussed yield, which I agree is the most important one for crop models. If the authors can have some discussions on other variables, it would be very interesting, even if the figures are dumped into supplementary materials.

RESPONSE: As this is the experiment description paper and the experiment is very comprehensive (experiments, different model types, different output variables), we tried to find a balance between producing a readable overview paper and one with exhaustive detail. However, we may have erred on the side of over-focusing on yield. We have therefore now added some examples of other outputs in the supplementary material and more discussion in the main text.

CHANGES:

- A figure illustrating the irrigation water response to warming has been added to the supplement (Figure S16).

COMMENT: More detailed comments are as following: COMMENT: P2, L30-L31: the transition to "Global crop model experiments are needed for systematic climate change assessments" is a little wired to me. Are you talking about the same point with last sentence or not?

RESPONSE: Yes, this sentence is a bit too condensed. We have updated the language accordingly.

CHANGES:

- Text modified on page 2, line 27-33 to clarify meaning.

COMMENT: P3, L22: Folberth et al. (2016); Porwollik et al. (2017)-> Folberth et al. (2016) and Porwollik et al. (2017)?

RESPONSE: Thank you for catching this! These references are now corrected; we have updated Folberth et al. 2016 to Folberth et al. 2019.

CHANGES:

- Correction made on page 3, line 28.

COMMENT: P3, L25: (C3MP Ruane et al., 2014; McDermid et al., 2015)-> (C3MP) (Ruane et al.,2014; McDermid et al., 2015)

RESPONSE: We have changed to "(C3MP, see Ruane et al., 2014; McDermid et al., 2015)"

CHANGES:

- Correction made on page 3, line 33.

COMMENT: P4, L26: an additional 84 for irrigated ($W\infty$)->an additional 84 for irrigated area ($W\infty$)?Are those 84 cases for irrigated area only with the assumption that the irrigated area will not change or also for rainfed area too (to get rid of water stress in rainfed regions)? Please clarify this point. It would be really interesting to have a no-water-limitation case for rainfed area. Moreover, how does each model trigger irrigation? Does the irrigated amount differ a lot among models?

RESPONSE: No, following the general GGCMI experiment design, fully rainfed and fully irrigated systems are simulated in all grid cells, independent of their actual distribution. This protocol allows for better analyses (e.g. simulations with and without water stress can be compared) and also for understanding the potential of optional cropland expansion. In the GGCMI Phase 2 setup, irrigated systems are also simulated during the rainfed growing seasons so that the simulation results are directly comparable and only differ with respect to water supply. We have modified the text to make this choice more clear.

CHANGES:

- Text added on page 6, lines 1-6 line 3 to clarify the irrigation protocol.

- Table added to supplement detailing individual model protocol for irrigation triggering (Table S1).

COMMENT: Table 1: There are three levels of applied nitrogen (10, 60, 200 kg/ha). Are those three levels uniformly applied for all the five crops? For soybean, we don't need that much nitrogen (200 kg/ha), right? For corn, is 10 kg/ha a too strong nitrogen limitation, especially for a few regions such as US?

RESPONSE: Yes, as discussed above, N levels in the experimental protocol are uniformly applied across all locations and all crops, and are not intended to be realistic. The experiment design is intended to span the full range of plausible input values, though the maximum of 200 kgN/ha may actually be a bit low for some crops and regions. Using uniform offsets and input levels allows structured analysis of the effects of each factor. As per answers above, we have now made this rationale more clear in the text.

CHANGES:

- Text added on page 6, lines 11 to clarify the intent of the N application levels.

COMMENT: P7, L5: it would be good to document the main differences related with crop growth among those sharing-a-common-base models, i.e. EPIC group (EPIC-IIASA, EPIC-TAMU, GEPIC, pEPIC), and LPJ group (LPJml, LPJ-GUESS).

RESPONSE: Yes, as discussed above, these differences are included in the table with details on model inputs, structure and setup that is now included in Supplementary Material. We feel this addition greatly strengthens the utility of the paper and thank the reviewer for the suggestion.

CHANGES:

- A new paragraph describing model differences has been added to page 12, lines 25-33 including some discussion about models that share a common genealogy.

- A new table has been added to the supplement (Table S1) detailing model differences.

COMMENT: P7, L24-L25: will the change of phenological parameters have a huge impact on yield for different models?

RESPONSE: Yes, model performance can be very sensitive to the parametrization of growing seasons. That is why the experiment protocol prescribes harmonized growing seasons so that it is easier to analyze model responses. We have amplified discussion of this point in the text.

CHANGES:

– Sentence added to page 8, line 28.

COMMENT: P7, L28: what's the "technical reasons" for CARAIB model? A note should be put on this.

RESPONSE: We agree that "technical reasons" does not adequately describe the issue at hand. In fact, the CARAIB team simply missed harmonizing this aspect. We think it is still of value to include their output in the archive, and any applications can exclude CARAIB results if required for their purposes. We have adjusted the sentence in question.

CHANGES:

– Text modified in the paragraph starting on page 8, line 24 to clarify the process here.

COMMENT: P7L35-P8L1: how did modelers adjust those parameters? Was it manual tuning or automatic tuning? And should this tuning be conducted for every year and each location? Ideally, there should be a section in the appendix for parameter tuning to include related details (parameter space, and tuning method)

RESPONSE: We now describe this procedure more clearly in the main text. The groups used manual parameter tuning to harmonize the growing seasons. First, parameters are adjusted for each crop in each location under the unperturbed AgMERRA baseline climate timeseries so that growing seasons in this 31-year period (1980-2010) reproduce specified observed average growing seasons for this period. For A0 simulations, the parameters are then left constant for all experiments, so that growing seasons alter under warming.

Note that because each crop model sets the growing season differently, the parameters modified will differ across models. Describing the exact procedure of the different modeling groups would require extensive discussion of the structure of each model. While we agree an appendix describing this would be useful to some readers, we feel it is out of scope for this paper. We hope that this need is satisfied instead by our links to the description papers for each individual model, which should cover their process of determining growing seasons.

CHANGES:

– Text modified in the paragraph on page 9, lines 8-22 to better describe the growing season calibration.

– New list of A1 case calibration measures added to the supplement on page 5.

COMMENT: P9, L10: please move "(Note that several models do not output the anthesis date.)"after "the dates of planting, anthesis, and maturity", i.e. the dates of planting, anthesis, and maturity (Note that several models do not output the anthesis date).

RESPONSE: Reviewer 1 also had problems with this sentence and it has been revised accordingly. It no longer includes parentheses.

CHANGES:

– Correction made on page 10, line 26.

COMMENT: P9, L8: 30-year or 31-year (1980-2010)? What's the model spin-up protocol?

RESPONSE: The spin up is very different across models. This is now documented in the new table on model inputs, structure and setup.

CHANGES:

- New table added to supplement (Table S1).

COMMENT: P11, L20: no italic text in Table 3!

RESPONSE: Thank you for catching this. These simulation sets are shown in bold in Table 3 (column "Sims per crop"), and the sentence is now corrected.

CHANGES:

- Correction made on page 14, line 13.

COMMENT: P11, L28: did you missed 510 ppm there?

RESPONSE: Yes, thanks for catching this; it is now corrected.

CHANGES:

- Correction made on page 14, line 21.

COMMENT: P13, L26: For example, global correlation coefficients for maize in Phase I and Phase II are 0.89 and 0.74, respectively; for wheat 0.67 and 0.64, and for soybeans 0.64 and 0.59. (Compare to Müller et al. (2017) Figures 1–4 and 6.)-> For example, global correlation coefficients in Phase I and Phase II are 0.89 and 0.74 for maize, 0.67 and0.64 for wheat, and 0.64 and 0.59 for soybeans, respectively (Phase I values are from Figures 1-4 and 6 in Müller et al. (2017))

RESPONSE: Corrected as suggested.

CHANGES:

- Correction made on page 15, line 28.

COMMENT: P13, L27: Figure 2 should be Figure 2(c)-2(f)

RESPONSE: Corrected as suggested.

CHANGES:

- Correction made on page 15, line 30.

COMMENT: Caption of Figure 5: There are two "all" in "Figure shows all all simulated grid cells for each model"

RESPONSE: Thanks for catching this; corrected.

CHANGES:

- Correction made to Figure 5 caption on page 20.

COMMENT: P19, L1: region. (For soybeans, temperature effects are more complex; see Supple-mental Figure S5.)-> region (for soybeans, temperature effects are more complex; see Supplemental Figure S5).

RESPONSE: Corrected as suggested.

CHANGES:

- Correction made on page 22, line 2.

COMMENT: P20, L10: Generally, the carbon fertilization effect (CFE) would be larger under drier condition than under wetter condition. Is this true in Fig. 6a and Fig. S7? McGrath, J.M., & Lobell, D.B. (2013). Regional disparities in the CO2 fertilization effect and implications for crop yields. Environmental Research Letters, 8, 014054

RESPONSE: The GGCMI-2 experiment is designed to allow diagnosis of this and other interaction effects! But, as this is the experiment description paper, we are not analyzing results in full depth. We hope instead that many analyses will follow, making use of the freely available data set that we describe here. We have now added this citation and mentioned this effect as the possible target of a future study.

CHANGES:

- Citation adde to page 22, line 23.
- Text added to page 24, line 8.

COMMENT: P21, L1-2: again, please check the use of parenthesis.

RESPONSE: We have removed parentheses here.

CHANGES:

- Correction made on page 23, line 11. Text modifications for clarity.

COMMENT: Section 5: I am glad that the authors discussed some of the limitations in the simulation exercise. One more point should be included there is about how to validate the simulated responses, especially considering that there are indeed some field experiments designed to measure the responses of crops to environmental manipulations.

RESPONSE: We now include more discussion of the fact that models have been individually and jointly evaluated, including against data from field experiments.

We also discuss the challenges from the artificial model setup in the GGCMI Phase 2 experiment more thoroughly, and now refer to the companion paper (Franke et al., 2020), in which we demonstrate that emulators built from this artificial setup can very well reproduce model behavior from crop yield simulations driven by more realistic future climate projections.

CHANGES:

- Text added to address model validation on page 15, lines 6-9.
- Text added to address model validation against realistic climate scenarios with the emulator on page 23, line 31.

Reference:

Franke J, Müller C, Elliott J, Ruane AC, Jägermeyr J, Snyder A, Dury M, Falloon P, Folberth C, François L, Hank T, Izaurralde RC, Jacquemin I, Jones C, Li M, Liu W, Olin S, Phillips M, Pugh TAM, Reddy A, Williams K, Wang Z, Zabel F, and Moyer E. 2020, The GGCMI phase II emulators: global gridded crop model responses to changes in CO2, temperature, water, and nitrogen (version 1.0), Geosci. Model Dev. Discuss., 2020, 1-31, doi: 10.5194/gmd-2019-365.

The GGCMI Phase ~~II~~ 2 experiment: global gridded crop model simulations under uniform changes in CO_2, temperature, water, and nitrogen levels (protocol version 1.0)

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

[1]Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA
[2]Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA
[3]Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany
[4]Department of Computer Science, University of Chicago, Chicago, IL, USA
[5]NASA Goddard Institute for Space Studies, New York, NY, United States
[6]Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria
[7]Department of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic
[8]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France
[9]Sino-French Institute of Earth System Sciences, College of Urban and Env. Sciences, Peking University, Beijing, China
[10]Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium
[11]Met Office Hadley Centre, Exeter, United Kingdom
[12]Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany
[13]Georg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modeling, Göttingen, Germany
[14]Department of Geographical Sciences, University of Maryland, College Park, MD, USA
[15]Texas Agrilife Research and Extension, Texas A&M University, Temple, TX, USA
[16]Department of Statistics, University of Chicago, Chicago, IL, USA
[17]EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[18]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden
[19]Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA
[20]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.
[21]Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK.
[22]Leibniz Centre for Agricultural Landscape Research (ZALF), D-15374 MÃŒncheberg, 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 and soil 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 to which 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~~ 2 experiment, an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), is designed to provide a systematic parameter sweep focused on climate change factors and their interaction with overall soil fertility, to allow both evaluating model behavior and emulating model responses in impact assessment tools. In this paper we describe the GGCMI Phase ~~II~~ 2 experimental protocol and its simulation data archive. Twelve crop models simulate five crops with systematic uniform perturbations of historical climate, varying CO_2, temperature, water supply, and applied nitrogen ("CTWN") for rainfed and irrigated agriculture, and a second set of simulations represents a type of adaptation by allowing the adjustment of growing season length. We present some crop yield results to illustrate general characteristics of the simulations and potential uses of the GGCMI Phase ~~II~~ 2 archive. For example, in cases without adaptation, modeled yields show robust decreases to warmer temperatures in almost all regions, with a nonlinear dependence that ~~indicates~~ means yields in warmer baseline locations have greater temperature sensitivity. Inter-model uncertainty is qualitatively similar across all the 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 (Foley et al., 2005; 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 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 effect of temperature, water and nutrient availability, and atmospheric CO2 concentration on the process of photosynthesis and the biology and phenology of individual crops, play a critical role in assessing the impacts of climate change.

Process-based models are necessary for understanding crop yields in novel conditions not included in historical data, including higher CO_2 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; Elliott et al., 2014a; Frieler et al., 2017), options for climate mitigation (Müller et al., 2015) and adaptation (Challinor et al., 2018), and future sustainable development (Humpenöder et al., 2018; Jägermeyr et al., 2017). ~~Process-based models also allow for the~~ They are a necessity for global gridded simulations~~needed for~~ , which allow 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~~ can strongly modulate the economic impacts of regional yield changes (Stevanović et al., 2016; Hasegawa et al., 2018). Global ~~crop model experiments are needed for systematic climate change assessments (Müller et al., 2017)~~simulations are especially necessary in studying agricultural effects of climate change (Müller et al., 2017), since systematic climate assessments must account for cultivation area changes

and crop selection switching (Rosenzweig et al., 2018; Ruane et al., 2018) and must consider inter-regional differences (e.g. Nelson et al., 2
.

Modeling crop responses, however, continues to be challenging, as crop growth is a function of complex interactions between climate inputs, soil, and management practices (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 and projections of shifts in yield under future climate scenarios (e.g. Elliott et al., 2015; Müller et al., 2017). 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; Jägermeyr and Frieler, 2018), 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; Schewe et al., 2019). Global models pose additional challenges due to variable input data quality and limited ability for model calibration. Long-term projections therefore 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., 2017) and to assess the representation of individual physical mechanisms such as water stress and CO_2 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 scenarios.

The Global Gridded Crop Model Intercomparison (GGCMI) Phase ~~II~~ 2 experiment is the first global gridded crop model intercomparison involving a systematic parameter sweep across critical interacting factors. GGCMI Phase ~~II~~ 2 is an activity of AgMIP, and a continuation of a multi-model comparison exercise begun in 2014. The initial GGCMI Phase ~~I~~ 1 (Elliott et al., 2015; Müller et al., 2017) compared harmonized yield simulations over the historical period, with 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~~. See also Folberth et al. (2019) and Porwollik et al. (2017) for more information. GGCMI Phase 2 compares simulations across a set of inputs with uniform perturbations to historical climatology, including CO_2, temperature, precipitation, and applied nitrogen~~(collectively referred to as "CTWN")~~, as well as adaptation to shifting growing seasons ~~. The CTWN~~ (collectively referred to as "CTWN-A"). The CTWN-A experiment is inspired by AgMIP's Coordinated Climate-Crop Modeling Project ~~(C3MP Ruane et al., 2014; McDermid et al., 2015)~~ (C3MP, see Ruane et al., 2014; McDermid et al., 2015) and contributes to the AgMIP Coordinated Global and Regional Assessments ~~(CGRA) (Ruane et al., 2018; Rosenzweig et al., 2018)~~(CGRA, see Ruane et al.,
.

In this paper, we describe the GGCMI Phase ~~II~~ 2 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 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~~ 2 is to provide a comprehensive, systematic evaluation of the response of process-based crop models to critical interacting factors, including CO_2, temperature, water, and applied nitrogen under two contrasting assumptions on growing season adaptation (CTWN-A). The dataset is designed to allow researchers to:

- Enhance understanding of models ~~'~~ sensitivity to climate and nitrogen drivers.
- Study the interactions between climate variables and nitrogen inputs in driving modeled yield impacts.
- Characterize differences in crop responses to climate change across the Earth's climate regions.
- Provide a dataset that allows statistical emulation of crop model responses for downstream modelers.
- Explore the potential effects on future yield changes of adaptations in growing season length.

2.2 Modeling protocol

The GGCMI Phase ~~I~~ 1 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 14 modeling groups. The GGCMI Phase ~~II~~ 2 protocol is substantially larger, involving over 1400 individual 30-year global scenarios, or over 42,000 model-years; 12 modeling groups nevertheless participated. To reduce the computational load, the GGCMI Phase ~~II~~ 2 protocol reduces 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 (FAO, 2018). This set of major crops has the advantage of historical yield data globally available at sub-national scale (Ray et al., 2012; Iizumi et al., 2014), and has been frequently used in subsequent analyses (e.g. Müller et al., 2017; Porwollik et al., 2017).

The Phase ~~II~~ 2 protocol involves a suite of uniform perturbations from a historical climate timeseries. The baseline climate scenario for GGCMI Phase ~~II~~ 2 is one of the weather products used in Phase ~~I~~1, daily climate inputs for 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

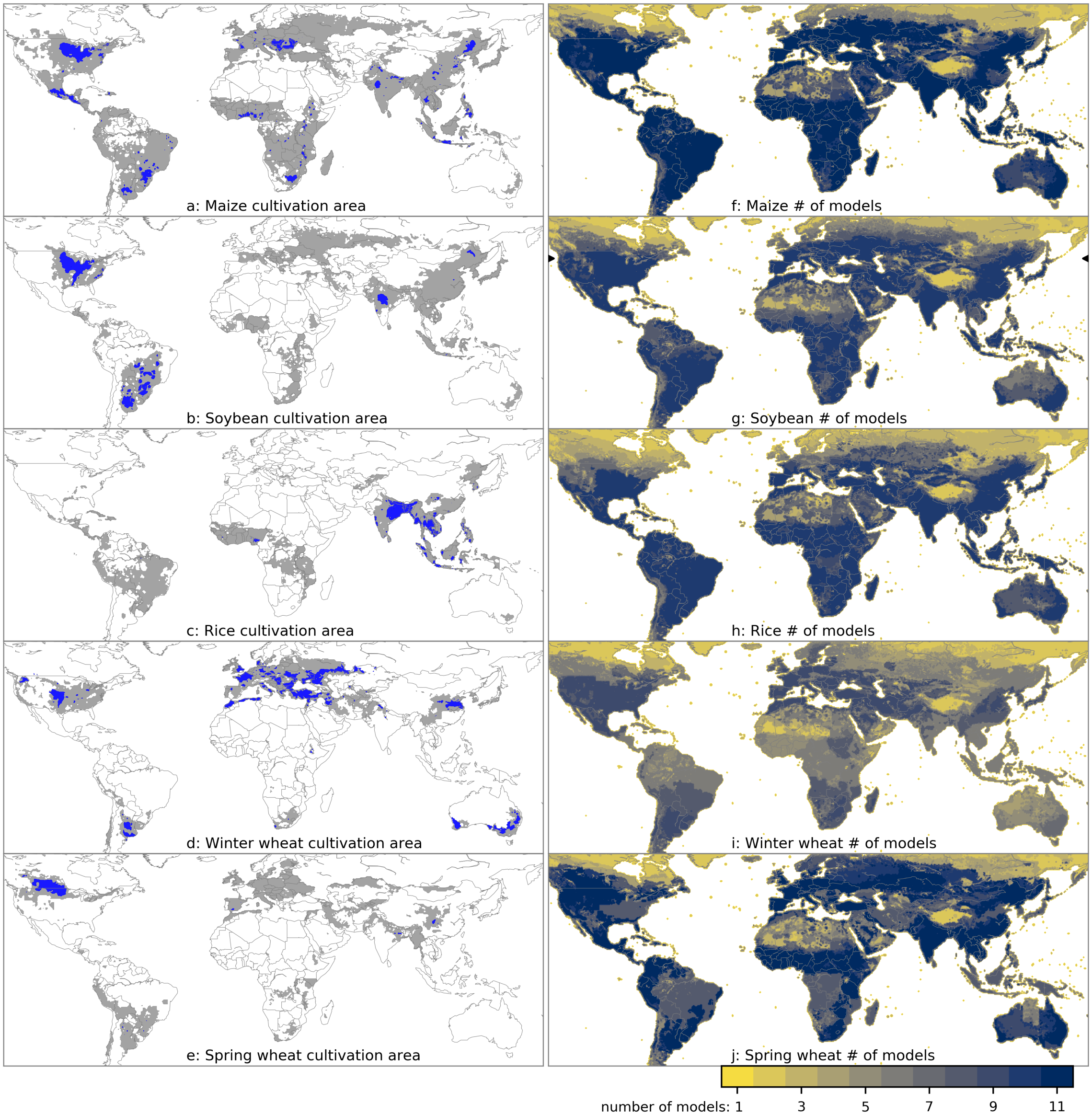

Figure 1. Left panel: Cultivated areas for maize, rice, and soybean ~~are taken~~ from the MIRCA2000 ("Monthly Irrigated and Rainfed Crop Areas around the year 2000") dataset (Portmann et al., 2010). Blue indicates grid cells with more that 20,000 hectares (10% of the equatorial grid cell) and gray contour shows gridcells with more that 10 hectares cultivated. Areas for winter and spring wheat areas are adapted from MIRCA2000 and two other sources; see text for details. For irrigated crops, see supplemental Figure S1. **Right panel:** Number of models providing simulations for each grid cell. All models provide the minimum areal coverage of the GGCMI Phase ~~II~~ 2 protocol, but some provide extra coverage at high latitudes or in arid or otherwise unsuitable areas.

Table 1. GGCMI Phase ~~II~~ 2 input parameter levels for each dimension. Temperature and precipitation values indicate the perturbations from the historical climatology. Irrigated (W_∞) simulations assume the maximum beneficial levels of water. Bold font indicates the ‘baseline’ or historical level for each dimension. One model provided simulations at the T + 5 level.

Input variable	Simulation input values	Unit
CO_2 (C)	**360**, 510, 660, 810	ppm
Temperature (T)	-1, **0**, 1, 2, 3, 4, 6	°C
Precipitation (W)	-50, -30, -20, -10, **0**, 10, 20, 30, (and W_∞)	%
Applied nitrogen (N)	10, 60, **200**	kg ha^{-1}
Adaptation (A)	**A0: none**, A1: new cultivar to maintain original growing season length	-

et al., 2015). The experimental protocol consists of 9 levels for water supply perturbations, 7 for temperature, 4 for CO_2, and 3 for applied nitrogen, for a total of 756 simulations (Table 1), 672 for rainfed agriculture and an additional 84 for irrigated (W_∞). ~~In irrigated simulations, crops are assumed to have no water constraints, i.e. all crop water requirements are fulfilled regardless of local water supply limitations. Given that the irrigated scenario (W_∞) is one element of the water supply levels, irrigated simulations use the same growing seasons and areas as all other simulations. All other water supply levels are implemented as relative variations of precipitation.~~ Values of climate variable perturbations are selected to represent reasonable ranges for changes over the medium term (to 2100) under business-as-usual emissions. Values for nitrogen application levels are intended to cover a wide range of potentials. The resulting GGCMI Phase ~~II dataset therefore~~ 2 dataset captures the distribution of crop model responses over ~~the~~ a wide range of potential future climate and management conditions.

~~While all~~ The protocol samples over all possible permutations of individual perturbations, i.e. all values are applied across all crops and regions, so that the protocol includes many combinations that are not realistic. For example, we simulate high N application to soybeans, which are N-fixers and need little fertilizer. This choice also means that CO_2 changes are applied independently of changes in climate variables, so that higher CO_2 is not associated with higher temperatures or other particular climate changes. The purpose of the experiment is not to produce individual scenarios that represent realistic future states, but to sample over a wide range of parameter space to enable understanding the factors that drive agricultural changes.

While all CTWN perturbations are applied uniformly across the historical timeseries, they are applied in different ways. ~~While precipitation perturbations are applied as fractional changes, temperature~~ CO_2 and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature timeseries for each grid cell~~. CO_2 and nitrogen levelsare specified as discrete values applied uniformly over all grid cells. The protocol samples over all possible permutations of individual perturbations . 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.~~ , and water perturbations are applied as fractional changes to daily precipitation. The irrigated scenario (W_{∞}) is a particular case of water supply levels, in which crops are assumed to have no water constraints. That is, all crop water requirements are fulfilled regardless of local water supply limitations. To facilitate comparison, irrigated simulations use the same growing seasons as all other simulations, even though in reality irrigated growing seasons may be different (Portmann et al., 2010), and both irrigated and rainfed cases are simulated with near-global coverage.

The uniform perturbations of the GGCMI Phase 2 protocol require some care in interpretation. Temperature and precipitation perturbations should be considered as differences from historical climatology within the growing season only. That is, a T+1 simulation represents a 1 °C warmer growing season, not a 1 °C warmer annual mean temperature. (The distinction is important because in climate projections, winters generally warm more than summers (e.g. Haugen et al., 2018).) In the GGCMI Phase 2 protocol, temperature and precipitation perturbations are applied uniformly in space, but future changes in temperature and precipitation will not be spatially or temporally uniform. In a realistic climate projection, higher latitudes generally warm more strongly than lower latitudes (e.g. Hansen et al., 1997), and the northern high latitudes warm more quickly than the southern ones. A GGCMI Phase 2 simulation therefore represents a possible future state that *could* occur in each grid cell, but not one that would in reality occur simultaneously in all grid cells across the globe. The GGCMI Phase 2 simulations are intended to be used for climate impact assessment not directly but instead as a "training set" for statistical emulation of each crop model. Once an emulator is constructed from the outputs described here, it can be driven with growing-season climate anomalies from any climate model projection. The GGCMI Phase 2 protocol does not involve any simulated changes in climate variability, but Franke et al. (2020) demonstrate that these effects are relatively minor and that GGCMI Phase 2 emulators can effectively reproduce crop model yields under realistic future climate scenarios.

~~Each model is run at 0.5 degree spatial resolution and covers all~~ The area simulated in the GGCMI Phase 2 protocol extends considerably outside currently cultivated areas~~and much of the uncultivated land area~~, because cultivation may shift under climate change. Figure 1 ~~, left, shows~~ shows both the present-day cultivated area of rainfed crops ~~and~~ (left) and model coverage (right). (See Supplemental Figure S1-2 ~~that for irrigated crops. Cultivated areas are provided by the MIRCA2000 (Monthly Irrigated and Rainfed Crop Area) data product (Portmann et al., 2010).Coverage extends considerably outside~~ for currently cultivated area for irrigated crops; model coverage is the same.) Each model covers all currently cultivated areas ~~because cultivation will likely shift under climate change~~and much of the uncultivated land area, run at 0.5 degree spatial resolution. To reduce the computational burden, ~~however,~~ the protocol requires simulation ~~only over~~ over only 80% of Earth land surface area~~. Areas are not simulated if they are~~ , omitting areas assumed to remain non-arable even under an extreme climate change~~; these regions include~~ , including Greenland, far-northern Canada, Siberia, Antarctica, the Gobi and Sahara Deserts, and Central Australia. The protocol also ~~eliminates~~ allows omitting 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. Currently cultivated areas are provided by the MIRCA2000 (Monthly Irrigated and Rainfed Crop Area) data product (Portmann et al., 2010). 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 (Figure 1, right).

2.3 Harmonization between models

The 12 models included in GGCMI Phase ~~II~~ 2 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~~ or EPIC families of models), they have subsequently developed independently. Wherever possible, the GGCMI Phase ~~II~~ 2 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~~ 2 use the AgMERRA historical daily climate data product, two models require sub-daily input data and thus use different baseline climate inputs: PROMET uses ERA-Interim reanalysis (Dee et al., 2011)~~;~~, and JULES uses a bias-corrected version of ERA-Interim, the 3-hour WFDEI (WATCH-Forcing-Data-ERA-Interim) (Weedon et al., 2014), ~~selecting~~ specifically the WFDEI version with precipitation bias-corrected against the CRU TS3.101/TS3.21 precipitation totals (Harris et al., 2014). The data products show some differences (Figures S3-S4, which compare data products over currently cultivated areas for each crop). For example, for ~~maize~~maize-growing areas, ERA-Interim daily precipitation is biased high from that in AgMERRA by 7% (< 1 sigma), while mean daily precipitation in WFDEI is only 3% higher. Precipitation differences are largest in wheat areas, where ERA-Interim is substantially wetter (+~~60mm~~ 60 mm year^{-1} or 10%). ~~Temperatures for maize are very similar between data products~~Temperature differences are largest for rice, with ERA-Interim ~~0.45~~1°C cooler ~~and WFDEI 0.1°C warmer. Differences are largest for rice,~~ than AgMERRA, and smaller for other crops, e.g. maize with ERA-Interim ~~1~~0.45°C cooler and WFDEI 0.1°C warmer. These differences are 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~~ (The exception is that Phase 2, unlike Phase 1, uses identical growing seasons for rainfed and irrigated ~~scenarios~~cases, to allow for direct ~~comparability~~ comparison of simulations along the W dimension~~, in which irrigation (W_∞) is one element. (See Table 1.) While~~ .) This harmonization is important because the parametrization of growing seasons can have strong effects on simulated yields (Müller et al., 2017; Jägermeyr and Frieler, 2018). In all the GGCMI Phase 2 crop models, sowing dates are prescribed directly~~and held fixed in models,~~, but the length of the growing season is a product of crop phenology, which ~~in turn is mostly driven~~ is driven mostly by phenological parameters and temperature. Modelers ~~are~~ were therefore asked to adjust ~~the~~ their phenological parameters so that the average growing season length of the baseline scenario (C=360, T=0, W=0) ~~on average matches~~ matched the harmonization target. ~~Given that temperature varies between years, individual years can vary from the harmonization target. Growing seasons are harmonized across models but are~~ (The one exception to this harmonization protocol involves CARAIB, whose team kept their own growing season specifications rather than tuning to standard lengths.) Two aspects of the procedure should be noted. First, the target growing seasons used in GGCMI Phase 2 are crop- and location-specific. For example, ~~at present~~ present-day maize is sown in March in Spain, in July in Indonesia, and in December in Namibia (Portmann et al., 2010). ~~The one exception to the harmonization protocol described above involves CARAIB, which for technical reasons kept their own growing season specifications rather than tuning to standard lengths~~Second, because temperature varies between years in the 30-year baseline climatology, realized growing season length will still vary in individual years even after harmonization.

~~Because harvest dates are a function of climate parameters , simulations with the harmonized phenological parameters described above generally result in shorter~~ The dependence of harvest dates on climate parameters means that growing seasons will alter under climate change in a model with phenological parameters tuned to match target growing seasons in ~~future warmer scenarios~~the baseline climate. In general, warmer future scenarios produce shorter growing seasons. We denote ~~these simulations~~ simulations that allow these future changes as “A0” experiments, where 0 denotes “no adaptation”. ~~To account for potential adaptation in crop cultivars, the GGCMI Phase II~~ The GGCMI Phase 2 protocol includes a second set of experiments, “A1”, that assume that future cultivars are modified to adjust to changes along the T dimension in the CTWN experiment. For these simulations, modelers adjust ~~parameters~~ phenological parameters for each temperature scenario to hold growing season length approximately constant~~across the different warming scenarios~~. (CARAIB simulations follow the same principle, fixing growing season length at their baseline levels.) That is, the A1 simulations require running a model with seven different choices of cultivar parameters, one per warming level. Parameter settings for T=0 are identical in both A0 and A1. The A1 simulations roughly capture the case in which adaptive crop cultivar choice ensures that crops reach maturity at roughly the same time as in 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 and is sufficiently easy to implement in a large model intercomparison project as GGCMI.

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), gap-filled with the MIRCA2000 crop calendar (Portmann et al., 2010) and, if no SAGE or MIRCA2000 data are available, with simulated LPJmL growing seasons (Waha et al., 2012) and are identical to those used in GGCMI Phase ~~I~~1 (Elliott et al., 2015). In GGCMI Phase ~~II~~2, we separately treat spring and winter wheat and so must define different growing seasons for each. As for the other crops, we use the SAGE crop calendar, which separately specifies spring and winter wheat, as the primary source for 69% 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 season is assigned to winter wheat if all of the following hold, and to spring wheat otherwise:

- the monthly mean temperature is below freezing point (<0°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°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)

Nitrogen (N) application is 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~~ 2 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~~ 2 protocol, the second fertilization date for winter wheat must lie at least 40 days after planting and – if not contradicting the distance to planting – no later than 50 days before maturity. If those limits permit, the second fertilization is set to the middle day of the first month after sowing that has average temperatures above 8°C.

All stresses in models are disabled other than those related to nitrogen, temperature, and water. For example, model responses to alkalinity, salinity, and non-nitrogen nutrients are all disabled. 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. In LPJmL, LPJ-GUESS and APSIM, soil mineralization is a part of model treatments of soil organic matter and cannot be disabled. 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 if available (Table 2, bold). For~~ 2 provide 30-year timeseries of annual crop yields for each scenario, 0.5 degree grid cell and crop, ~~models provide 30-year timeseries of annual crop yields~~ in units of tons ha^{-1} year^{-1}~~, as well as~~. They also provide all available variables of the following 6: total above-ground biomass yield; the dates of planting, anthesis ~~,~~ and maturity; applied irrigation water in irrigated scenarios; and total ~~evapotranspirtation. (Note that several~~ evapotranspiration. We term these 7 variables the “mandatory” outputs, but note that some models do not ~~output~~ compute all of them, e.g. CARAIB does not compute the anthesis date. ~~) Besides these mandatory~~ Besides these 7 ~~data products,~~ “mandatory” data products (Table 2, bold), the protocol requests any or all of 18 ~~optional~~ “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. File names follow the naming conventions of GGCMI Phase ~~I~~ 1 (Elliott et al., 2015), which themselves are derived from those of ISIMIP (Frieler et al., 2017). File names are specified as

Table 2. Output variables, naming convention, and units in the GGCMI Phase ~~II~~ 2 protocol. Items in **bold** are the mandatory minimum requirements (if model capacities allow for these outputs). Other variables are optionally provided depending on availability and participating modeling groups provided between 3 (PEPIC) and 18 (APSIM-UGOE) of these optional variables.

Variable	variable name	units
Yield	**yield_<crop>**	**t ha^{-1} yr^{-1} (dry matter)**
Total above ground biomass yield	**biom_<crop>**	**t ha^{-1} yr^{-1} (dry matter)**
Actual planting date	**plant-day_<crop>**	**day of year**
Anthesis date †	**anth-day_<crop>**	**days from planting**
Maturity date	**maty-day_<crop>**	**days from planting**
Applied irrigation water	**pirrww_<crop>**	**mm yr^{-1}**
Evapotranspiration (growing season sum)	**etransp_<crop>**	**mm yr^{-1} (W_∞ scenarios only)**
Transpiration (growing season sum)	transp_<crop>	mm yr^{-1}
Evaporation (growing season sum)	evap_<crop>	mm yr^{-1}
Runoff (total growing season sum, subsurface + surface)	runoff_<crop>	mm yr^{-1}
Total available soil moisture in root zone *	trzpah2o_<crop>	mm yr^{-1}
Total root biomass	rootm_<crop>	t ha^{-1} yr^{-1} (dry matter)
Total Reactive Nitrogen (Nr) uptake (growing season sum)	tnrup_<crop>	kg ha^{-1} yr^{-1}
Total Nr inputs (growing season sum)	tnrin_<crop>	kg ha^{-1} yr^{-1}
Total Nr losses (growing season sum)	tnrloss_<crop>	kg ha^{-1} yr^{-1}
Gross primary production (GPP)	gpp_<crop>	gC m^{-2} yr^{-1}
Net primary production (NPP)	npp_<crop>	gC m^{-2} yr^{-1}
CO_2 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	nrnpp_<crop>	- {0..1}
Other nutrient response scaler on NPP	ornpp_<crop>	- {0..1}
CO_2 response scaler on transpiration	co2trans_<crop>	- {0..1}
Maximum stress response scaler	maxstress_<crop>	- {0..1}
Maximum Leaf Area Index (LAI)	laimax_<crop>	m^2 m^{-2}

* growing season sum, basis for computing average soil moisture

† provided where possible

$$[model]\_[climate]\_hist\_fullharm\_[variable]\_[crop]\_global\_annual\_[start-year]\_[end-year]\_[C]\_[T]\_[W]\_[N]\_[A].nc4$$

Here $[model]$ is the crop model name; $[climate]$ is the original climate input dataset (typically AgMERRA); $[variable]$ 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" for winter wheat); and $[start-year]$ and $[end-year]$ specify the first and last years recorded on file. $[C]$, $[T]$, $[W]$, $[N]$ and $[A]$ indicate the CTWN-A settings, each represented with the respective uppercase letter and the number indicating the level (e.g. "C360_T0_W0_N200" see Table 1). Except for the CTWN-A letters, the entire file name needs to be in small caps. All filenames include the identifiers $global$ and $annual$ to distinguish them as global, annual model output, following the updated ISIMIP file naming convention (Frieler et al., 2017).

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 reported as "missing values" (defined as 1.e+20 in the netCDF files). 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~~1 standards, the first entry in each file describes the first complete cropping cycle simulated from the given climate input timeseries. In the AgMERRA timeseries used for GGCMI Phase ~~II~~2, the first year provided is 1980 but the date of the first entry can vary by crop and location. In the northern hemisphere, for summer crops like maize (sown in spring 1980 and harvested in fall 1980), the first harvest record would be of 1980, but for winter wheat (sown in fall 1980 and harvested in spring 1981) the first harvest record would be of 1981. Output files report the sequence of growing periods rather than calendar years. While there is generally one sowing event per calendar year (since simulations with harmonized growing seasons do not permit double-cropping), in some cases harvest events may skip or repeat within a calendar year. For example, because soybeans in North Carolina are typically harvested well into December, some calendar years may include no harvest (if it is not completed until after Dec. 31) or two harvests (one in January and one 11 months later in the following December).

3 Models contributing

The 12 models participating in GGCMI Phase 2 are listed in Table 3. Models differ substantially in structure and parameterization and can be separated into two broad categories: site-based (field-scale) models, and global ecosystem models. The 6 site-based models are APSIM, pDSSAT, and the EPIC family of models; the 6 ecosystem models are LPJmL, LPJ-GUESS, PROMET, CARAIB, ORCHIDEE, and JULES. Models employ a variety of approaches for the core modules such as primary production or evapotranspiration. For primary production, site-based models employ light use efficiency approaches and ecosystem models use photosynthesis approaches. For evapotranspiration, most models use Priestley-Taylor, Penman-Monteith or Hargreaves schemes, but JULES and PROMET utilize a land surface model approach instead. Note that models that share a common genealogy may still use different schemes for evapotranspiration: for example, EPIC-TAMU uses Penman-Monteith and EPIC-IIASA uses Hargreaves. To describe soils, most models use either the Harmonized World Soil Database (HWSD)

Table 3. Models included in GGCMI Phase ~~II~~ 2 and the number of CTWN-A simulations performed. The maximum number is 756 for A0 (no adaptation) experiments, and 648 for A1 (maintaining growing length) experiments, since T0 is not simulated under A1. “N-Dim.” indicates whether the models are able to represent varying nitrogen levels. Each model provides the same set of CTWN simulations across all its modeled crops, but some models omit individual crops. (For example, APSIM-UGOE does not simulate winter wheat.)

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

from the FAO (Fischer et al., 2008) or the ISRIC-WISE database (Batjes, 2005) or a derivation thereof. Supplemental Table S1 provides details on these model characteristics as well as on implementation, including spin-up, calibration other than growing season, residue management, and irrigation rules.

~~The simulation output contributions of the 12 crop models~~ Table 3 also describes the simulation output contribution of each model to the GGCMI Phase ~~II archiveare described in Table 3~~2 archive. Not all modeling groups provided simulations for 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 most differed in CTWN coverage of A0 and A1 scenarios. Since the adaptation dimension was defined as a secondary priority for GGCMI Phase ~~II~~2, some models provided a more limited set of A1 scenarios. Of these, EPIC-IIASA, JULES, and ORCHIDEE-crop provided no A1 scenarios.

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)

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

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, 510, 660, 810 ppm) and 3 N (10, 60, 200 kg ha^{-1} yr^{-1})
- *low* as only 4 pairs: C360_N10, C360_N200, C660_N60, C810_N200

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

4 Results

To illustrate the properties of the GGCMI Phase ~~II~~ 2 model simulations, we provide an evaluation of model performance by comparing model and historical yields, and show example results that demonstrate the spread of model responses to climate and management inputs.

4.1 Evaluation of model performance

All models participating in GGCMI Phase 2 have be evaluated against historical yields and site specific experimental data. Most models (9 of 12, all but CARAIB, JULES, and PROMET) have been evaluated in their global setup in the GGCMI Phase 1 evaluation exercise (Müller et al., 2017), and many have used the GGCMI Phase 1 online tool to similarly evaluate subsequent model versions (e.g. von Bloh et al., 2018). Evaluating the performance of crop models in the GGCMI Phase ~~II~~ 2 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 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), developed for GGCMI Phase ~~I.~~ 1. Müller et al. (2017) evaluate how well model crop yield responses in a historical run capture real-world yield variations driven by year-to-year temperature and precipitation variations. Following this approach, we compare yields in the GGCMI Phase ~~II~~ 2 baseline simulations with detrended historical yields from the Food and Agriculture Organization of the United Nations (FAO, 2018) by calculating the Pearson product moment correlation coefficient over 26 years of yield. 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 timeseries are shifted by one year to account for discrepancies in FAO or model year reporting. Because the GGCMI Phase ~~II~~ 2 protocol imposes fixed, uniform nitrogen application levels that are not realistic for individual countries, we evaluate control runs for each model at multiple N levels whenever possible. Nine of the GGCMI Phase ~~II~~ 2 models (Table 3) provide historical runs for all three nitrogen levels (10, 60, and. 200 kg ha^{-}1 yr^{-}1).

As expected due to the unrealistic features described above, correlation coefficients for the GGCMI Phase ~~II~~ 2 simulations are slightly lower than those found in the Phase ~~I~~ 1 evaluation, but models show reasonable fidelity at capturing year-over-year variation (Figure 2). For example, global correlation coefficients ~~for maize in Phase I and Phase II~~ in Phase 1 and Phase 2, respectively, are 0.89 and 0.74 ~~, respectively; for wheat~~ for maize, 0.67 and 0.64 ~~, and for soybeans~~ for wheat, and 0.64 and 0.59 ~~. (Compare to Müller et al. (2017)~~ for soybeans. (Phase 1 values are from Figures 1–4 and ~~6.~~6 in Müller et al. (2017).) Differences in fidelity between regions and crops exceed differences between models: that is, Figure 2(c)–2(f) 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

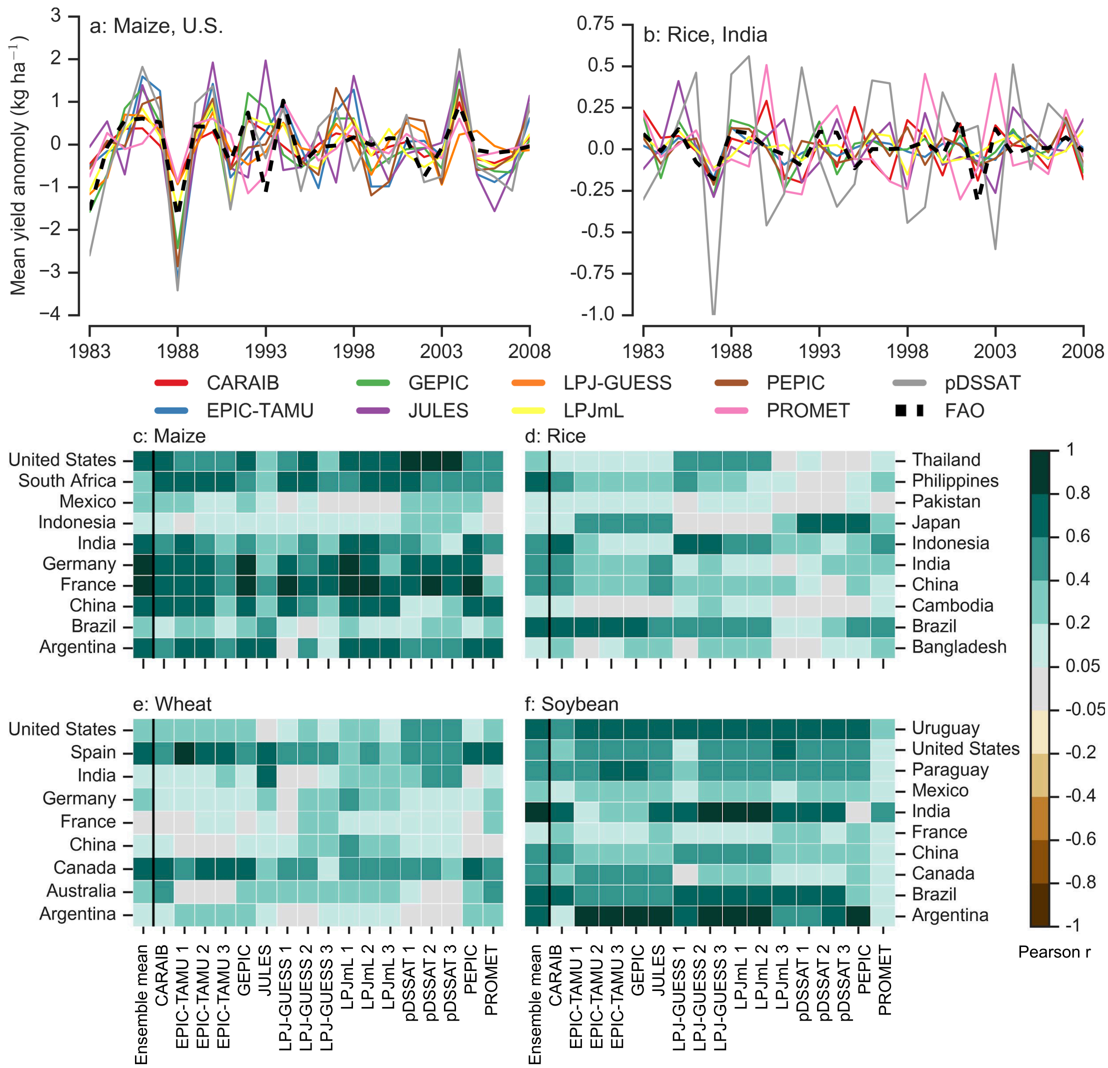

Figure 2. Assessment of crop model performance in GGCMI Phase ~~II~~2, following the protocol of GGCMI Phase ~~I~~ 1 (Müller et al., 2017). **Top:** example timeseries comparison between simulated crop yield and FAO country statistics (FAO, 2018) at the country level for two example high production countries: US maize, and rice in India, both for the 200 kg ha^{-1} nitrogen application level. **Bottom:** heatmaps illustrating the Pearson r correlation coefficient between the detrended simulated and observed country-level mean yields for the top 10 countries by production for each crop, of those countries with continuous FAO data over 1981-2010. We show separate comparisons for simulations with the three different nitrogen application levels, denoted 1, 2, 3 for 10, 60, and 200 kg N ha^{-1}, respectively. Left column shows correlation of ensemble mean yields with FAO data Because FAO does not distinguish between wheat types, we sum simulated spring and winter wheat for models that provide both (See Table 3.). Note that differences by region and crop are stronger than difference between models, e.g. horizontal bars are more similar in color than vertical bars. Countries are ordered alphabetically, not by production quantity.

with higher variability. In some cases, especially in the developing world, low correlation coefficients may point to reporting problems in the FAO statistics and to real-world variability caused by variations in management rather than weather (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, pDSSAT is the best model for maize in the US, LPJmL and GEPIC are best in Germany, PROMET is best in Argentina, and PEPIC and LPJ-GUESS are best in France.

4.2 Model crop yield responses under CTWN forcing

Crop models in the GGCMI Phase ~~II~~ 2 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 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 patterns (Figure 3, right, which shows fractional yield differences between the baseline and T+4 ~~scenario and the baseline scenario with historical climatology~~A0 scenarios). In general, models agree most on yield response in regions where yield potentials are currently high and therefore where crops are currently grown. ~~Models~~ In A0 simulations, models show robust decreases in yields at low latitudes, and highly uncertain ensemble mean increases at most high latitudes. ~~Models~~ Low latitude yield reductions are due in part to shortening of the growing season under warming and in part to the direct effects of higher temperature. In A1 simulations, where growing seasons length does not change, temperature-related reductions in yield are more muted (see Supplemental Figure S14). In both A0 and A1 simulations, models show some increases in high mountain regions that are currently cold-limited.

Projections of strong yield growth at higher latitudes should be treated with caution, since the effects evident in Figure 3 are due in part to inaccuracies in model representations of present-day crop yields. For example, at latitudes north of 45°, the GGCMI Phase ~~II~~ 2 models collectively suggest strong (but uncertain) growth in soybean yields under warmer conditions (Figure 3, g). However, model differences are greater in the baseline than future simulations, and greatest in currently-cultivated areas (Figure 4). Both the mean projected growth and the inter-model spread are driven by three models that show almost zero present-day potential soybean yields across the entire high-latitude region, even in locations where soybeans are currently grown (Figure 4, left). PROMET, for example, involves a stronger response to cold than other models (e.g. LPJmL) with frost below -8 °C irreversibly killing non-winter crops and prolonged periods of below-optimum temperatures also leading to complete crop failure. Over the high-latitude 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). These spurious low baseline yields result in very large fractional changes in the T+4 warming scenario, when all models agree that conditions become favorable for soybeans. Those models that most accurately reproduce present-day high-latitude soybean yields of 1-2 ton ha^{-1} (Ray et al., 2012) in fact show a slight decrease in yield under a warming scenario (Figure 4, left). Apparent future yield increases in the multi-model mean are driven by the least realistic simulations.

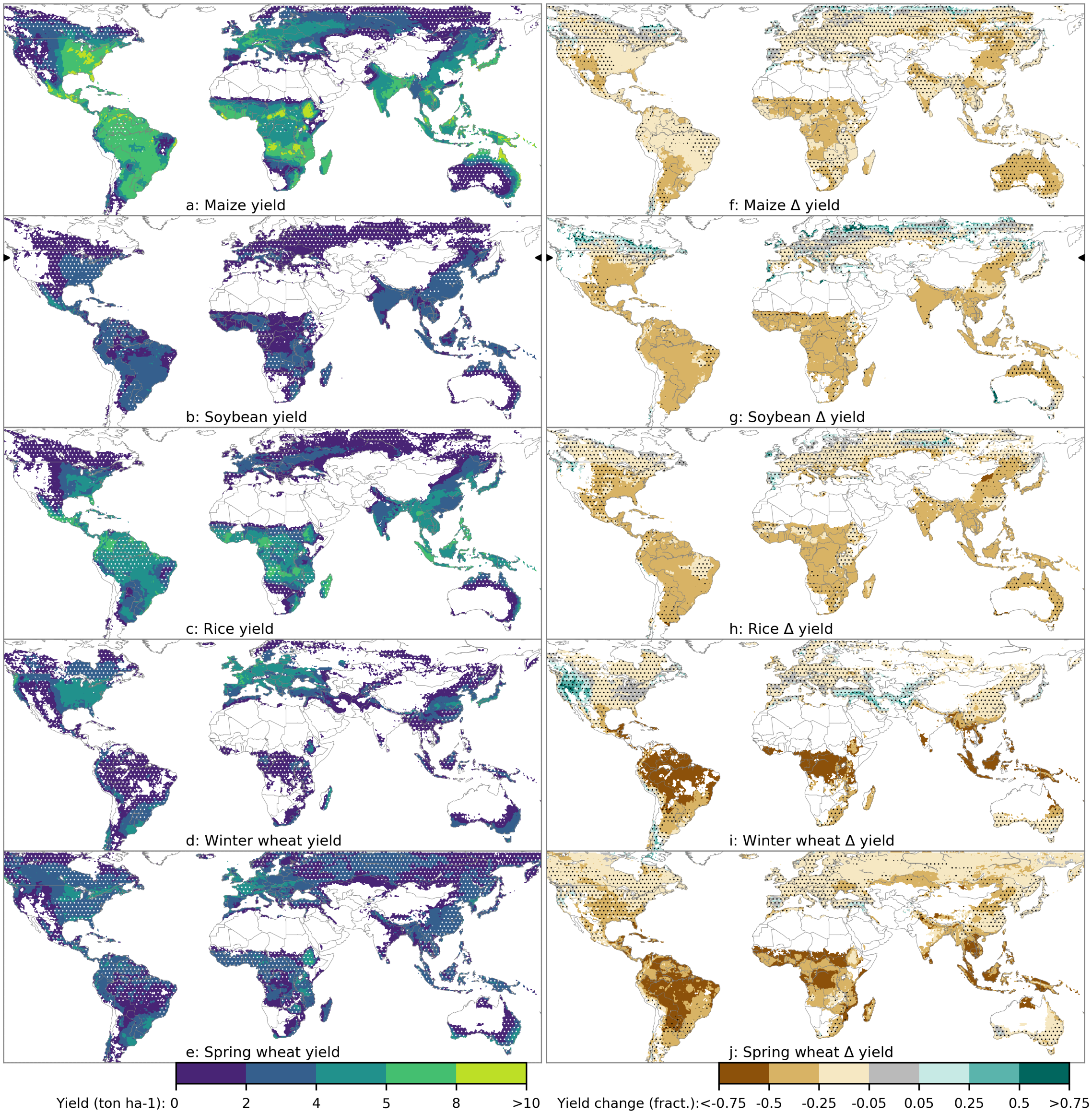

Figure 3. Illustration of the spatial patterns of baseline yields (left) and yield changes (right) in the GGCMI Phase ~~II~~ 2 ensemble. Left column shows multi-model climatological(30 year) median yields for the baseline scenario, with white stippling indicating areas where these crops are not currently cultivated. Areas with less ~~that~~ than 0.5 ton ha^{-1} in the baseline are masked. Absence of cultivation aligns well with the lowest yield contour (0-2 ton ha^{-1}). Right column shows multi-model mean fractional yield changes in the T+4 °C scenario relative to the baseline scenario. Areas without stippling are those where models agree on changes: the multi-model mean fractional change exceeds the standard deviation of changes in individual models. Stippling indicates areas of low confidence ($\Delta < 1\sigma$). Some spatial structure in projected changes at high latitudes may be due to differences in model coverage; see Figure 1.

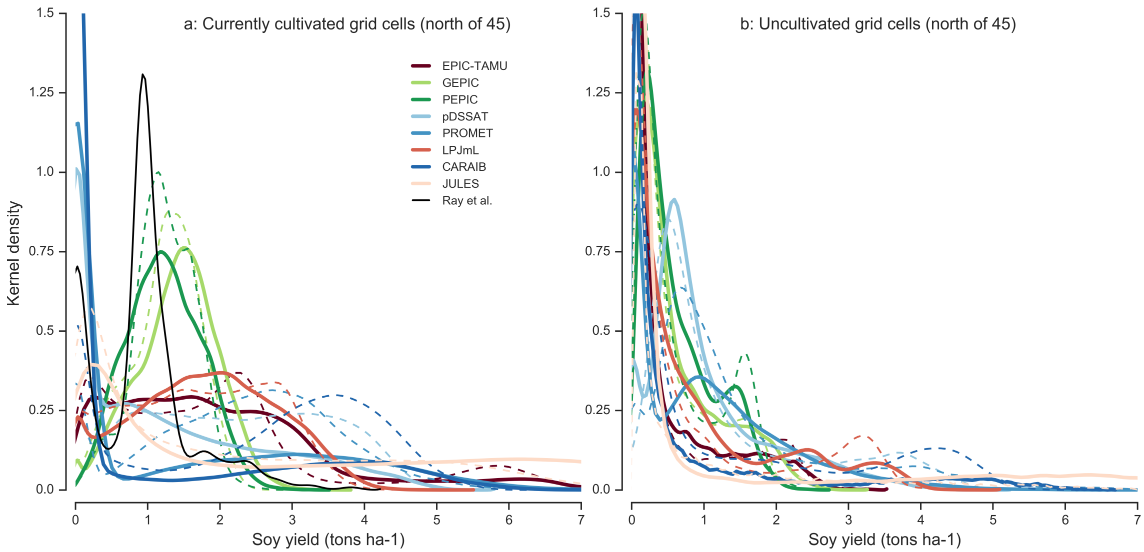

Figure 4. Model probability densities for soybean yields at latitudes north of 45° in historical and warming simulations in the A0 case. While 10 GGCMI Phase ~~II~~ 2 models provide simulations (Table 3); we show 8 representative models for clarity. Probability density functions are estimated separately for locations with some current cultivation (left, approximately 2500 grid cells, unweighted by cultivated area) and for uncultivated locations (right, approximately 1500 grid cells), for baseline historical (solid) and T+4 (°C) (dashed) simulations. Black line in left panel shows actual yields from 1997-2003 derived from Ray et al. (2012). For historical simulations, models agree on low potential yields in currently uncultivated areas (right) but disagree widely on yields in currently cultivated areas (left). Color code groups models into those with realistic yield distributions peaking at 1-2 ton ha^{-1} (green), those with flatter distributions extending to unrealistically high values (red), and those with predominantly zero yields (blue). "Green" models show slight decreases under T+4 warming, "red" models moderate increases, and "blue" models large increases.

The GGCMI Phase ~~II~~ 2 exercise offers the opportunity to examine and characterize not just crop response to a single temperature change but nonlinearities in responses and interactions between factors. We illustrate a few of these relationships in Figures 5-6 ~~, choosing~~ using A0 simulations to capture maximum climate effects. We choose crops and factors whose effects are reasonably well understood, and show that these are reproduced in models. It is expected, for instance, that increases in precipitation should buffer the effects of warmer temperatures and that CO_2 increases should reduce damage to crops in scenarios where water is limited. Models generally confirm expected behavior but also provide insight into unforeseen interactions. To show geographic effects, we divide model responses in Figures 5-6 by the primary Köppen-Geiger climate regions (Rubel and Kottek, 2010), showing the yield changes across all simulated grid cells in each region. In each panel we examine relationships between two factors, showing yield response against one for several scenarios of the other, in box plots that show the inter-model spread. The responses highlighted here are qualitatively similar across all crops included in this study (Supplemental Figures ~~S5 - S8~~S5–S9 for all simulated area and S10–S13 for cultivated area only).

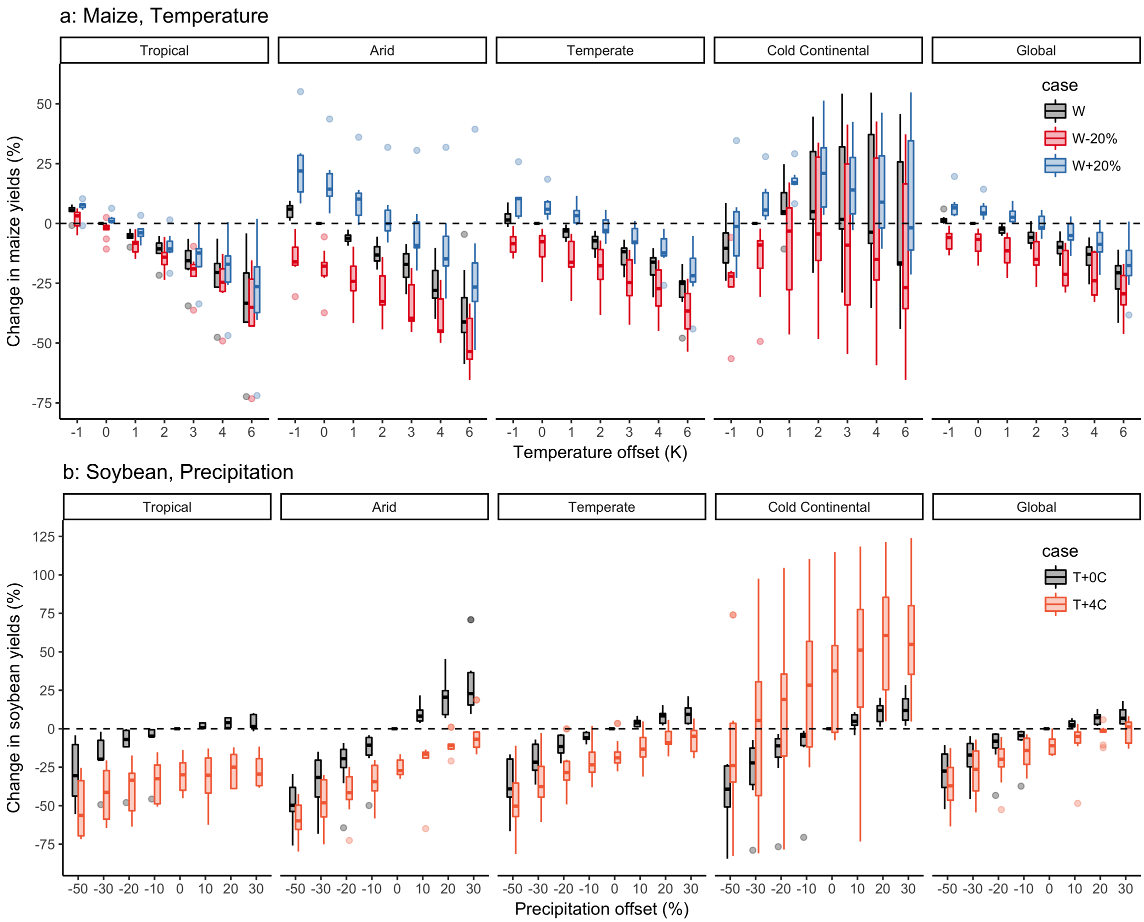

Figure 5. Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions, and with global response in rightmost panel. Y-axis is the fractional change in the regional average climatological (30-year mean) potential yield relative to the baseline. Box-and-whiskers plots show distribution across models, with median marked; edges are first and third quartiles and whiskers extend to 1.5·IQR. Figure shows all ~~all~~ simulated grid cells for each model; see Supplemental Figure S10-S13 for only currently-cultivated land. We highlight responses to individual factors; note that results are not directly comparable to simulations of realistic projected climate scenarios with identical global mean changes. Models generally agree outside high-latitude regions, with projected changes exceeding inter-model variance. **Top:** Response of 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 pmm and 200 kg ha^{-1} yr^{-1}). Outliers in the tropics (strong negative impact of higher T) are the pDSSAT model; outliers in the arid region (strong positive impact of higher P) are JULES. **Bottom:** Response of rainfed soybeans to applied uniform precipitation perturbations, for two discrete temperature levels. Cases with reduced precipitation show greater inter-model spread than those with increased precipitation. At very large precipitation increases, yield changes level out: benefits saturate once water availability is no longer limiting. Precipitation changes are more important in the arid region, as expected. Note the large uncertainty in the cold continental region, also illustrated in Figures 3 and 4.

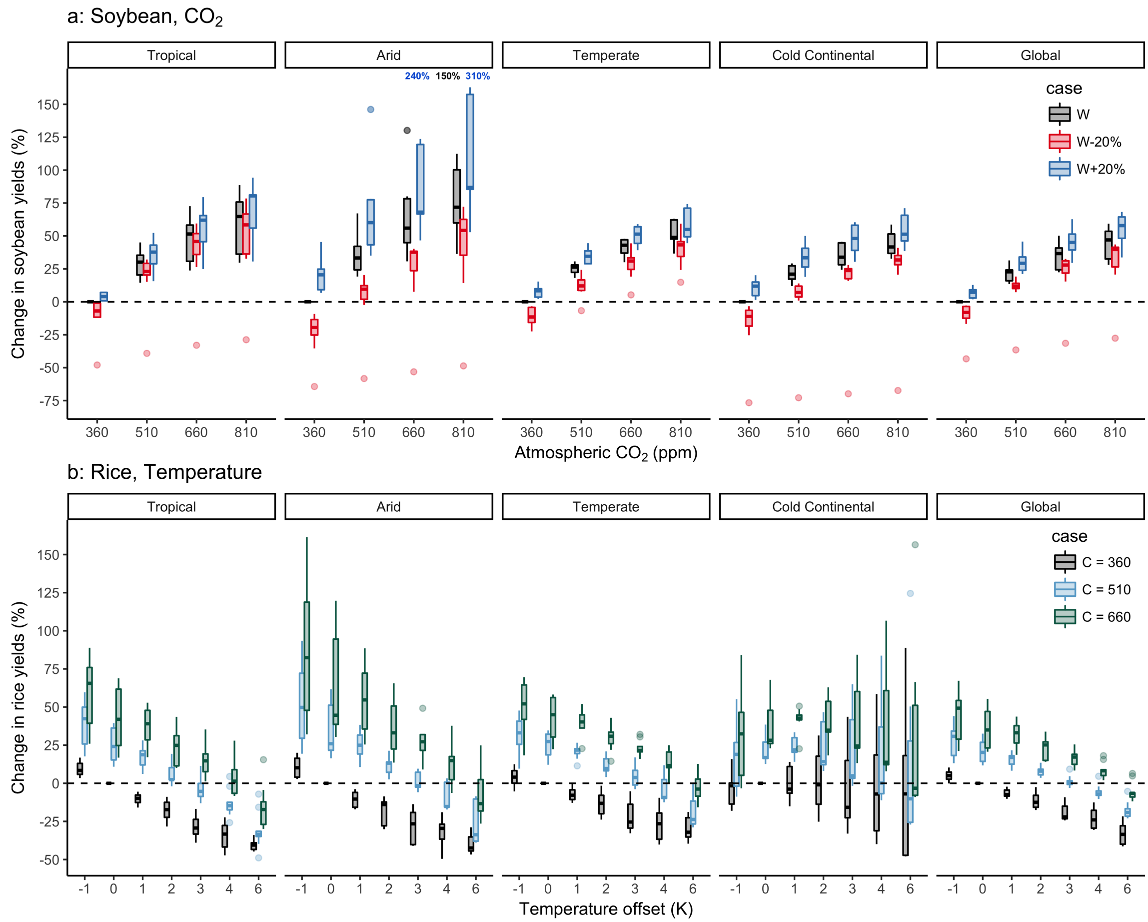

Figure 6. Illustration of the distribution of regional yield changes across the multi-model ensemble, here for soybeans and rice for the A0 case. Conventions as in Figure 5. **Top:** Response of rainfed soybeans to atmospheric CO_2, for three discrete precipitation perturbation levels with temperature and nitrogen held constant at baseline values. Low outliers are the EPIC-TAMU model and the high outliers in the Arid region are the JULES model. Reduced precipitation tends to steepen the CO_2 response and increased precipitation tends to flatten it, as expected. Reduced precipitation tends to increase the inter model spread, especially at the highest CO_2 levels. **Bottom:** Response of 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.

For all crops, warming scenarios with precipitation held constant produce yield decreases in most regions. These impacts are robust for even moderate climate perturbations. For rainfed maize, even a 1°C temperature increase with other factors held constant produces a median regional decline in potential yield that exceeds the variance across models, in all but the "cold-continental" regions (Figure 5a). The remaining areas ("warm temperate", "equatorial", and "arid") account for nearly three-quarters of global maize production. In the high-latitude "cold-continental" region, potential yield changes are positive but highly uncertain, for the reasons discussed previously; uncertainties are larger even for maize than for soybeans. (Compare

Figures 5a and 5b.) Temperature effects are somewhat nonlinear, with the largest impacts for maize in the warm "tropical" region ~~. (For~~ (for soybeans, temperature effects are more complex; see Supplemental Figure S5~~.)~~). Precipitation effects on rainfed crops are more strongly nonlinear. The curvature of the precipitation response can be seen by eye in Figure 5b: soybean yields are strongly negatively impacted by reduced rainfall, peak under increased precipitation of 20%, and actually decline at higher precipitation levels.

As expected, precipitation and temperature effects interact, with increases in precipitation buffering yield responses to temperature. Increased rainfall mitigates the negative impacts of warmer temperatures caused by increased evapotranspiration (e.g. Allen et al., 1998). For maize, the effect is relatively modest outside the "arid" regions (Figure 5a). Globally, a 4°C temperature rise with no change in precipitation results in median loss of ∼13% of rainfed maize, with all models showing a negative response. With a 20% increase in precipitation, the median yield loss is ∼8%. For soybeans, the equivalent values are ∼11% and 1%, respectively. Decreased rainfall, on the other hand, amplifies yield losses and also increases inter-model variance. That is, models agree that the response to decreased water availability is negative in sign but disagree on its magnitude. Outside of arid regions, the interaction effect itself shows little nonlinearity (i.e. response slopes in Figures 5a and 5b are roughly parallel). As expected, irrigated crops are more resilient to temperature increases in all regions, especially so where water is already limiting (other than winter wheat, Supplemental Figure S9).

Increased CO_2 boosts yields overall through the well-known CO_2 fertilization effect (Figure 6). The effect is strongest for the C3 crops (wheat, soybeans, and rice), while maize, a C4 grass, has a comparatively muted response. We show irrigated rice and rainfed soy in Figure 6 as representative C3 crops. The effect of CO_2 on yields is nonlinear, as expected, with significant benefit from small increases but with effects plateauing at higher concentrations (Figure 6). CO_2 and temperature effects show minimal interaction. This effect is seen in Figure 6a, which shows nearly parallel response slopes at different CO_2 levels. That is, CO_2 fertilization does little to change the nature of the temperature response. On the other hand, CO_2 and precipitation effects interact strongly, as expected since higher CO_2 levels allow reduced stomatal conductance and evapotranspiration losses, mitigating the effect of reduced rainfall (e.g. McGrath and Lobell, 2013). This interaction is seen in Figure 6b as smaller yield losses from reduced rainfall when CO_2 levels are higher. For example, for soy, raising CO_2 to 510 ppm actually outweighs the multi-model median damages caused by a 20% precipitation reduction in all climate regions. All crops show similar behavior, but note that model uncertainties for wheat are substantially higher than those for other crops. (Compare Figure 6a for soy and Supplemental Figure S7 for wheat).

We show some additional cases in Supplemental Material. As noted previously, the A1 adaptation simulations involve significantly moderated temperature impacts relative to the A0 simulations shown here (Supplemental Figure S14). Supplemental Figures S15 and S16 show the response in the nitrogen dimension and an irrigation water demand response example.

5 Discussion and Conclusions

The GGCMI Phase ~~II~~ 2 experiment provides a database designed to allow detailed study of crop yields in process-based models under climate change. While previous crop model intercomparison projects in the climate change context have focused on simulations along realistic projected climate scenarios ~~(e.g Rosenzweig et al., 2014)~~(e.g. Rosenzweig et al., 2014), the use of systematic input parameter variations in GGCMI Phase ~~II (~~2, with up to 756 scenarios~~)~~ , allows not only comparing yield sensitivities to changing climate and management inputs but also evaluating the complex interactions between important driving factors: CO_2, temperature, water supply, and applied ~~Nitrogen~~nitrogen. The global extent of the experiment also allows identifying geographic shifts in high potential yield locations. With 12 participating models and 31 simulation years per scenario, the complete database constitutes over 150,000 years of gridded global yield simulation output for each crop.

Preliminary results shown here highlight some of the insights facilitated by the simulation exercise and lend confidence in the models. In validation tests of ~~simulations of the historical scenario~~historical simulations, year-over-year correlations in modeled and actual country-level yields are similar to those of GGCMI Phase ~~I. In simulations on~~ 1. In simulations of scenarios with perturbed climate and management factors, models ~~generally~~ broadly agree on changes ~~driven~~ outside the high latitudes, with the magnitude of changes at nearly all perturbation levels exceeding the inter-model spread. ~~(~~At high latitudes, differences ~~are often driven by differences in model treatment of crop response to~~ between models may result from differences in their assumed yields in current cold conditions. ~~)~~In simulations with multiple perturbations, interactions between major yield drivers (e.g. temperature and precipitation in Figure 5, or precipitation and CO_2 in Figure 6) generally follow expectations and produce physically reasonable responses in crop yields.

Users should however be aware of some limitations of the GGCMI Phase 2 experiment that affect its potential applications. First, absolute ~~model~~ yield values in the ~~historical scenario~~baseline scenario, driven by 1981–2010 historical climate, will generally not match observed yields over this time period. In order to match current yields, process-based models must ~~generally~~ be re-tuned to account for the constant evolution of crop cultivar genetics and management practice (e.g. Jones et al., 2017). ~~The historical scenario also includes no trend in CO_2, and scenarios assume unrealistic globally uniform nitrogen application levels (Elliott et al., 2015). GGCMI Phase II~~ GGCMI Phase 2 is intended as a study of ~~model-projected changes under broad climate change, which may not be as sensitive~~ model responses to changes in climatic conditions, which are assumed insensitive to the adjustments needed to reproduce present-day yields. ~~Note however that the models used in this exercise cannot simulate some potential climate-related yield changes: those due to factors such as pests, diseases, and weeds~~The baseline scenario also includes no trend in CO_2, and no individual case involves realistic country-specific nitrogen application levels (Elliott et al., 2015).

The second major caveat is that no individual GGCMI Phase ~~II~~ 2 simulation is itself a realistic future yield projection. The uniform applied offsets in temperature and precipitation sample over potential changes, ~~but projections of climate change involve~~ and do not individually capture the spatially heterogeneous warming and precipitation changes expected in realistic climate projections. GGCMI Phase ~~II~~ 2 simulation results can be used for impacts projection, but only with the construction of an emulator of crop yield response to climatological changes, ~~that is then driven by a realistic climate projection. However, note that the~~ which can then be driven by arbitrary climate scenarios. Such emulators are shown to accurately reproduce crop model output under realistic climate projections, even though the GGCMI Phase 2 experiment does not sample over potential changes in the higher-order moments in ~~the distributions in~~ temperature and precipitation ~~.~~ distributions (Franke et al., 2020). Note that some factors that may affect future climate-driven yield impacts cannot be captured by the GGCMI Phase 2 models in any usage, since models do not include representations of pests, diseases, and weeds. Off-line crop model simulations (i.e. with prescribed rather than dynamically simulated atmospheric conditions) can also not capture any feedbacks on the climate from land use, such as irrigation impacts on humidity (e.g. Decker et al., 2017).

We expect that the GGCMI Phase ~~II~~ 2 simulations will yield multiple insights in future studies. Potential applications include, as mentioned, the construction of emulators and yield response ~~surface, as well as studies of issues such as~~ surfaces that can be used for both model diagnosis and impacts assessment. Specific studies could include analyzing the drivers of temperature-related yield losses (which may be due to both direct thermal effects or to shortening growing seasons); the benefits of adaptation~~, interactions between the~~ ; interactions between CO_2 and water or other CTWN factors affecting yield~~,~~ ; changes in nitrogen use efficiency~~,~~ ; geographic shifts in regional production~~, investigation of core sensitivities to CTWN/A by region and farm system, identification of~~ ; regional differences in yield sensitivities to CTWN-A factors. Emulators based on the dataset can be used to identify hotspots of crop system vulnerability, and to conduct rapid assessment of new climate projections~~, and many others~~. In general, the development of multi-model ensembles involving systematic ~~parameters~~ parameter sweeps has large promise both for increasing understanding of potential future crop responses and for improving process-based crop models.

6

5.1 ~~Data Access~~

~~Simulation yield output datasets can be found at the DOIs located in table 4. Data are published in crop- and GGCM-specific packages, in order to break down the overall data amount into manageable packages (<50GB per archive).~~

~~DOI's for model data outputs. All model output data can be found at https://doi.org/10.5281/zenodo/XX. Where XX is the value listed in the table.~~ ~~**ModelMaizeSoybeanRiceWinter wheatSpring wheatAPSIM-UGOE**2582531258253525825332582537258253~~

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-SynG-610028 Imbalance-P. P.F. and K.W. were supported by the Newton Fund 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. 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).

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., 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.

Code and data availability. The simulation outputs of the mandatory 7 output variables (Table 2) are available on zenodo.org. See Table 4 for data DOIs. Data are published in crop- and GGCM-specific packages, in order to break down the overall data amount into manageable packages (<50GB per archive). All other simulation output variables are available upon request to the corresponding author.

Table 4. DOI's for model data outputs. All model output data can be found at https://doi.org/10.5281/zenodo/XX. Where XX is the value listed in the table.

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

The scripts for 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 athttps://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 whereas all other growing season data (maize, rice, soybean) are the same as in Phase 1 (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

References

Allen, R., Pereira, L., Dirk, R., and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements, FAO Irrigation and drainage paper 56, 1998.

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., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, 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., Stratonovitch, 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.

Batjes, N.: ISRIC-WISE global data set of derived soil properties on a 0.5 by 0.5 degree grid (ver. 3.0), 2005.

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.

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.

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.

Decker, M., Ma, S., and Pitman, A.: Local land–atmosphere feedbacks limit irrigation demand, Environmental Research Letters, 12, 054 003, https://doi.org/10.1088/1748-9326/aa65a6, 2017.

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., Ouberdous, 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., Deryng, D., Müller, C., Frieler, K., Konzmann, M., Gerten, D., Glotter, M., Floerke, M., Wada, Y., Best, N., Eisner, S., Fekete, B., Folberth, C., Foster, I., Gosling, S., Haddeland, I., Khabarov, N., Ludwig, F., Masaki, Y., and Wisser, D.: Constraints and potentials of future irrigation water availability on agricultural production under climate change, Proceedings of the National Academy of Sciences of the United States of America, 111, 3239–3244, https://doi.org/10.1073/pnas.1222474110, 2014a.

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, 2014b.

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: FAOSTAT Database, http://www.fao.org/faostat/en/home, 2018.

FAO/IIASA: Global Agro-ecological Zones and FAO-GAEZ Data Portal(GAEZ v3.0), http://www.gaez.iiasa.ac.at/, 2011.

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H., Verelst, L., and Wiberg, D.: Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008), http://webarchive.iiasa.ac.at/Research/LUC/, 2008.

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., Izaurralde, R. C., Jones, C. D., Khabarov, N., Liu, W., Reddy, A., 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.

Folberth, C., Elliott, J., Müller, C., Balkovič, J., Chryssanthacopoulos, J., Izaurralde, R. C., Jones, C. D., Khabarov, N., Liu, W., Reddy, A., 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.: Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble, PLOS ONE, 14, 1–36, https://doi.org/10.1371/journal.pone.0221862, https://doi.org/10.1371/journal.pone.0221862, 2019.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., and Snyder, P. K.: Global Consequences of Land Use, Science, 309, 570–574, https://doi.org/10.1126/science.1111772, 2005.

Franke, J., Müller, C., Elliott, J., Ruane, A., Jägermeyr, J., Snyder, A., Dury, M., Falloon, P., Folberth, C., François, L., Hank, T., Izaurralde, R., Jacquemin, I., Jones, C., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T., Reddy, A., Williams, K., Wang, Z., Zabel, F., and Moyer, E.: The GGCMI phase II emulators: global gridded crop model responses to changes in CO_2, temperature, water, and nitrogen (version 1.0), Geoscientific Model Development, https://doi.org/10.5194/gmd-2019-365, in open review, 2020.

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, 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., Stratonovitch, 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.agsy.2017.08.004, 2018.

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.

Hansen, J., Sato, M., and Ruedy, R.: Radiative forcing and climate response, Journal of Geophysical Research: Atmospheres, 102, 6831–6864, https://doi.org/10.1029/96JD03436, 1997.

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.

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.

Haugen, M., Stein, M., Moyer, E., and Sriver, R.: Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression, Journal of Climate, 31, 8573–8588, https://doi.org/10.1175/JCLI-D-17-0782.1, 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.: APSIM – 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, https://doi.org/10.1175/2009JAMC2225.1, 2010.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanenkov, V., Oettli, 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.

Izaurralde, R., Williams, J., Mcgill, W., Rosenberg, N., and Quiroga Jakas, M.: Simulating soil C dynamics with EPIC: Model description and testing against long-term data, Ecological Modelling, 192, 362–384, https://doi.org/10.1016/j.ecolmodel.2005.07.010, 2006.

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.

Jägermeyr, J., Pastor, A., Biemans, H., and Gerten, D.: Reconciling irrigated food production with environmental flows for Sustainable Development Goals implementation, Nature Communications, 8, 15 900, https://doi.org/10.1038/ncomms15900, http://www.nature.com/articles/ncomms15900, 2017.

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, 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, 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, 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/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., 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, 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 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, 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.

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.

McGrath, J. M. and Lobell, D. B.: Regional disparities in the CO2fertilization effect and implications for crop yields, Environmental Research Letters, 8, 014 054, https://doi.org/10.1088/1748-9326/8/1/014054, https://doi.org/10.1088%2F1748-9326%2F8%2F1%2F014054, 2013.

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.

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., Izaurralde, 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.

Nelson, G. C., Mensbrugghe, D., Ahammad, H., Blanc, E., Calvin, K., Hasegawa, T., Havlik, P., Heyhoe, E., Kyle, P., Lotze-Campen, H., Lampe, M., Mason d'Croz, D., Meijl, H., Müller, C., Reilly, J., Robertson, R., Sands, R. D., Schmitz, C., Tabeau, A., Takahashi, K., Valin, H., and Willenbockel, D.: Agriculture and climate change in global scenarios: why don't the models agree, Agricultural Economics, 45, 85–101, https://doi.org/doi.org/10.1111/agec.12091, 2014.

Olin, S., Schurgers, 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_2 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., Stratonovitch, 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., Izaurralde, 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, 12 608, 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.

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 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 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 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, 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., Bindi, M., Carter, T., Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, S., 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/10.1088/1748-9326/10/8/085010, https://doi.org/10.1088%2F1748-9326%2F10%2F8%2F085010, 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 CO2, Agricultural and Forest Meteorology, 112, 217 – 231, 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., 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.