Review Response: The Global Gridded Crop Model Intercomparison: Data and modeling protocols for Phase 1 (v1.0)

J. Elliott[1], C. Müller[2], D. Deryng[3], J. Chryssanthacopoulos[4], K. J. Boote[5], M. Büchner[2], I. Foster[1], M. Glotter[6], J. Heinke[7,2], T. Iizumi[8], R. C. Izaurralde[9], N. D. Mueller[10], D. K. Ray[11], C. Rosenzweig[12], A. C. Ruane[12], and J. Sheffield[13]

General Comments:

The authors discuss the protocols established for the Agricultural Model Intercomparison and Improvement Project, specifically for the Global Gridded Crop Model Intercomparison.

The paper addresses the types of experiments to be conducted for each participating modeling group, with a description the atmospheric forcing and other input data for model harmonization, and an overview of the methods for evaluating performance.

The paper is well written and addresses a growing need for a comparison among the increasing number of agricultural models. I believe the methodology is sound, with well established guidelines, but the manuscript would benefit from a few clarifications.

One example is with the setup of the default model configuration – does that include the atmospheric forcing from the protocols or the standard forcing data normally used for each model? The paper alludes to using the protocol forcing data, but doesn't explicitly state how it should be configured.

- ***That is a good point and we were unaware that this could be interpreted ambiguously. The "default" setting refers to assumptions on crop management only. Even though there may be standard atmospheric forcing used by individual models, we don't intend to compare model simulations across these. We will make clear throughout the manuscript that default refers to management options, not to atmospheric forcing.***

I'm not sure I see the benefit of growing crops everywhere. While this may provide some useful incite to possible future land use scenarios and yield expectations, there is no means of validating any of the models and therefore comparing the productivity between models isn't very useful. The authors mention data won't be considered in regions where crop growing season is considered unreasonable; therefore it might make more sense to only consider where crops are currently grown.

- ***This is also a valid point and our approach is both habit and opportunity driven. The standard in the ISI-MIP fast track was to simulate crops everywhere and models that have participated in that have been set up that way, so it seemed reasonable to just stick to that setting. However, there may also be some analyses possible for the historic period that require crop yield simulations in places where these crops are currently not grown. For examples see the use of gridded crop model simulations in agro-economic models (Nelson et al. 2014, ISI-MIP PNAS special feature + special issues in Agricultural Economics (2014, volume 1). As we intend to facilitate a broad spectrum of analyses we don't want to constrain options by being too restrictive on current cropping patterns and experience from the fast-track showed that this was not a major challenge to modeling groups.***

Specific Comments:

1. P. 4388, L. 6-8: Do the three (or more) models used for intercomparison need to be comparable models? I think a DGVM and an empirical model will respond differently, especially to the harmonized forcing. Comparing the simulated yield might not reveal any useful information if the processes that go into determining that yield are represented completely differently. This brings up another issue – how are the site models run, globally or will site based output be aggregated to global levels?

- ***The models are run on a global grid with 0.5 degree resolution (longitude, latitude), see also section 3.4 on input data. We will make this clearer in the revised version of the manuscript. With respect to the comparability of model types, we do not constrain analysis to a minimum of 3 models per type. In fact, exploring whether model types are actually responsible for differences in simulations is part of the intended analysis and insights on these general differences will be generated from simulations for the priority 1 crops. If we find substantial differences between model types, this will inform the analysis and interpretation of model intercomparison of priority-2 crops.***

2. P. 4388, L. 16: I can understand running the model without nitrogen stress to compare with models that don't consider nitrogen, however, one concern is in some models, the carbon and nitrogen are coupled and removing the nitrogen stress can cause a decoupling of the carbon-nitrogen system, which might lead to less than desirable model behavior.

- ***All crop models should be able to simulate high-input systems (with no effective nitrogen stress) without jeopardizing model stability or functionality. Indeed, carbon and/or nitrogen cycle dynamics could be distorted by the assumption of superfluous nitrogen supply, but these are not analyzed here.***

3. P. 4389 L. 24: The word "minimum" used here and later on P. 4390, L. 6 should be replaced with "standard" since it seems that there are exceptions to the required simulations depending on model capabilities.

- ***Yes there are exceptions depending on model capabilities. We don't want to allow for exceptions other than model capabilities such as lack of resources to avoid dilution of the simulation set, while we also don't want to preclude models from an intercomparison on wheat simulations only because the model is not capable of also simulating soybean. For all models that have the capability to simulate these crops is thus indeed a non-negotiable minimum.***
  The "minimum" on page 4390, L. 6 is referring to our expectations and we will replace it with "at least" to avoid confusion.

4. Sect. 3.1: This section is not very clear. Are all the datasets daily or are some monthly? What about models that require a higher temporal resolution? How should models that require long spinup periods begin the simulation – cycle through the generic pre-industrial atmospheric forcing (for hundreds of years) before using the Princeton data or can an initial conditions file be used from a previous simulation?

- ***All weather inputs come in daily resolution and we have 1 WFDEI available in 3-hourly resolution (and working on a 3-hourly version of AgMERRA) for the (few) models that require sub-daily resolution. Spinup procedures are not strictly harmonized and will be handled by the modelers based on their modeling standard and experience. We discourage model initialization without spin-up as mismatches between initialization and driving data can lead to unwanted model behavior. We will clarify these points in the manuscript.***

What period are you using for the analysis – just the period that all datasets cover, or the entire period for each individual dataset?

- ***That actually depends on the analysis. Certainly the period with complete overlap between driving data will be the main focus in most analyses, but individual analyses will e.g. look at historic extreme events and will thus also analyze all data sets that cover the respective years even if these are not included in all datasets.***

5. P. 4393, L. 2: Using maturity dates to harmonize harvest is tricky since some models use a GDD based approach to determine maturity (and growth phases of crops). Depending on atmospheric inputs of a given year, the maturity dates could differ greatly between the model and the dataset. Do you have a suggested approach for those models?

- ***That is true. We suggest that modelers compute the required GDD per grid cell and crop for a single weather dataset and use these variety parameters in all simulations. We will not be able to fully harmonize growing seasons across models as there are fundamental differences between models that cannot be harmonized without greatly interfering with the model's functioning. Models are requested to report planting and harvest dates, though, so that we're able to assess how well growing seasons have been aligned and to consider this in the analyses. We will make this clearer in the manuscript.***

6. P. 4393, L. 8-9: It would be nice to have a brief description of the rule-based approach used to estimate planting and harvest dates when data isn't available.

- ***Good point. These rules are the standard rules as implemented in LPJmL and as described by Waha et al. 2012. We will include this reference.***

7. Sec. 3.2.1: Would it be possible to put a flag in the dataset to indicate which data source is being used for the planting date for each crop? It might give a confidence or quality level for the data.

- ***Yes, that is possible. We will provide that data as well.***

8. P. 4397, L. 18-23: What is the reasoning for applying fertilizer in regions that are not currently applying fertilizer? Even if it's for currently uncultivated lands, the way it is described, that methodology is counter to the current fertilization practices.

- ***For grid cells where a given crop is actually grown, the goal of the harmonized fertilizer product is to produce a dataset which reflects, as best as possible, the actual average fertilizer applied. In regions where a given crop is not currently grown, the goal is to produce a plausible best guess of the fertilizer level that would be likely to be used if the crop was grown there. We consider that the best determinant for how a specific crop is likely to be grown in one location where its not currently grown is approximated by simply looking at how its grown on average in the country/region. We do a similar thing for planting dates and growing season length. For countries where the crop in question isn't grown at all so that no information is available on average fertilizer use, we consider that countries with similar economic profiles are most likely to have similar fertilizer availability and practices. We could elaborate on this method by considering soil properties and similar factors that affect the need for and availability of nutrients, but in this project phase we have made no attempt at harmonizing on these soil characteristics.***

9. Table 2: What is the "# models" column – is that the expected number of models that will be contributing (does that include different model versions)?

- ***Yes indeed. That's the number of models expected to contribute and it includes different groups running the same model or different versions of one model. We will clarify that in the table's caption.***

10. Table 8: Both Planting Window and Automatic Planting are listed for the harmonized runs, but the dataset includes just one plant date – how should this be used, perhaps clarify in Section 3.2.2?

- ***We have removed this for clarification.***

11. Table 8: The irrigation protocol isn't mentioned in the paper (assuming each model uses its own), but in Table 8 an automatic irrigation protocol is included. The authors should include a paragraph explaining how this should be implemented in the harmonized runs.

- ***Yes indeed. This is a recommendation on how to implement irrigation rules if similar parameters are used by the models to trigger irrigation events. We will make clear that there is this recommendation in Table 8.***

12. Table 10: My understanding is that in the WFDEI dataset, pr does not contain snow, it must be added to prsn.

- ***Yes, indeed. Will be corrected.***

13. Fig. 7: Will the authors make three figures, one for each run – default, fullharm, and harmnon, or have a means of knowing which run was considered "best".

- ***This is an exemplary figure of how the evalution metrics could look like. How the evaluation will best be presented in the paper presenting the evaluation results will be determined there.***

Technical:

1. P. 4393, L. 9: should be Waha et al., 2012.
2.

- ***Yes. Will be corrected.***

J. Ramirez-Villegas:
General comments:
This paper presents a project's approach to global gridded simulations for the period 1948-2012. The paper should be a useful reference for both crop modellers involved in the project and more broadly also for other scientists that aim at using the project's public outputs for their analyses. The methods and data sources presented in the paper can also be of use to other researchers conducting regional or global-scale crop simulations. The paper provides a great deal of detail on many of the assumptions that will go into the project's simulations, including clear descriptions of weather and crop data. The GGCMI project is mainly an improvement over the work presented in the so-called 'fast track' (mainly Rosenzweig et al. 2014).
My main concern is that the authors do not demonstrate the methodology, or even parts of it. The paper is currently limited to showing the input data. This is fine, but maybe not enough for a scientific paper. For instance, one can think of some evaluation exercise of the Rosenzweig et al. (2014) model output over the historical period, using either the Iizumi et al. (2014) or the Ray et al. (2012) datasets, or both. This can provide an idea of whether there is scope for improvement in model skill through using better model inputs or scope for uncertainty reduction by 'harmonising' inputs. Taking advantage of the same simulations, authors can also show the type of extreme-event analysis that would be done. This can help the authors in framing / contextualising a bit better their objectives, and would improve substantially the paper. I suggest some revisions be made mainly targeted at removing ambiguities and better contextualising phase 1 within the project and the project's objectives more broadly in the context of climate change impacts research.

- ***The GGCMI is not necessarily an improvement but a follow up exercise to the fast-track which basically only reported on model differences. The objective of this paper presented here is not to describe the methodology of the analyses conducted in GGCMI and with GGCMI data, but to provide a clear description of the modeling protocol and the model input data provided by GGCMI. We describe data sets that will be used for model evaluation and show examples of how this evaluation could look like, but the intention of this paper is not to be a comprehensive methods section for the evaluation publication that is to follow in one or several following papers. Further, its generally not possible to use the fast-track outputs for the types of analyses considered here, because the "historical period" in the fast-track is just from climate model output rather than observation or even reanalysis-based weather. For this reason the results of the fast-track cannot be directly compared to observation-based yield estimates like Iizumi et al (2014) or Ray et al (2012). Indeed this is a significant motivation for the design of the GGCMI.***

Specific comments:
1. Relevance / context of the project. GGCMI phase 1 will conduct global simulations of as many crops as possible for a historical period with four main objectives. Authors could expand a bit on the three-year GGCMI project so that the reader gets a clearer idea of how next phases will build upon phase 1. It would also be useful to see at least a brief discussion (in Sect. 6) of how this project overlaps / feedbacks from / contributes to regional assessments that are currently being carried out / funded by AgMIP itself or by other programs (e.g. CCAFS). Moreover, the context of these analyses (i.e. global gridded simulations) within the impacts research literature should also be stated (also see point 2 below).

- ***This is certainly a great suggestion. We have expanded the discussion on phases 2 and 3 of the project in order to provide greater context for this first project phase, and clarified to a greater extent how each phase will build off the ones before. We have also expanded discussions in section 6 to clarify how the outcomes of GGCMI are expected to facilitate other projects within and beyond AgMIP and ISI-MIP, including global and regional agro-economic and biophysical climate impact assessments.***

2. Relevance / context of project objectives. It is not entirely clear, why are some of these four objectives being researched. While items (2) and (4) are clear overarching needs and/or knowledge gaps, the hypothesis and/or context behind item 1 should be stated more explicitly. More specifically, what new knowledge is expected to be generated by running models with harmonised and non-harmonised inputs? For item 3 (uncertainties) it is not clear which uncertainties or why do the authors choose to quantify these? is there evidence suggesting they may be a major source of uncertainty in yield hindcasts? On the input weather one can also think of bias correction of climate model meteorology? why are these not being researched (from a climate change perspective they may be at least as relevant)?

- ***The motivation for item 1 includes exploring how important varying assumptions on growing seasons and fertilizer inputs (or inclusion of nitrogen dynamics) actually are for simulated dynamics. Historic simulations allow for assessing how well observed variability can be reproduced by the models and how strongly this depends on assumptions on management. However item 1 also includes comparisons of some more fundamental model choices, such as the method uses to calculate evapotranspiration within the models. In phase 1, we are performing a detailed intercomparison of different ET methodologies using the fact that some participating models (pDSSAT, pAPSIM, and the EPIC-based models) have the ability to simulate multiple ET methods with all other elements held fixed. This was mentioned only briefly in the initial submission, but that oversight has now been corrected and this example of***

deep model intercomparison has now been highlighted to clarify our motivations.
Item 3: the uncertainties are to be derived from the differences between models and scenarios (weather datasets, management assumptions) also in order to facilitate a targeted attempt to improve model skills. The point is not to understand the uncertainty in yield hindcasts but to assess model skills from their ability to simulate historic yield dynamics and spatial patterns. The uncertainty in bias correction is certainly also an important one but we have put the focus on the different weather data sets available. We include, however, 2 raw reanalysis products that can shed some light on the general importance of bias correction, even though not on different methods of doing so.

3. L20-25 P4388: having in mind the four objectives stated at the beginning of Sect. 2 it does seem that running crop models where crops are not currently grown is unnecessary. Particularly for climate variability (obj. 4) and model evaluation (obj. 2) assessments. Maybe authors have a purpose for this (e.g. for further comparison to any future simulations that will be done in a follow up phase). However, as of now, why not just use some prescribed "crop mask" per crop and so in this way do not waste computational resources and facilitate further analyses? This is particularly important for northern hemisphere cereals such as wheat and barley whose climate requirements are unlikely to exist in large areas of the tropics. Vice versa for tropical crops not adapted to cold (e.g. cassava). The niches of the crops need to be maintained somehow. This brings confusion to the reader: for instance, in Fig. 4 (right) of this paper one can already see wheat in the Sahel.

- *See also response to Beth above (point 2). We note that figure 4 (right) is produced not from simulation output but instead from national and sub-national observations compiled by Ray et al (2012). Additionally, the MIRCA dataset of crop covers that is used throughout the project does indicate that there is a small but nonzero amount of wheat grown in this region (see Figure 1 below).*

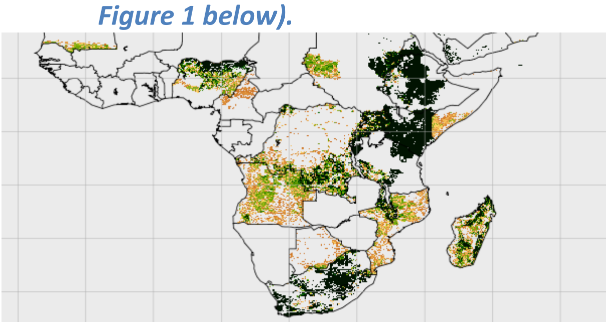


Figure 1: MIRCA land-cover dataset for rainfed wheat area in sub-Saharan Africa. The global M3-crops dataset (Monfreda et al, 2008) shows a similar result).

4. L1-10 P4389: crop duration is a key output for understanding differences across models, particularly when these are driven by mean temperatures. All annual crop models should be capable of providing this as an output. In addition, perhaps authors should somehow indicate how many models (or by percentage) can provide each output.

- *Indeed, crop modelers are asked to report planting and harvest dates, which allows for deriving crop duration (see Table 4). We can provide information on some models with respect to intended outputs, but those models that have merely indicated their interest have also not provided much information on what variables they will actually report. This information will clearly be reported in publications using the datasets provided by the crop models.*

Technical corrections:
1. L5, P4386: unless described briefly (i.e. what it is and how is it different to GGCMI) a reference to AgGRID may confuse the readership.

- ***OK, will briefly expand the description of AgGRID (or possibly scratch it)***

2. L21, P4386: consider using regional-scale process-based models. Hybrid may be too ambiguous.

- ***These models are not regional-scale models as they will be run at the global scale. Also, even though this may be true for some, not all "hybrid" models are developed for specific regions (e.g. Pegasus). The classification is certainly ambiguous, and its usefulness will have to be proven. Here we just want to highlight that we have field-scale models, land-surface/DGVM type models and other global gridded crop models, that we subsume under "hybrid" as they typically have a larger share of empirical relationships than field-scale or DGVM type models.***

3. L22, P4386: ditto above, why not just use 'statistical models', instead of 'purely empirical'?

- ***Done, thanks.***

4. L27, P4386: 'modelling groups', rather than 'modelers'

- ***Correct.***

5. L6, P4387: "such as" brings about some unnecessary ambiguity. Be specific. List clearly which uncertainty sources are being quantified.

- Done, thanks.

6. L10 P4387: productivity, not production

- Yes. Will be corrected.

7. L19-20 P4387: one would expect a relationship between the two measures (importance to food security / economies / livestock feed and number of models, or likelihood a model exists). It is likely that each criterion would yield the same list separately, hence it seems redundant to use both (with FS and/or economic importance being the independent variable). Besides, it seems reasonable to think that, as long as >=3 models simulate a particular crop (to allow for inter-comparison), the existence of many models should exert little impact on establishing the scientific problem / priorities. Also, the brackets on "(primarily global)" seem unnecessary.

- While there is certainly expected to be a correlation between the most modeled and most "important" crops, there are certainly circumstances where this is not the case. Many crops that are very important in economic terms (such as various cash crops, including coffee and tomatoes) or essential for nutrition in important regions (as e.g. sorghum, teff) are not modeled as frequently as some other crops.

8. Table 2: # models for priority 1 states 15-20 models. How can a crop achieve 20 individual model simulations when Table 1 lists 18 crop models?

- ***GGCMI is constantly growing and accepting new members and participants, so it's somewhat difficult to say precisely how many models will contribute in any given phase. At least 2 new models have joined the group since initial submission with the intention of contributing results in time to participate in one or more paper for phase 1, so once these are added to the table the 15-20 estimate is more logical.***

9. L18 P4388: "For the purposes of various analyses". Which analyses? if described in this paper please ref. the section. If not described in this paper then please do so, or state briefly what is meant by "various".

- ***OK, will do that. We have generally tried to make it clear that this modeling protocol lays the basis for many analyses, several of which have been scoped but not yet strictly planned in detail, for which we will try to provide suitable data.***

10. L16 P4389: or maybe also to be able to interpret the differences in simulated yields?

- ***Certainly a good example of a future analysis not anticipated in advance would be the proposal and evaluation of a hypothesis of what is driving yield differences that has not yet been considered.***

11 L18-20 P4391: This is unclear. While it makes sense to think of a growing season for comparability across models, observational datasets are generally based on the reporting standard of FAO, which uses whatever the countries report. In this scheme, yields reported in one year correspond to crops harvested in that year. It is not "artificial", as authors state. Authors are advised to cross-check their statement against the FAO reporting standard.

- ***According to the FAO glossary (http://faostat.fao.org/site/375/default.aspx), faostat yield estimates are usually produced by collecting production and area data and taking the ratio. For crop production, the definition in the glossary says the following:***

  > ***Crop production data refer to the actual harvested production from the field or orchard and gardens ... When the production data available refers to a production period falling into two successive calendar years and it is not possible to allocate the relative production to each of them, it is usual to refer production data to that year into which the bulk of the production falls. The procedure implemented by FAO is to assign the production to a given calendar year based on when that production is reported. In some countries this date can actually come significantly after the date of harvest.***
- ***Many crop models use a similar definition but this runs into problems when you're trying to compare among models or indeed when trying to compare to FAO. For example, in areas where harvest occurs near the new year, it may fall in some years in December and in other years in January. This often leads to calendar years with twice the normal production and other years with none. Clearly in this case assigning production strictly to the calendar year in which is falls is not the best option. Furthermore, models typically don't say much about when harvest of crops actually occurs, but instead only when the crops are matured. This is further complicated by the fact that FAO assigns production to a calendar year based not on harvest but instead on when the production data is reported to FAO, which as they note can "come significantly after the date of harvest". There is thus no consistent way to reproduce the FAO definition within a model protocol. The approach we have chosen comes as close as possible to being an unambiguous request to the model groups and leaves the difficult step of re-aligning outputs to match FAO to be done as part of the output processing pipeline, where different methods can be implemented and evaluated for relative performance.***

12. It does seem a bit strange that the paper first describes simulation outputs and only after that describes the inputs.

- ***The goal of the paper is to describe output formats and protocols, not simulation outputs themselves.***

13. L25-27 P4392: this statement is inconsistent with (actually contradicts) the purpose of the comparison of input meteorological datasets itself.

- ***Variable substitution is only required in very rare circumstances and for variables of secondary importance (long wave radiation may be the only example in fact, and its only used in a few models).***

14. Table 11 should clarify whether ‘standard‘ (for wheat and barley) means spring.

- ***This has been clarified.***

15. L6 P4394: sugarcane is harvested beyond 12 months in many places across the tropics

- ***Yes, but if we use the cropping calendar of MIRCA2000, sugarcane grows for exactly 365 days. For consistency and lack of better data with sufficient coverage, we stick to this. Also, cropping cycles >12 months would interfere with the annual character of agricultural systems that is embedded in many of the participating models***

16. L13 P4394: LAI will not be zero for indeterminate crops

- ***True. For those we simply describe harvest dates and make no effort to adjust for maturity.***

17. L3-12 P4394: it does seem like too many assumptions for areas in which no model evaluation can anyway be performed, and for which little scope exists for inter annual variability assessments.

- ***For various reasons described above, we want to produce a best guess for what the planting date and growing season length will be in each grid cell, even in grid-cells where a particular crop is not historically grown. We have tried to come up with a simple hierarchy for picking this best guess based on the data that is available at a given point.***

18. L1-4 P4396: unclear whether this is done for each input meteorology dataset or using which met data?

- ***Yeah. That was criticized above as well. We should make clear that it should be done for one assuming that differences in temperature are not that severe to account for many days.***

19. L21-25 P4397: why has this been done? clearly, it will affect simulations of models that account for nutrient availability and/or uptake, mainly across the developing world. If this procedure is inconsistent with observations then what is the expectation with regards to model evaluation?

- ***See previous answers and also for the other review. This is for the extrapolation to currently uncultivated land and will thus not affect model evaluation. However for various purposes we need to produce a best guess for what management practices would be in a grid cell if a given crop were grown there.***

20. Sect. 4.1. Perhaps it would be good to include some basic quality checking for the yield data (see for instance wheat in the Sahel, Fig. 4 right). In addition, FAOSTAT reported yields also have known issues.

- ***Yes that will be part of the evaluation study. Actually, strong disagreement with all models could be an indication of poor data quality in the reference data sets (although of course there a many other possible reasons for disagreement).***

21. L17 P4399: "various analyses". Please specify

- ***Clarified.***

22. Sect. 4.2.2. Detrending of FAOSTAT data may imply the need to detrend yield simulations as well, if climate change driven yield trends for the period analysed are observed in the simulations.

- ***Indeed, trends are removed from both the observation and simulation sets. For consistency, the same method is used to correct both (matching linear-detrended observations with linear-detrended simulations, etc.).***

23. Sect. 4.2.3 be consistent with terminology: validation vs. evaluation. Validation suggests universality (not this case), hence it seems best to use the term evaluation.

- ***Agreed, thanks.***

24. L6-8 P4401: It is unclear how this will be achieved only with yield simulations and observations. You need an entire series of prognostic variables and measurements in order to conduct such an assessment. It also seems unlikely that regional-scale evaluation of yield simulations can drive model improvement. Far more detailed data are needed for such task.

- ***Agreed. But as a first step, we try to identify areas (crops, regions, events) where crop models performance is weak. Once these cases have been identified, we can try to find general patterns and supplement additional targeted analyses for these. The global gridded crop models are intended to work at regional scale, so an assessment should work at the scale of application and any model deficiency at the scale of application can certainly inform targeted model improvement.***

25. L14 P4401: "stakeholder", please clarify / expand.

- ***Clarified, thanks.***

The Global Gridded Crop Model Intercomparison: Data and modeling protocols for Phase 1 (v1.0)

J. Elliott[1], C. Müller[2], D. Deryng[3], J. Chryssanthacopoulos[4], K. J. Boote[5], M. Büchner[2], I. Foster[1], M. Glotter[6], J. Heinke[7,2], T. Iizumi[8], R. C. Izaurralde[9], N. D. Mueller[10], D. K. Ray[11], C. Rosenzweig[12], A. C. Ruane[12], and J. Sheffield[13]

[1]{University of Chicago & Argonne Natl. Lab Computation Institute, Chicago, Illinois}

[2]{Potsdam Institute for Climate Impact Research, Potsdam, Germany}

[3]{Tyndall Centre, University of East Anglia, Norwich, United Kingdom}

[4]{Columbia University Center for Climate Systems Research, New York, New York}

[5]{University of Florida Department of Agronomy, Gainesville, Florida}

[6]{University of Chicago Department of Geophysical Science, Chicago Illinois}

[7]{International Livestock Research Institute, Nairobi, Kenya}

[8]{National Institute for Agro-Environmental Sciences, Tsukuba, Ibaraki, Japan}

[9]{University of Maryland Dept. of Geographical Sciences, College Park, Maryland}

[10]{Harvard University Center for the Environment, Cambridge, Massachusetts}

[11]{University of Minnesota Institute for the Environment, Saint Paul, Minnesota}

[12]{NASA Goddard Institute for Space Studies, New York, New York}

[13]{Princeton University Dept. Civil & Environ. Engineering, Princeton, New Jersey}

Correspondence to: J. Elliott (jelliott@ci.uchicago.edu) and C. Müller (cmueller@pik-potsdam.de)

Abstract

We present protocols and input data for Phase 1 of the Global Gridded Crop Model Intercomparison, a project of the Agricultural Model Intercomparison and Improvement Project~~’s~~ (AgMIP~~’s~~). ~~Gridded Crop Modeling Initiative (AgGRID).~~ The project includes global simulations of yields, phenologies, and many land-surface fluxes by 12-15 modeling groups for many crops, climate forcing datasets, and scenarios over the historical period from 1948-2012.

The primary outcomes of the project include 1) a detailed comparison of the major differences and similarities among global models commonly used for large-scale climate impact assessment, 2) an evaluation of model and ensemble hindcasting skill, 3) quantification of key uncertainties from climate input data, model choice, and other sources, and 4) a multi-model analysis of the impacts to agriculture of large-scale climate extremes from the historical record.

1 Introduction

Climate change presents a significant risk for agricultural productivity in many key regions, even under relatively optimistic scenarios for near-term mitigation efforts (Rosenzweig et al., 2014). Consistent global scale evaluation of crop productivity is essential for assessing the likely impacts of climate change and identifying system vulnerabilities and potential adaptations. Over the last several years, many research groups around the world have developed Global Gridded Crop Models (GGCMs) to simulate crop productivity and climate impacts at relatively high spatial resolution over continental and global extents, with a huge diversity of methodologies and assumptions leading to a wide range of results.

In 2012 and 2013, the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013), led a global Fast-Track climate impact assessment in coordination with the Inter-Sectoral Impacts Model Intercomparison Project (ISI-MIP) (Warszawski et al., 2013) that brought together a group of GGCMs to simulate future crop productivity under various climate change and farm management scenarios (Elliott et al., 2014a;Rosenzweig et al., 2014;Piontek et al., 2014;Nelson et al., 2014). Increased application of crop growth models for global-scale analyses and the wide variation in model assumptions and projected outputs found in the Fast-Track, inspired the launch of the AgMIP GRIDded crop modeling initiative (Ag-GRID) and the Global Gridded Crop Model Intercomparison (GGCMI). We define here the simulation protocol for the first phase of the GGCMI, which is designed to, among other things, enable a comprehensive evaluation of model and ensemble skill – with respect to yield levels, variability, and large-scale extreme events – based on comparisons of simulations and observations over the last several decades.

The GGCMI Phase 1 simulation protocol includes participants that run a number of gridded crop models (listed with contacts and short descriptions in Table 1), driven with consistent inputs based on multiple weather data products (to evaluate uncertainties from weather data) and harmonized management practice data (planting date, growing season length, and fertilizer inputs). The results of these different simulation runs will then be compared to 3 distinct reference data sets derived from census and remote sensing data sources (Ray et al., 2013;Iizumi et al., 2013;FAOSTAT data, 2013). GGCMI is a protocol-based simulation experiment for gridded crop models and is open to participation by any model group that simulates crop productivity at the global scale, including models developed for field-scale application, biogeochemical dynamic global vegetation and land-surface scheme models, empirical-process-based hybrid models, and statistical ~~purely empirical~~ models.

In the modeling protocol presented here, we describe the simulation experiments and priorities, central inputs provided to modelers, required outputs to be provided by ~~modelers~~modeling groups, and data format conventions. GGCMI protocols are designed to overlap as much as possible with and contribute to the refinement of the modeling protocols of the next phase of ISI-MIP (ISI-MIP2). Modelers participating in GGCMI can directly participate in ISI-MIP2 if they so desire.

2 Simulation experiments, models, and objectives

The primary goals of Phase 1 of the GGCMI are:

1) intercomparison of models with and without harmonized inputs and assumptions, and with and without explicit nitrogen stress;
2) evaluation of model and ensemble skill over the historical period;
3) detailed characterization of important uncertainties (~~, such as~~ weather data, ~~and~~ management systems, evapotranspiration methods, and output processing techniques~~)~~) in historical crop yield ~~assessment~~analysis and the implication of these for future climate impact assessment; and

4) multi-model, multi-forcing analysis of the impacts to agriculture of large-scale extremes (primarily drought and heat events) in the historical record.

Groups are asked to simulate agricultural productivity~~ion~~ for various crops under purely rain-fed as well as fully irrigated conditions for different driving input data sets on weather and management. To avoid overtaxing of modeling groups, we define simulation priorities to facilitate central analyses with an as broad as possible group of GGCMs as well as additional analyses of more specific questions~~, such as~~ (the performance of crop models for crops beyond wheat, maize, rice, soy;~~,~~ ~~and~~ the influence of weather data uncertainty on model performance; and the impact of different evapotranspiration methodologies on model response and model skill in different regions and agro-climatic zones).~~.~~

2.1 Crops and management systems to simulate

We define a two-tiered priority structure that takes into account both the crops that are most important for questions of (primarily global) food security and economics, and the crops that are most commonly simulated in available models. The three main cereal crops (maize, wheat, and rice) alone account for about 43% of total food energy intake (FAOSTAT data, 2013). Along with soybean, which is the largest single source of oilseeds globally and an essential source of protein and animal feed, these crops have been the focus of most crop yield and climate impact modeling work, and are generally simulated by all the models participating in GGCMI. Thus, we define them as our Priority-1 crops, representing the minimum set for our analyses (Table 2). Many other crops are important staple food, feed, or energy crops in economically or climate-sensitive regions, and most contributing models within GGCMI do simulate one or more of these secondary (or Priority-2) crops. In order to consider as many crops as possible, we ask modelers to supply data on all crops that they can simulate, and consider any crop simulated by at least three models as valid for a multi-model intercomparison analysis. The participating models cover a broad range of annual crops as well as managed grassland, but provide no modeling capacities for perennial crops (Table 2).

We define three distinct types of model configurations (Table 3) for the simulations in Phase 1. First, each group is to develop their own 'default' configuration based on the management and technology assumptions and inputs they typically use for simulations in the historical period.

Each group must also prepare a 'harmonized' configuration using input data, parameters, and definitions provided by the GGCMI coordinators. Finally, each model that considers nitrogen (whether with explicit fertilizers or an empirical calibration) is also to be run in a configuration without nitrogen stress, 'harmnon', to allow for direct comparison with models that do not explicitly consider the nitrogen cycle. W~~For the purposes of various analyses, especially in the context of so-called yield gaps, w~~e define the '*hamrnon_firr*', which has zero (or near-zero) stress from both nitrogen and water, as 'potential yield' for the purpose of defining yield gaps and related analyses. ~~d.'~~

All modelers are asked to simulate all crops across the globe, irrespective of current cropping areas for purely rain-fed as well as irrigated conditions. This approach allows for addressing uncertainties in assumed distributions of cropland in post-processing analysis. The minimum spatial extent of historical simulations is current agricultural land, and we require that all crops be simulated on all agricultural lands, rather than just on the land where they are currently grown.

We assume that irrigated systems are not limited by freshwater availability and have no water losses during conveyance and application. While the latter assumption has no implications for crop growth, it helps to make reported irrigation water quantities comparable across models.

Table 4 summarizes the outputs requested from GGCMI simulations. We require that all models provide two central outputs, dry matter equivalent crop yield and (for irrigated scenarios) total irrigation water requirements. Due to the unique characteristics of different models, few other output variables are available to be contributed by all groups. Rather than limit the project only to those variables that are universally produced (crop yields and applied irrigation water), we list in Table 4 many additional optional outputs that are to be provided as possible. These optional outputs include, for example, aboveground biomass, accumulated water applied and transpired, accumulated nitrogen applied and lost through leaching, key phenological dates, and growing season climate characteristics. This approach will facilitate better analyses and interpretation of results and will allow GGCMI participants to further leverage the archives for scientific deliverables and overall project impacts.

We ask that modelers archive model versions used for the simulations and all primary outputs generated, in order to allow for reproducibility and facilitate extraction of additional or more detailed (e.g., higher temporal resolution) data that may be found to be necessary for analyses not yet planned.

As far as possible for the models, all modelers should supply yield and irrigation water amounts for at least the four main crops: wheat, maize, rice and soy (Table 2). Simulations should be conducted for default and harmonized management assumptions as well as for different weather data sets. If modeling capacities are constrained, modelers should supply at least the four priority 1 crops (Table 2) and selected weather-management combinations to allow for a comprehensive model intercomparison across a limited set of scenarios and for analyses of input and assumption uncertainties with those models that contributed (Table 5). *Priority 1* denotes the minimum simulations required for participation unless model capacities do not allow for covering the full spectrum of priority 1 simulations (e.g., because not all crops are implemented, or because a model requires special weather data inputs).

Priority 2 includes two distinct simulation tracks designed around specific science objectives and expected publications. Simulations in the “climate track” (Priority 2.1) are designed to evaluate differences among the forcing products through an agro-climatic lens, enabling assessment of the relative importance of different reanalysis products, bias correction techniques, and target datasets used for bias-correction. The “crop track” (Priority 2.2) will allow us to expand our analysis to crops that have not been studied as thoroughly as the primary four food crops or that are only important regionally or in non-food contexts (such as energy crops). ~~At minimum~~ T~~t~~his expanded set is expected to include managed grass, sugarcane, sorghum, millet, rapeseed, sugar beet, and cassava.

2.2 Conventions for simulation outputs

In order to facilitate analysis, portability, and processing of outputs, results will be collected in compressed, self-describing NetCDF v4 files with consistent and relatively simple data, meta-data, and file-naming conventions described below.

File names: Each file must contain a single output variable and be named according to the following convention (see definitions in Table 6):

[*model***]_[***climate***]_[***clim.scenario***]_[***sim.scenario***]_[***variable***]_[***crop***]_[***timestep***]_[***start-year***]_[***end-year***]**.nc4

For example:

pdssat_watch_hist_default_noirr_yield_mai_annual_1958_2001.nc4

Geographical extent: Data must be submitted for the ranges 89.75 to -89.75 degrees latitude, and -179.75 to 179.75 degrees longitude. Thus, each file will contain 360 rows and 720 columns for a total of 259,200 grid cells. All ocean grid cells must be filled with the fill value (Table 7). Modelers need not simulate Greenland, the Arctic, or Antarctica but must submit output completely filled for the entire range from latitude 89.75 to -89.75. Output data must be reported row-wise starting at 89.75 and -179.75, and ending at -89.75 and 179.75. As is standard in NetCDF files, latitude, longitude and time must be included as variables in each file explicitly defining their extent.

Date reporting convention: The analysis of inter-seasonal variability of crop yields is complicated by reporting conventions involving the assignment of reported production to calendar years. This issue is especially problematic in the southern hemisphere, where harvest sometimes occurs in a window around December 31st so that assignment to calendar years based on the harvest date gives double harvests (e.g., one in early January and the next in late December of the same calendar year) in some years and no harvest in others. The data reporting convention for GGCMI thus is not calendar year but growing season based. That is, results are to be reported as a sequence of growing seasons, irrespective of whether that growing season actually spans two calendar years or if harvests occur just before or just after December 31st. Cumulative growing season variables as e.g., actual evapo-transpiration or precipitation are to be accumulated over the growing season, again irrespective of any calendar year definitions, and are to be reported in the same sequence as the harvest events (yield, above ground biomass). The unit of the time dimension of the NetCDF v4 output file is thus "growing seasons since YYYY-01-01 00:00:00" (Table 7). The first season in the file (with value time=1) is then the first *complete* growing season of the time period provided by the input data without any assumed spin-up data, which equates to the growing season with the first planting after this date. This convention roughly corresponds to an annual reporting scheme but allows for a better separation and analysis of outputs. The artificial separation of harvest seasons into two different calendar years may, however, also be present in observational data and may complicate evaluation of model skills in these regions anyway.

3 Central input data

In order to ensure comparability of simulation results across models and to investigate the importance of uncertainties with respect to weather and management data, we supply central input data to all participating modelers. The GGCMI Phase 1 protocols include a set of assumptions, definitions, and input data products that will be used to harmonize participating models as closely as possible in the *fullharm* and *harmnon* configurations (Table 8). During project pre-planning we have established data sharing arrangements with leading agricultural data groups that will contribute global high-resolution crop-specific data on key management inputs covering sowing dates, growing season length, fertilizer application rates (including nitrogen, phosphorus, and potassium), manure use, and historical atmospheric CO_2 concentration. We will also harmonize a set of definitions and parameter choices among models, ensuring that output data is directly comparable to the greatest extent possible.

3.1 Weather data inputs

In total we will use six historical retrospective-analysis-based forcing datasets (bias-corrected at monthly time-scales against observational products such as CRU and GPCC) and two raw (non-bias-corrected) reanalysis products (Table 9). Within the cropping areas of the major crops, these weather products display some uncertainty with respect to mean and variability of weather variables such as temperature (Figure 1) and precipitation (Figure 2). We do not strictly harmonize on spin-up procedures for those models that require it, however ~~For models that require spin-up periods,~~ we provide ~~will use~~ the Princeton global forcing dataset for years after 1948, and a decade of generic pre-industrial weather that can be used for all preceding years. We ~~will~~ also consider two versions of WFDEI, with biases corrected separately using either the GPCC or CRU data as targets, for a total of nine distinct data products and about 350 years of daily data. In total, this collection provides one or more weather data inputs for every year from 1948 to 2012. All products cover the 30-year period from 1980-2009 (which will serve as our primary analysis period) except WATCH (1958-2001) and Princeton (1948-2008). Each dataset is provided at daily resolution and one product (WFDEI) is additionally provided at 3-hourly resolution for those models that require sub-daily data.

Different GGCMs can require different weather variables, which are supplied by the different forcing data sets. Models that require weather variables not included in some data products (e.g., long-wave downward radiation, Table 10) should use the equivalent variable from another data set. As weather variables are bias-corrected individually and there is consequently no consistency between the individual variables within one data set, and as all data refer to the historic period, we assume that the errors introduced by this approach are small.

3.2 Harmonized growing season definitions

We supply harmonized growing season data (planting and maturity dates) for all priority 1 crops (wheat, maize, rice, soybean, see Table 2) plus data for the priority 2 crops barley, cassava, groundnuts, millet, potatoes, pulses, rapeseed, rye, sorghum, sugarbeet, sugarcane, and sunflower. Of the priority 2 crops, we lack information for cotton, while managed grassland is assumed to grow all year round. We compile growing season data from two existing global crop calendars, MIRCA2000[1] (Portmann et al., 2010) and SAGE[2] (Sacks et al., 2010), supplementing those data by a rule-based approach as implemented in LPJmL[3] (Waha et al., 2012) ~~(Waha et al., 2013)~~ to provide as much coverage of the global land surface as possible.

3.2.1 Methodology

We use data from two global cropping calendars, MIRCA2000 (Portmann et al., 2010) and SAGE (Sacks et al., 2010) for current cropping regions (or administrative units with cropping activity). To fill areas not covered by MIRCA2000 and SAGE, we use the planting and harvest dates as computed by LPJmL (Waha et al., 2012) as implemented for the ISI-MIP Fast-Track (Müller and Robertson, 2013;Rosenzweig et al., in press). Table 11 shows the availability of crops in the crop calendar data sets and the crops used from LPJmL.

---

[1] Available for download at ftp://ftp.rz.uni-frankfurt.de/pub/uni-frankfurt/physische_geographie/hydrologie/public/data/MIRCA2000/growing_periods_listed/CELL_SPECIFIC_CROPPING_CALENDARS_30MN.TXT.gz

[2] Available for download at http://www.sage.wisc.edu/download/sacks/netCDF0.5degree.html

[3] Available for download at the ISI-MIP fast-track archive http://esg.pik-potsdam.de

MIRCA2000 data supply up to five growing periods per pixel, each with a specific area. For each pixel, we choose the growing period with the largest area. SAGE data supplies median planting and harvest dates as well as beginning and end of planting/harvest. We use the median dates. Because MIRCA2000 has monthly resolution only, assuming the first of the month for planting dates and the last of the month for harvest dates, we use SAGE data with daily resolution where available, and MIRCA2000 data only in regions where no SAGE data is available. We ignore MIRCA2000 data if growing seasons are longer than 330 days (e.g., wheat in large parts of Russia), except for sugarcane, which is recorded to grow all year round in MIRCA2000. Finally, we use LPJmL data to fill remaining areas globally with climate-driven rule-based estimates covering a large subset of priority 1 and 2 crops.

To estimate growing season length, we use harvest dates from the same data set selected for planting dates. In order to estimate the maturity date (which characterizes crop varieties) from the harvest date, we correct for crop-specific times between harvest and maturity, assuming that maturity in models refers to the development stage in which the green LAI is zero ("fully ripe"; BBCH code 89)[4]. Where no information on differences between harvest and maturity dates could be found, we assume no difference (Table 11 contains details by crop).

In regions where neither crop calendar supplies data, we use simulated phenology from LPJmL. Here, we mask planting dates as unreasonable if planting in cool regions occurs before day 90 or after day 274 in the northern hemisphere or between days 152 and 304 in the southern hemisphere. We define *cool regions* as those in which the annual mean of monthly maximum temperatures according to the WATCH data average for 1991-2000, is only 3°C above the crop-specific base temperature. In these areas, GGCMI modelers can chose any planting date or skip the simulation as results will not be evaluated. Generally, all anticipated analyses will consider current cropland areas only, for which data is generally available from crop calendars. Data filling with rule-based algorithms is only meant to harmonize assumptions among models and to enable standard all-crops-everywhere simulations.

We also mask harvest dates as unreasonable where crops in regions filled with rule-based LPJmL data do not reach maturity within a prescribed crop-specific maximum growing season length,

[4] http://en.wikipedia.org/wiki/BBCH-scale_%28cereals%29

where crops die after less than 60 days, where freezing (Tmin of WATCH data average for 1991-2000 below 0°C) occurs in the month prior to maturity, or where planting dates are unreasonable.

If the LPJmL growing season occurs in very hot seasons (defined as those for which Tmax of WATCH data average for 1991-2000 in one of the growing season months is > 38°C), we assume that the growing season of temperate cereals (barley, rye, wheat) is offset by 6, +3 or -3 months to avoid the heat. Offsets are tested in this sequence and the first that actually reduces maximum monthly temperatures to at least below 36°C is selected. Avoidance of heat is not part of the rules implemented in LPJmL (Waha et al., 2012) and may imply that corrected sowing happens not during the wettest season. Since these areas are not currently cropped (otherwise there would be crop calendar data), it seems justifiable to correct sowing dates for cooler seasons for harmonized simulation data.

SAGE calendar data are uniform within administrative units. If the SAGE data set suggests that planting in currently unused grid cells would occur in autumn but mean monthly temperatures are already below 5°C, we correct planting dates for planting of spring varieties. For this correction, we select the first month, starting in January for the northern hemisphere and in July for the southern hemisphere, in which average monthly temperatures (Tas of WATCH data average for 1991-2000) rise above 5°C.

The R processing script that we used to generate these data is available in the appendix and in the GGCMI software repository at https://github.com/RDCEP/ggcmi/.

3.2.2 Implementation instructions for growing season dates

GGCMI modelers should implement planting dates per grid cell, per crop, and per irrigation system (purely rain-fed vs. irrigated) either directly or with a given flexibility within model-specific planting windows. In regions in which the harmonized planting dates as supplied here are masked as unreasonable, crop modelers may either set planting dates to any date or simply skip simulations, whatever is easier to implement. These data will not be considered in GGCMI analyses.

Crop variety parameters (e.g., required growing degree days to reach maturity, vernalization requirements, photoperiodic sensitivity) should be adjusted as much as possible to roughly match reported maturity dates supplied here for the average of the period 1991-2000. In regions in which harvest dates are masked as unreasonable, modelers should parameterize their fastest maturing crop variety as these stand best chances to reach maturity at all.

3.3 Harmonized fertilizer inputs

We supply average annual nitrogen (N-equivalent), phosphorus (P_2O_5-equivalent), and potassium (K_2O-equivalent) application rates (kg ha^{-1} yr^{-1}) for 15 crops and all locations. We supply crop-specific fertilization rates for the Priority 1 crops (Table 1) as well as a broad set of Priority 2 crops (cassava, cotton, groundnut, millet, potato, rapeseed, sorghum, sugarbeet, sugarcane, sunflower) as well as for one perennial crop, coffee. Fertilizer data is based on published data on mineral fertilizers and manure applications (Mueller et al., 2012;Potter et al., 2010;Foley et al., 2011). These data are available for currently cropped areas and have been extrapolated in space to cover the entire land surface.

3.3.1 Methodology

We compiled and harmonized fertilizer data in a four-step procedure. First, we disaggregated manure data to crop-specific application rates. This was done by assigning a proportion of the manure nutrient production from (Potter et al., 2010) to croplands as outlined in (Foley et al., 2011). Of manure applied to croplands, crop-specific application was determined by dividing manure application in each grid cell between all crops present in the grid cell, in proportion to harvested area of each crop.

We aggregate data from the original five arcminute resolution to the GGCMI simulation grid of 0.5°x0.5°. The political units in the original mineral fertilizer dataset differ for each crop type and cover current crop-specific growing area, up to 473 units for the maize nitrogen fertilizer data (Mueller et al., 2012). Therefore we harmonized the administrative boundary units across crop and nutrient types for the interpolation procedure here. Data on manure application (Potter et al., 2010) have resolution finer than political units, as they are based off a gridded livestock dataset. Thus, the manure nutrient maps were simply aggregated to each of the 372 administrative units as an area-weighted average.

In a third step, we harmonized the reference units between organic and inorganic fertilizers (manure). Original manure data is reported in terms of atomic nitrogen (N) and phosphorus (P) and assumed to contain no potassium (Potter et al., 2010) whereas inorganic fertilizer data is reported as N, phosphate (P_2O_5) and potassium oxide (K_2O). The conversion from P manure to P_2O_5 is based on atomic masses

$$P_2O_5\text{-eq.} = P /31 * (31*2+5*16). \tag{1}$$

Nutrients from manure are generally less available to plants than mineral fertilizers. We assume 60% of applied N-manure and 75% of applied P-manure to be plant-available (Rosen and Bierman, 2005).

In the final step, we extrapolated fertilizer application rates to currently uncultivated land. The original data on mineral fertilizers (Mueller et al., 2012) cover only crop-specific harvested areas. First, we assigned the national average nutrient-specific fertilizer rate (area-weighted) to all administrative units that do not apply any mineral fertilizer or manure in the original data but are within a country actually reporting fertilizer application. Second, for all other countries that do not currently apply fertilizer to grow the specific crop, we attributed estimated nutrient-specific application rates by averaging fertilizer application rates over the corresponding income level group. We base income level groups on the World Bank's definition to classify countries by income level: economies are divided according to 2012 GNI per capita, calculated using the World Bank Atlas method (World Bank 2013). The groups are: low income, $1,035 or less; lower middle income, $1,036 - $4,085; upper middle income, $4,086 - $12,615; and high income, $12,616 or more. We averaged fertilizer application rates for all countries with fertilizer application larger than zero within the income level group and applied those rates to all countries without fertilizer data within that group.

3.3.2 Implementation instructions

All fertilizer data supplied here should be treated as mineral fertilizer; organic fertilizer (manure) has been reduced to account for limited plant-availability and combined with data on inorganic fertilizer applications.

3.4 Other data and parameter recommendations

In addition to management drivers, we harmonize on historical CO_2 levels based on the Mauna Loa Observatory time-series (Thoning et al., 1989). We also provide instructions for how to measure growing seasons, and provide guidance on parameter choices for automatic irrigation algorithms (where applicable).

3.5~~4~~ Data format conventions of input data

All input data is supplied in gridded form at 0.5° x 0.5° spatial resolution in a compressed NetCDF4 file format. Weather data is available at daily time steps and at 3-hourly values for WFDEI (which is required for some participating land-surface models). Management data is available for only one time period and are assumed to apply for all historic time periods since data is lacking on changes in management over time (all comparisons are done between detrended observation and simulation time-series, which greatly reduces, but certainly does not eliminate the effect of changes management practices and technology over time).

4 ~~Validation~~ Evaluation datasets and procedures

4.1 Historical yield data

We will use three yield data products at multiple scales to ~~validate~~ evaluate our simulation outputs~~analysis~~, Iizumi (Iizumi et al., 2014), Ray (Ray et al., 2012) ~~(Ray et al., 2013)~~, and FAOSTAT (FAOSTAT data, 2013). Iizumi (Figure 4, left) provides a hybrid of national statistics and satellite derived Normalized Difference Vegetation Index (NDVI) at a nominal resolution of 1.125 degrees, covering maize, soy, wheat, and rice, and spanning 1982-2006. Ray (Figure 4, right) covers the same four crops using national, sub-national and sub-subnational statistics, spans 1961-2008, and is provided at a nominal resolution of five arcminutes by distributing yield statistics from administrative units to grid cells evenly based on the approximate distribution of crop areas in the unit, without any proxy measures of the relative distribution of attained yields. To fill in the gaps of crops and years that are not available in these first two datasets, we will compare aggregated simulation outputs at the national level directly with statistics from FAOSTAT.

4.2 Open-source processing and evaluation pipeline

In order to ensure consistency and encourage consensus in GGCMI products, we are developing all output processing software utilities within an open software repository available at https://github.com/RDCEP/ggcmi/. Additionally, we permanently archive the intermediate and final results of each step in the output processing pipeline on the GGCMI data servers. These data will be made available along with the data supplied by GGCMI modeling groups at the time of public release. The key stages of the pipeline are described in sections 4.2.1-4.2.4.

4.2.1 Aggregation

All simulated data is first aggregated up to administrative and environmental boundaries~~ for the purpose of various planned evaluations and analyses~~, including state/province (GADM[5] level 1), country (GADM level 0)~~-~~, river basins and Food Producing Units (FPUs; river basins crossed with countries (Cai and Rosegrant, 2002)), Koeppen-Geiger climate regions (Peel et al., 2007) (example shown in Figure 5), and large-scale continental or sub-continental regions.

4.2.2 Detrending

In order to compare FAOSTAT observations with simulation results, we must remove trends from the statistics. As there are several methods to remove trend from observed data and no one method works best in all situations, we employ four distinct detrending methods: we take the linear or quadratic trends from a least-squares regression (Fig. 6, right), we take a 7 year moving mean trend, and we calculate the fraction first differences, $Y_t / Y_{t-1} - 1$, of the series and remove a linear trend (Figure 6, right). All conclusions and results are then checked for robustness against all the detrending method used.

4.2.3 Multi-metric ~~validation~~evaluation

[5] http://gadm.org/

GGCMI uses a varied approach to ~~model validat~~evaluate model outputs~~ion~~ over the evaluation period, comparing reference data and simulations using a number of metrics and methodologies. In preliminary analysis, metrics evaluated include the time-series correlation, root-means-square error, ratio of simulated and observed coefficients of variation, and the top and bottom hit-rates (number of years in the top and bottom quintile of the observation series that are reproduced in the simulated series). The metrics are formalized in the output processing pipeline in a set of multi-dimensional metric files, which are provided along with a plotting application that produces 2-dimensional cross-sections by selecting, averaging, or optimizing over any combination of dimensions (an example array is shown in Figure 7).

4.2.4 Multi-model ensembles

In a final processing step, we aim to produce multi-model ensemble versions of the output to evaluate, for example, how well the ensemble performs relative to individual models, highlighting individual model skill and deficiencies vs. model community skills and deficiencies. This step uses the multi-metrics files to produce versions of the simulated variables that aggregate all the models into various combinations. Ensembles range in complexity from simple averages (all models weighted equally) to weighted averages using one or more evaluation metric, and from all models included in the average to the inclusion of only the top-performing model. Finally, we produce evaluation multi-metric files for the ensemble combinations to easily facilitate comparison of the ensemble measures with individual models. This will be the basis for identifying central processes in models that are responsible for differences in model performance as well as general model deficiencies that require improvements in all models and in understanding. This phase will likely require additional simulations with modified models.

5 GGCMI data archive and crediting

GGCMI computing and data services are housed at the University of Chicago Research Computing Center (RCC) and the German Climate Computing Center (DKRZ). GGCMI will host an archive of all project inputs and outputs and will work continuously with research and stakeholder communities, for example through engagement processes established as part of frequent regional and global workshops hosted by AgMIP, to improve archive access and usability. During each phase of the project (i.e. before public launch of the resulting archive), all inputs and outputs generated belong to the GGCMI as a team (i.e., all GGCMI modelers) and must not be used, distributed, presented, or published in any individual or selected study without the consent of the group of contributing GGCMI modelers. During this time, presentations and publications will be led by GGCMI team members and will be coordinated through the GGCMI coordinators. The publications must acknowledge each individual contribution, including providers of not publicly available input or reference data, via co-authorship or other agreed acknowledgement.

Because GGCMI acts as the sectoral coordinator for crop modeling in phase 2 of the ISI-MIP project (ISI-MIP2), we have designed the GGCMI protocols to overlap with (planned) ISI-MIP2 simulations as closely as possible. Upon the data submission deadline as defined by ISI-MIP2, GGCMI data will automatically be transferred to ISI-MIP2, unless otherwise specified by participating modelers. At this time, GGCMI modelers become ISI-MIP2 participants and additional restrictions or specifications for data availability, as negotiated between ISI-MIP2 and GGCMI coordinators and modelers, may apply at this time.

6 Discussion

The core outcome of GGCMI is the creation and maintenance of an international community of modelers focusing on climate impacts and relationships to food security, resources, economics, land-use change, and climate feedbacks at continental and global scales. As has been amply demonstrated in processes like CMIP (Taylor et al., 2012), the Energy Modeling Forum (Weyant et al., 2006), AgMIP projects such as the wheat pilot (Asseng et al., 2013), and the ISI-MIP fast-track recently completed (Warszawski et al., 2013;Rosenzweig et al., 2014;Elliott et al., 2014a;Nelson et al., 2014), the bringing together of modelers working independently on complex dynamic phenomena to compare and synthesize outputs can generate substantive insights and innovations that are not generally possible otherwise. A key observation from the AgMIP/ISI-MIP Fast-Track and other recent model intercomparisons (Rosenzweig et al., 2014;Nelson et al., 2014;Challinor et al., 2014), and a key motivation for GGCMI, is the importance of harmonization on input data and assumptions.

Each phase of GGCMI will include planning, simulation, analysis, and publication components that will build on the inputs, science, and deliverables of the previous phase. In Phase 2, ~~analysis of CTWN sensitivity,~~ GGCMI participants will conduct a multi-dimensional sensitivity study of model response to carbon dioxide, temperature, water, and nitrogen (CTWN) organized around a set of simulations driven by perturbed versions of the historical and harmonization data products prepared in Phase 1. Results will be used both to analyze model sensitivity and to develop high-resolution multi-dimensional response surfaces that can be aggregated to arbitrary administrative or environmental boundaries and ~~used~~will be tested for suitability as efficient multi-model emulators. In Phase 3, GGCMI participants will conduct a comprehensive assessment of climate vulnerabilities, impacts, and adaptations using a new set of future climate forcings from CMIP5 and CORDEX and a detailed set of adaptation scenarios developed in the AgMIP Representative Agricultural Pathways (RAPs) framework. GGCMI also builds on other existing AgMIP projects, such as the Coordinated Climate-Crop Modeling Project (Ruane et al., 2014), and cross-cutting themes such as uncertainty and spatial scaling/aggregation.

We intend that during GGCMI's three year duration, the community will create a new standard for research on global change vulnerabilities, impacts, and potential adaptations. Data products, analyses and insights are to be published in peer-reviewed scientific journals and will thus be accessible to the scientific community. Due to the open and accessible structure of the project and its data distribution architecture, we expect important scientific outcomes and deliverables to evolve and develop during and well beyond the planned project lifetime. GGCMI leverages, and relies on, the contributions of many partners that typically lack funding for this project. However, the tremendous enthusiasm that this project has generated among participants and user communities makes us confident that GGCMI will succeed in its stated goals—and, with high likelihood, greatly surpass those goals. In addition, close partnership with the AgMIP and ISI-MIP networks, and the active participation of leaders from those groups, will help ensure that GGCMI is highly visible within and beyond the scientific community. The GGCMI team will also work with potential end-users to facilitate usage of GGCMI results downstream in economic models and global and regional integrated assessments. For this purpose we are developing several use cases for the existing fast-track archive (Nelson et al., 2014) and working with economic modeling communities such as EMF and GTAP[6] and actively seek funding for GGCMI activities and cooperation with other groups.

The standardized, protocol-based model intercomparison described here will be the basis for a clear analysis of model skills and deficiencies, identification and reduction of crop model uncertainties, and identification of future development paths to improve models and assessments. Clearly, more work than is envisioned here is needed in analyzing and improving crop modeling skills for gridded large-scale applications. Still, the first phase of GGCMI will provide a solid basis for future work by providing not only standardized inputs and reference data but also open-access data processing and analysis tools. During this first part of the project, we expect that key conditions for the next phase of analysis will take shape, by identifying the main sources of uncertainty and model-disagreement. We hope to support all large-scale crop modeling efforts with the insights and analysis tools that are produced in GGCMI, and we invite all agricultural scientists to contribute to the development and framing of the next phases of the project and protocols.

Acknowledgements

JE acknowledges financial support from the National Science Foundation under grants SBE-0951576 and GEO-1215910. CM acknowledges financial support from the KULUNDA project (01LL0905L) and the FACCE MACSUR project (031A103B) funded through the German Federal Ministry of Education and Research (BMBF). Computing and data resources provided through the University of Chicago Research Computing Center.

References

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P. J., Rotter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P. K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A. J., 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., Naresh Kumar, S., Nendel, C., O/'Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D.,

[6] https://www.gtap.agecon.purdue.edu/

White, J. W., Williams, J. R., and Wolf, J.: Uncertainty in simulating wheat yields under climate change, Nature Climate Change, 3, 827-832, 10.1038/NCLIMATE1916, 2013.

Balkovič, J., van der Velde, M., Schmid, E., Skalský, R., Khabarov, N., Obersteiner, M., Stürmer, B., and Xiong, W.: Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation, Agricultural Systems, 120, 61-75, http://dx.doi.org/10.1016/j.agsy.2013.05.008, 2013.

Bolton, D.: The Computation of Equivalent Potential Temperature, Monthly Weather Review, 108, 1046-1053, 10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2, 1980.

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Global Change Biology, 13, 679-706, 10.1111/j.1365-2486.2006.01305.x, 2007.

Cai, X. M., and Rosegrant, M. W.: Global water demand and supply projections part - 1. A modeling approach, Water International, 27, 159-169, 2002.

Challinor, A. J., Wheeler, T. R., Craufurd, P. Q., Slingo, J. M., and Grimes, D. I. F.: Design and optimisation of a large-area process-based model for annual crops, Agricultural and Forest Meteorology, 124, 99-120, 2004.

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., and Chhetri, N.: A meta-analysis of crop yield under climate change and adaptation, Nature Clim. Change, 4, 287-291, 10.1038/nclimate2153, 2014.

de Noblet-Ducoudre, N., Gervois, S., Ciais, P., Viovy, N., Brisson, N., Seguin, B., and Perrier, A.: Coupling the Soil-Vegetation-Atmosphere-Transfer Scheme ORCHIDEE to the agronomy model STICS to study the influence of croplands on the European carbon and water budgets, Agronomie, 24, 397-407, 2004.

Deryng, D., Sacks, W. J., Barford, C. C., and Ramankutty, N.: Simulating the effects of climate and agricultural management practices on global crop yield, Global Biogeochem. Cycles, 25, GB2006, 10.1029/2009GB003765, 2011.

Deryng, D., Conway, D., Ramankutty, N., Price, J., and Warren, R.: Global crop yield response to extreme heat stress under multiple climate change futures, Environmental Research Letters, 9, 034011, 2014.

Drewniak, B., Song, J., Prell, J., Kotamarthi, V. R., and Jacob, R.: Modeling agriculture in the Community Land Model, Geosci. Model Dev., 6, 495-515, 10.5194/gmd-6-495-2013, 2013.

Elliott, J., Deryng, D., Müller, C., Frieler, K., Konzmann, M., Gerten, D., Glotter, M., Flörke, M., Wada, Y., Best, N., Eisner, S., Fekete, B. M., Folberth, C., Foster, I., Gosling, S. N., Haddeland, I., Khabarov, N., Ludwig, F., Masaki, Y., Olin, S., Rosenzweig, C., Ruane, A. C., Satoh, Y., Schmid, E., Stacke, T., Tang, Q., and Wisser, D.: Constraints and potentials of future irrigation water availability on agricultural production under climate change, Proceedings of the National Academy of Sciences, 111, 3239-3244, 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, Environmental Modelling & Software, Published online ahead of print, http://dx.doi.org/10.1016/j.envsoft.2014.04.008, 2014b.

FAOSTAT data: http://faostat.fao.org/ 2013.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O/'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockstrom, J., Sheehan, J., Siebert, S., Tilman, D., and Zaks, D. P. M.: Solutions for a cultivated planet, Nature, 478, 337-342, 10.1038/nature10452, 2011.

Gueneau, A., Schlosser, C. A., Strzepek, K. M., Gao, X., and Monier, E.: CLM-AG: An Agriculture Module for the Community Land Model version 3.5, MIT Joint Program on the Science and Policy of Global Change, 2012.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanernkov, 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, early online, 10.1111/geb.12120, 2013.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanernkov, 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, 10.1111/geb.12120, 2014.

Izaurralde, R. C., Williams, J. R., McGill, W. B., Rosenberg, N. J., and Jakas, M. C. Q.: Simulating soil C dynamics with EPIC: Model description and testing against long-term data, Ecological Modelling, 192, 362-384, 2006.

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., and Ritchie, J. T.: The DSSAT cropping system model, European Journal of Agronomy, 18, 235-265, 2003.

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., Huth, N. I., Hargreaves, J. N. G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J. P., Silburn, M., Wang, E., Brown, S., Bristow, K. L., Asseng, S., Chapman, S., McCown, R. L., Freebairn, D. M., and Smith, C. J.: An overview of APSIM, a model designed for farming systems simulation, European Journal of Agronomy, 18, 267-288, http://dx.doi.org/10.1016/S1161-0301(02)00108-9, 2003.

Levis, S., Bonan, G. B., Kluzek, E., Thornton, P. E., Jones, A., Sacks, W. J., and Kucharik, C. J.: Interactive Crop Management in the Community Earth System Model (CESM1): Seasonal Influences on Land-Atmosphere Fluxes, Journal of Climate, 25, 4839-4859, 10.1175/JCLI-D-11-00446.1, 2012.

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 Syst. Dynam., 4, 385-407, 10.5194/esd-4-385-2013, 2013.

Liu, J. G., Williams, J. R., Zehnder, A. J. B., and Yang, H.: GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scale, Agricultural Systems, 94, 478-493, 10.1016/j.agsy.2006.11.019, 2007.

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., and Foley, J. A.: Closing yield gaps through nutrient and water management, Nature, 490, 254-257, 10.1038/nature11420, 2012.

Müller, C., and Robertson, R.: Projecting future crop productivity for global economic modeling, Agric. Econ., in press, 2013.

Müller, C., and Robertson, R.: Projecting future crop productivity for global economic modeling, Agric. Econ., 45, 37-50, 10.1111/agec.12088, 2014.

Nelson, G. C., Valin, H., Sands, R. D., Havlík, P., Ahammad, H., Deryng, D., Elliott, J., Fujimori, S., Hasegawa, T., Heyhoe, E., Kyle, P., Von Lampe, M., Lotze-Campen, H., Mason d'Croz, D., van Meijl, H., van der Mensbrugghe, D., Müller, C., Popp, A., Robertson, R., Robinson, S., Schmid, E., Schmitz, C., Tabeau, A., and Willenbockel, D.: Climate change effects on agriculture: Economic responses to biophysical shocks, Proceedings of the National Academy of Sciences, 111, 3274-3279, 10.1073/pnas.1222465110, 2014.

Okada, M., Iizumi, T., Hayashi, Y., and Yokozawa, M.: Modeling the multiple effects of temperature and radiation on rice quality, Environmental Research Letters, 6, 034031, 10.1088/1748-9326/6/3/034031, 2011.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633-1644, 10.5194/hess-11-1633-2007, 2007.

Piontek, F., Müller, C., Pugh, T. A. M., Clark, D. B., Deryng, D., Elliott, J., Colón González, F. d. J., Flörke, M., Folberth, C., Franssen, W., Frieler, K., Friend, A. D., Gosling, S. N., Hemming, D., Khabarov, N., Kim, H., Lomas, M. R., Masaki, Y., Mengel, M., Morse, A., Neumann, K., Nishina, K., Ostberg, S., Pavlick, R., Ruane, A. C., Schewe, J., Schmid, E., Stacke, T., Tang, Q., Tessler, Z. D., Tompkins, A. M., Warszawski, L., Wisser, D., and Schellnhuber, H. J.: Multisectoral climate impact hotspots in a warming world, Proceedings of the National Academy of Sciences, 111, 3233-3238, 10.1073/pnas.1222471110, 2014.

Portmann, F., Siebert, S., Bauer, C., and Döll, P.: Global data set of monthly growing areas of 26 irrigated crops, Institute of Physical Geography, University of Frankfurt, Frankfurt am Main, GermanyFrankfurt Hydrology Paper 06, 400, 2008.

Portmann, F. T., Siebert, S., and Döll, P.: MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global Biogeochemical Cycles, 24, Gb1011, 10.1029/2008gb003435, 2010.

Potter, P., Ramankutty, N., Bennett, E. M., and Donner, S. D.: Characterizing the Spatial Patterns of Global Fertilizer Application and Manure Production, Earth Interactions, 14, 1-22, 10.1175/2009EI288.1, 2010.

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., and Foley, J. A.: Recent patterns of crop yield growth and stagnation, Nat. Commun., 3, 7, 10.1038/ncomms2296, 2012.

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A.: Yield Trends Are Insufficient to Double Global Crop Production by 2050, Plos One, 8, 8, 10.1371/journal.pone.0066428, 2013.

Rosen, C. J., and Bierman, P. M.: Using manure and compost as nutrient sources for fruit and vegetable crops, Publication of the Department of Soil, Water, and Climate University of Minnesota, 2005.

Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburne, P., Antle, J. M., Nelson, G. C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., and Winter, J. M.: The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies, Agricultural and Forest Meteorology, 170, 166-182, 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, 10.1073/pnas.1222463110, 2014.

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 of the United States of America, in press, in press.

Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W., Romero, C. C., and DeWayne Cecil, L.: Carbon–Temperature–Water change analysis for peanut production under climate change: a prototype for the AgMIP Coordinated Climate-Crop Modeling Project (C3MP), Global Change Biology, 20, 394-407, 10.1111/gcb.12412, 2014.

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, 10.1111/j.1466-8238.2010.00551.x, 2010.

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecology and Biogeography, 10, 621-637, 2001.

Song, Y., Jain, A. K., and McIsaac, G. F.: Implementation of dynamic crop growth processes into a land surface model: evaluation of energy, water and carbon fluxes under corn and soybean rotation, Biogeosciences, 10, 8039-8066, 10.5194/bg-10-8039-2013, 2013.

Stehfest, E., Heistermann, M., Priess, J. A., Ojima, D. S., and Alcamo, J.: Simulation of global crop production with the ecosystem model DayCent, Ecological Modelling, 209, 203-219, 10.1016/j.ecolmodel.2007.06.028, 2007.

Stöckle, C. O., Donatelli, M., and Nelson, R.: CropSyst, a cropping systems simulation model, European Journal of Agronomy, 18, 289-307, http://dx.doi.org/10.1016/S1161-0301(02)00109-0, 2003.
Supit, I., Hooijer, A. A., and van Diepen, C. A.: System description of the WOFOST 6.0 crop simulation model implemented in CGMS, EUR;15956 EN, Office for Official Publications of the European Commission, Luxembourg, 1994.
Tao, F., and Zhang, Z.: Climate Change, High-Temperature Stress, Rice Productivity, and Water Use in Eastern China: A New Superensemble-Based Probabilistic Projection, Journal of Applied Meteorology and Climatology, 52, 531-551, 10.1175/JAMC-D-12-0100.1, 2012.
Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485-498, 10.1175/BAMS-D-11-00094.1, 2012.
Thoning, K. W., Tans, P. P., and Komhyr, W. D.: Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–1985, Journal of Geophysical Research: Atmospheres, 94, 8549-8565, 10.1029/JD094iD06p08549, 1989.
Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., and Martiné, J. F.: Modeling sugarcane yield with a process-based model from site to continental scale: uncertainties arising from model structure and parameter values, Geosci. Model Dev., 7, 1225-1245, 10.5194/gmd-7-1225-2014, 2014.
Van den Hoof, C., Hanert, E., and Vidale, P. L.: Simulating dynamic crop growth with an adapted land surface model - JULES-SUCROS: Model development and validation, Agricultural and Forest Meteorology, 151, 137-153, 10.1016/j.agrformet.2010.09.011, 2011.
van Diepen, C. A., Wolf, J., van Keulen, H., and Rappoldt, C.: WOFOST: a simulation model of crop production, Soil Use and Management, 5, 16-24, 10.1111/j.1475-2743.1989.tb00755.x, 1989.
Waha, K., van Bussel, L. G. J., Müller, C., and Bondeau, A.: Climate-driven simulation of global crop sowing dates, Global Ecology and Biogeography, 21, 247-259, 10.1111/j.1466-8238.2011.00678.x 2012.
Waha, K., Müller, C., Bondeau, A., Dietrich, J. P., Kurukulasuriya, P., Heinke, J., and Lotze-Campen, H.: Adaptation to climate change through the choice of cropping system and sowing date in sub-Saharan Africa, Global Environmental Change, 23, 130-143, 10.1016/j.gloenvcha.2012.11.001, 2013.
Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework, Proceedings of the National Academy of Sciences of the United States of America, early view, 10.1073/pnas.1312330110, 2013.
Weyant, J. P., de la Chesnaye, F. C., and Blanford, G. J.: Overview of EMF-21: multigas mitigation and climate policy, The Energy Journal, 27, 1-32, 2006.

Table 1: Models and groups engaged thus far for GGCMI.

Model	Lead Institution	Contact(s)	Model type and notes
pDSSAT [†|E]	U of Chicago, USA	jelliott@ci.uchicago.edu	Site-based process (Elliott et al., 2014b) (DSSAT 4.5, Jones et al., 2003)
EPIC-Boku*[|E]	Boku, Austria	erwin.schmid@boku.ac.at	Site-based process (EPIC v0810) (Balkovič et al., 2013)
GEPIC*[|E]	EAWAG, Switzerland	folberth@iiasa.ac.at	Site-based process (EPIC v0810) (Liu et al., 2007)

pAPSIM[†E]	U of Chicago, USA	jelliott@ci.uchicago.edu	Site-based process (APSIM v7.5) (Elliott et al., 2014b;Keating et al., 2003)
EPIC-IIASA*[E]	IIASA, Austria	khabarov@iiasa.ac.at	Site-based process (EPIC v0810) (Balkovič et al., 2013)
EPIC-TAMU*[E]	TAMU and UMD, USA	cizaurra@umd.edu	Site-based process (EPIC v1102) (Izaurralde et al., 2006)
CropSyst[O]	WSU, USA	stockle@wsu.edu	Site-based process (Stöckle et al., 2003)
DAYCENT[O]	Colorado State, USA	dennis.ojima@colostate.edu	Site-based process (Stehfest et al., 2007)
LPJmL[ǀ]	PIK, Germany	cmueller@pik-potsdam.de	DGVM (Bondeau et al., 2007;Müller and Robertson, 2014)
ORCHIDEE	IPSL, France	nathalie.de-noblet@lsce.ipsl.fr	DGVM (de Noblet-Ducoudre et al., 2004)
ORCHIDEE-crop	LSCE-IPSL, France	philippe.ciais@lsce.ipsl.fr	DGVM ((Valade et al., 2014))
LPJ-GUESS[ǀ]	KIT, Germany	almut.arneth@kit.edu	DGVM (Lindeskog et al., 2013;Smith et al., 2001)
JULES-crop[O]	Met Office, UK	pete.falloon@metoffice.gov.uk	DGVM (Van den Hoof et al., 2011)
CLM-Crop	LBNL, USA	adjones@lbl.gov	DGVM (Levis et al., 2012;Drewniak et al., 2013)
PEGASUS[ǀ]	Tyndall, UEA, UK	d.deryng@uea.ac.uk	Empirical/process (Deryng et al., 2011;Deryng et al., 2014)
GLAM[O]	SEE, Leeds, UK	a.j.challinor@leeds.ac.uk	Empirical/process (Challinor et al., 2004)
CGMS	WUR, NL	allard.dewit@wur.nl	Empirical/process (WOFOST) (van Diepen et al., 1989;Supit et al., 1994)
PRYSBI-2	NIAES, Japan	iizumit@affrc.go.jp	Empirical/process (Okada et al., 2011)
MCWLA[O]	IGSNRR, China	taofl@igsnrr.ac.cn	Empirical/process (Tao and Zhang, 2012)
ISAM	UIUC, USA	jain1@illinois.edu	DGVM ((Song et al., 2013))
DLEM-Ag	Auburn U, USA	renwei@auburn.edu	DGVM ((Gueneau et al., 2012))

[†] pDSSAT and pAPSIM are both part of the pSIMS framework, using inputs and assumptions as closely harmonized as is possible, allowing for a more direct comparison of inter-model differences.

* Four contributing GGCMs are built from the field-scale EPIC model and will be used for detailed explorations of the effects of different assumptions and configurations even within the same model.

[ǀ] Model participating in the 2012/2013 AgMIP/ISI-MIP Fast-Track.

[E] EPIC, DSSAT, and APSIM-based models will perform additional scenarios using alternative methods to model evapotranspiration in order to better understand the effect this important model choice has on assessments

[O] Models expected to participate starting in Phase 2.

Table 2: Priority 1 and 2 crops in phase 1, along with the number of models expected to contribute results for each crop.

Priority	Crops	Labels	# models	Notes
1	Wheat, maize, soy, rice	whe, mai, soy, ric	15-20	Required for all objectives
2	All others: Managed grass*, sugarcane, sorghum, millet, rapeseed, sugar beet, barley, cassava, field peas, sunflower, groundnuts, drybean, cotton, potato	mgr, sug, sor, mil, rap, sgb, bar, cas, pea, sun, nut, ben, cot, pot	Based on availability (>2)	Priority 2 crops will be considered case-by-case (require at least 3 model submissions)

* We consider only managed grassland productivity, not unmanaged pasture.

Table 3: General simulation configurations for phase 1.

Config	Long name	Description
Default	Default configuration	Simulations using default “best guess” choices for all inputs.
fullharm	Fully harmonized configuration	Simulations using harmonized inputs and assumptions.
harmnon	Harmonized with no nitrogen	Harmonized inputs with no nitrogen stress

Table 4: Output variables to be collected during GGCMI Phase 1. The first two variables are to be provided by every model; other variables are to be provided as possible by each model

Variable	Variable name*	Units (and notes)
Mandatory variables to be provided for all simulations		
Crop yields	yield_<crop>	t ha-1 yr-1 (dry matter)
Applied irrigation water	pirrww_<crop>	mm yr-1 (firr only, assume loss-free conveyance/application)
Additional variables below are to be provided as possible by each model		
Total Above ground biomass yield	biom_<crop>	t ha-1 yr-1
Actual growing season evapotranspiration	aet_<crop>	mm yr-1 (season only)
Actual planting date	plant-day_<crop>	day of year
Days from planting to anthesis	anth-day_<crop>	days from planting
Days from planting to maturity	maty-day_<crop>	days from planting
Nitrogen appl. Rate	initr_<crop>	kg ha-1 yr-1

Nitrogen leached	leach_<crop>	kg ha-1 yr-1
Nitrous oxide emissions	sn2o_<crop>	kg N2O-N ha-1
Accumulated precip, plant to harvest	gsprcp_<crop>	mm ha-1 yr-1 (season only)
Growing season incoming solar	gsrsds_<crop>	w m-2 yr-1 (season only)
Sum of daily mean temps, planting to harvest	sumt_<crop>	deg C-days yr-1 (season only)

* <crop> refers to the three-letter variable codes (whe, mai, ric, etc.) from Table 2.

Table 5: Simulation priorities for phase 1. For climate product descriptions see Table 9.

Priority	Crops	Climate product	Scenarios	Goal
Priority1	P1	WFDEI.GPCC, AgMERRA	*Default*, *fullharm*, *harmnon*	Establish key minimal yield estimates and comparisons
Priority 2	P1	WATCH.GPCC, PGF, GRASP, AgCFSR	*fullharm*	Extend range of years and characterize uncertainty due to multiple forcing products.
2.1 Climate track	P1	WFDEI.CRU, ERA-I and CFSR	*fullharm*	Evaluate the effects of different drivers (pure reanalysis, GPCC vs. CRU target for bias-correction, etc.)
2.2 Crop Track	P2	WFDEI.GPCC, AgMERRA	*fullharm*	Evaluate other crops that have a sufficient number of models and interest.

Table 6: Filename conventions for standardized model outputs.

Filename tag []	Values
[model]	pdssat, epic-iiasa, lpjml, etc. (see Table 1)
[climate]	watch, wfdei.gpcc, wfdei.cru, grasp, agmerra, agcfsr, Princeton (see Table 9)
[clim.scenario]	Hist
[sim.scenario]	default_firr, fullharm_noirr, etc. (see Table 3)
[variable]	yield, pirrww, plant-day, anth-day, etc. (see Table 4)
[crop]	mai, soy, whe, ric, mil, sor, etc. (see Table 2)
[timestep]	annual
[start-year]_[end-year]	1958_2001, 1980_2009, 1980_2010, etc. (see Table 9)

Table 7: NetCDF file dimension, variable, and attribute info.

Dimension/variable	Fill value	# type	Units	Range
Lon	NA	double	degrees east,	-179.75…179.75
lat	NA	double	degrees north	89.75…-89.75
time	NA	double	"growing seasons since YYYY-01-01 00:00:00" (YYYY varies, see Table 9)	1..T (T varies, see Table 9).
[variable]_[crop]	1.e+20f	Float	Varies (see Tables 2 and 4).	Varies

Table 8: Harmonized input variable sources for *fullharm* and *harmnon* configurations in Phase 1.

Variable	Source	Units	Notes
Planting window	(Sacks et al., 2010;Portmann et al., 2008;Portmann et al., 2010) & environment-based extrapolations	Julian days (Jan1= 1,…)	Crop calendar data (planting and maturity) for primary seasons.
Approximate maturity	(Sacks et al., 2010;Portmann et al., 2008;Portmann et al., 2010) & environment-based extrapolations	Days/GDD from sowing	Growing season length provided in number of days.
Fertilizers and manure	(Mueller et al., 2012;Potter et al., 2010;Foley et al., 2011)	kg ha-1 yr-1	Average nitrogen, phosphorus, and potassium application rates in each grid cell.
Historical [CO_2]	Mauna Loa/RCP historical	ppm	Annual and monthly [CO_2] values from 1900-2013.
Definition of time variable	Protocol choice	"growing seasons since YYYY-01-01"	YYYY is just the first year in the file. For a run 1958-2001, YYYY=1958. Values of time are independent of how to map growing season to calendar.
Season Definition	Protocol choice	Definition	AET and PirrWW defined as accumulated over the growing season, not over the calendar year.
Automatic irrigation	Guidance for parameter choices	Definition	Management depth = 40cm / Efficiency = 100% Lower event trigger threshold = 90% Max single AND annual volume = Unlimited
~~Automatic planting~~	~~Guidance for parameters choices~~	~~Definition~~	~~Min/max soil H2O at planting (40 cm) = 40/100%~~

~~Min/max soil temp at planting (10 cm) = 10/40 C~~

1

Table 10: Weather variables supplied per data set.

Variable	long name	Unit	WATCH	WFDEI	GRASP	AgMERRA	AgCFSR	PGF	CFSR	ERA-I	Notes
tas	daily mean temperature	°C	x	x	x	x	x	x	x	x	
tasmin	daily min. temperature	°C	x	x	x	x	x	x	x	x	
tasmax	daily max. temperature	°C	x	x	x	x	x	x	x	x	
pr	daily avg. precip. flux rate	~~K~~kg/m^2/s	x	gpcc (‘~~20~~10) cru (‘~~20~~12)	x	x	x	x	x	x	(incl. snow)
rsds	short wave downward	W/m^2	x	x	x	x	x	x	x	x	
rlds	long wave downward	W/m^2	x	x	NA	NA	NA	x	x	x	
wind	wind speed	m/s	x	x	x	x	x	x	x	x	
hur	relative humidity	%	x	x	x	at Tmax & Tavg	at Tmax & Tavg	*	x	*	
hus	specific humidity	kg/kg	x	x	NA	NA	NA	x	NA	x	
vap	vapor pressure	Pa	*	*	x	*	*	*	*	*	
ps	surface pressure	Pa	x	x	NA	NA	NA	x	NA	x	

x These variables are directly provided by the climate data provider.
* These variables are not directly provided but can be calculated using standard relationships (Bolton, 1980) which we implement in GGCMI.
NA These variables are not available from the given dataset.

◦ WATCH and WFDEI provide rainfall and snowfall separately. In the final version of the dataset used for GGCMI, these have been combined.

Table 11: combination of crop calendar data in GGCMI data sets.

GGCMI crop	MIRCA2000	SAGE	LPJmL	Days maturity to harvest
Barley	Barley	Barley ~~standard~~spring+winter	Wheat	7[7]
Cassava	Cassava	Cassava	Cassava	assuming 0[8]
Groundnuts	Groundnuts	Groundnuts	Groundnuts	0[9]
Maize	Maize	Maize	Maize	1-28[10] here 21
Millet	Millet	Millet	Millet	assuming 0
Potatoes	Potatoes	Potatoes	Sugarbeet	assuming 0
Pulses	Pulses	Pulses	Pulses	assuming 0
Rapeseed	Rapeseed	Rapeseed, winter	Rapeseed	same as wheat=7
Rice	Rice	Rice	Rice	0[11] or 8-12[12], here 7[1]
Rye	Rye	Rye, winter	Wheat	7[1]
Sorghum	Sorghum	Sorghum	Millet	0[13]
Soybean	Soybean	Soybean	Soybean	7-21[14] here 21
Sugarbeet	Sugarbeet	Sugarbeet	Sugarbeet	assuming 0
Sugarcane	Sugarcane	NA	Sugarcane	assuming 0
Sunflower	Sunflower	Sunflower	Sunflower	0[15]
Wheat	Wheat	Wheat, ~~standard~~ spring+-winter	Wheat	3[16] to 8[17] here 7

1

[7] Assuming quick harvests for barley, rice, rye and wheat as they are all threatened by pre-harvest sprouting, see e.g., http://www.dpi.nsw.gov.au/data/assets/pdf_file/0010/445636/farrer_oration_1981_nf_derera.pdf but allowing some time to dry after full maturity

[8] Can be anything from 0 days to up to 6 months, harvest on demand

[9] http://www.interaide.org/pratiques_old/pages/agro/3cultures/Phalombe_Mlwi_crop_management_2010.pdf, p 8

[10] http://www.smartgardener.com/plants/4159-corn-cherokee-white-flour/harvesting

[11] http://agris.fao.org/agris-search/search/display.do?f=1990%2FPH%2FPH90013.xml%3BPH8811720

[12] http://www.interaide.org/pratiques_old/pages/agro/3cultures/Phalombe_Mlwi_crop_management_2010.pdf, p 13

[13] http://www.interaide.org/pratiques_old/pages/agro/3cultures/Phalombe_Mlwi_crop_management_2010.pdf, p 14

[14] http://agris.fao.org/agris-search/search/display.do?f=2009%2FJP%2FJP0932.xml%3BJP2009005739

[15] http://www.interaide.org/pratiques_old/pages/agro/3cultures/Phalombe_Mlwi_crop_management_2010.pdf, p12

[16] http://agris.fao.org/agris-search/search/display.do?f=2009%2FJP%2FJP0938.xml%3BJP2009007527

[17] http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_windowLabel=T94008&_urlType=action&_pageLabel=_dwdwww_klima_umwelt_phaenologie shows that there is 16 days between “hard dough” stage (BBCH87) and harvest in Germany, and http://www.dwd.de/bvbw/generator/DWDWWW/Content/Landwirtschaft/Dokumentation/AgroProg/Kornfeuchte,templateId=raw,property=publicationFile.pdf/Kornfeuchte.pdf shows that there are about 8 days between “hard dough” and “fully ripe” (BBCH89) stages, so that the difference between “fully ripe” and harvest is 8 days as well.

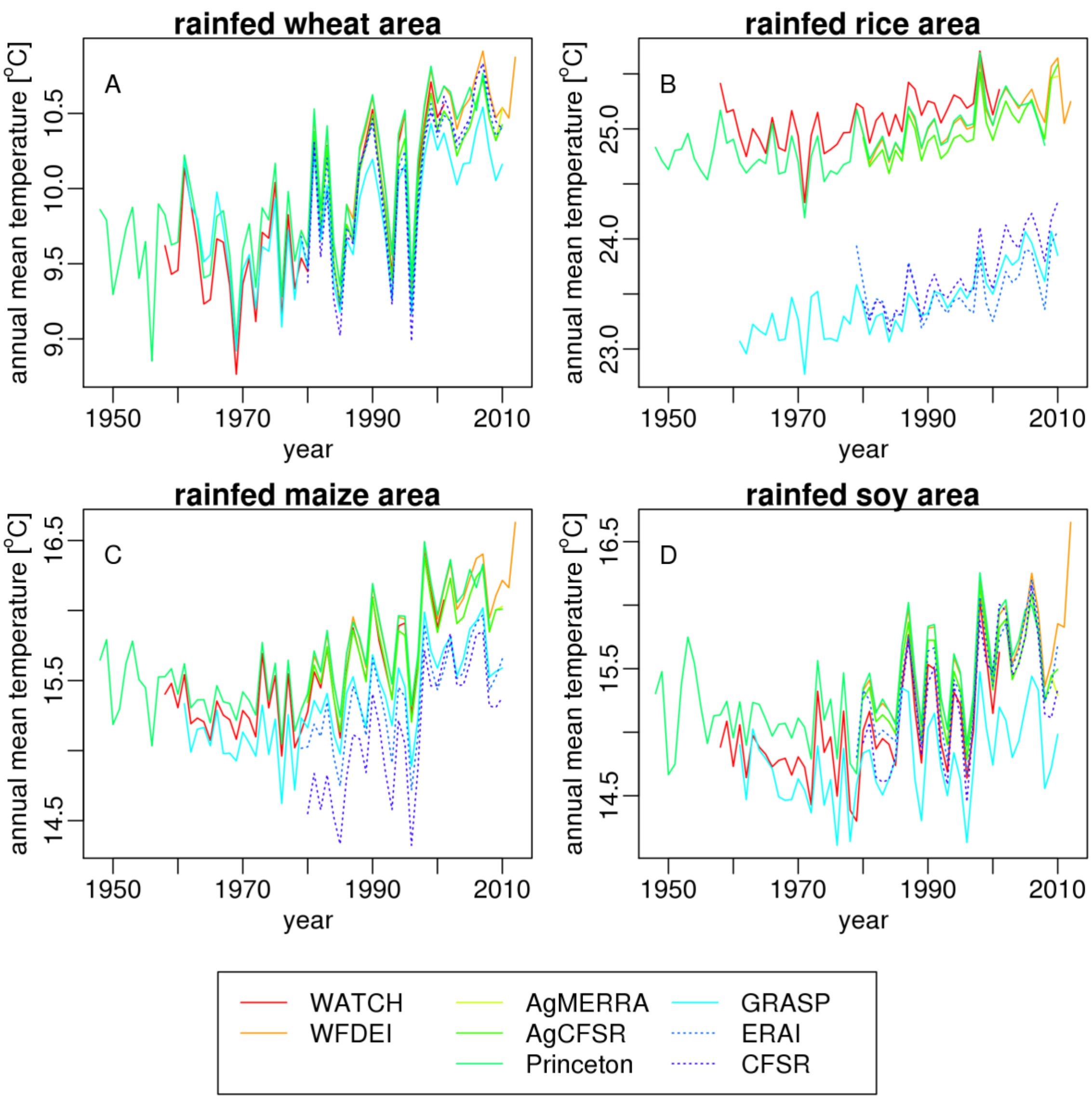


Figure 1: Area-weighted mean of annual temperatures [°C] for cropping areas for rain-fed wheat (A), rice (B), maize (C), and soy (D).

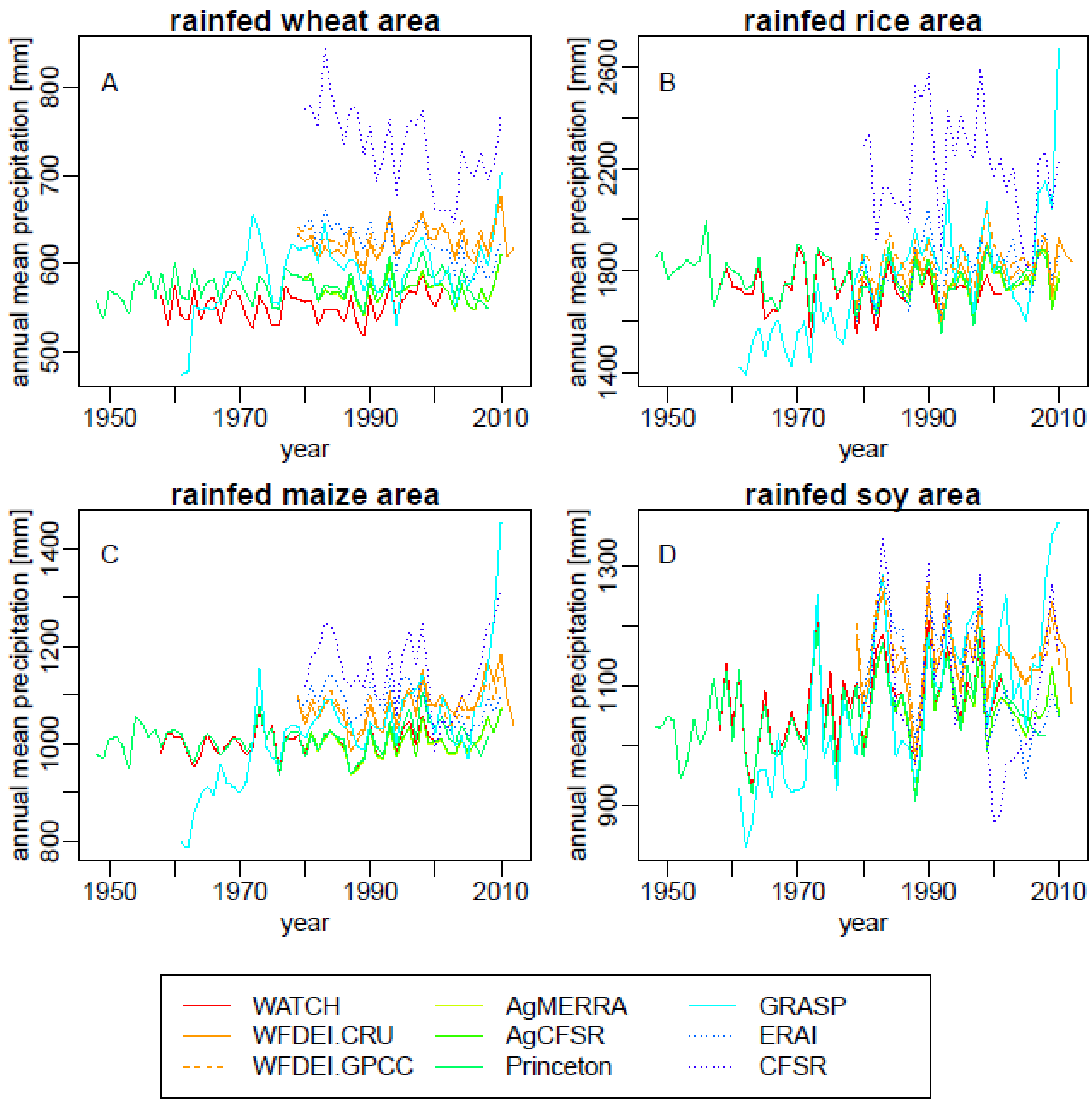


Figure 2: Area-weighted mean of annual precipitation [°C] for cropping areas for rain-fed wheat (A), rice (B), maize (C), and soy (D).

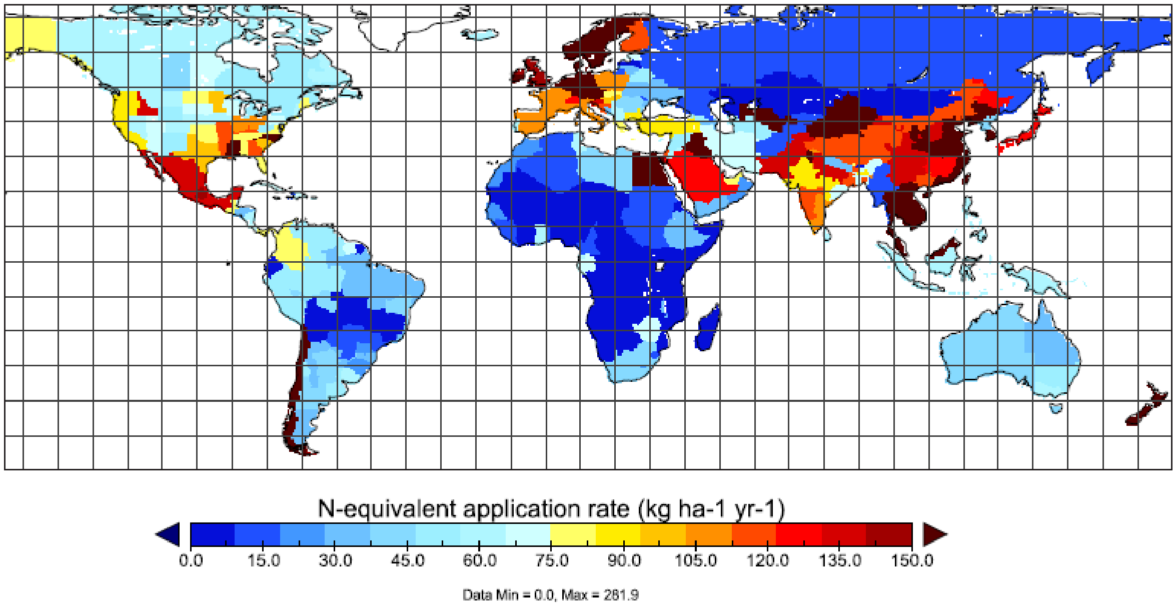
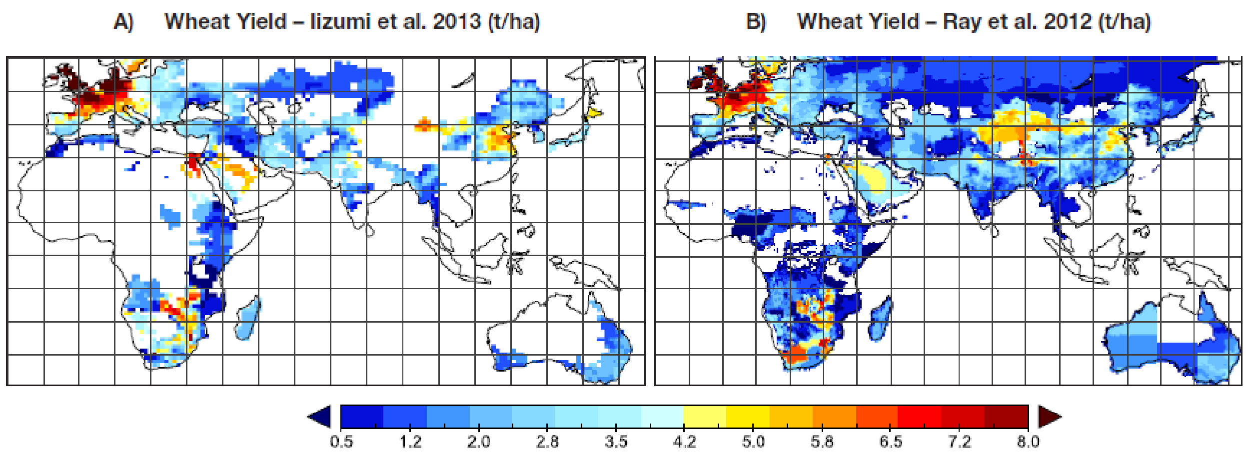


Figure 3: N-equivalent application rate of nitrogen fertilizers for the production of wheat.



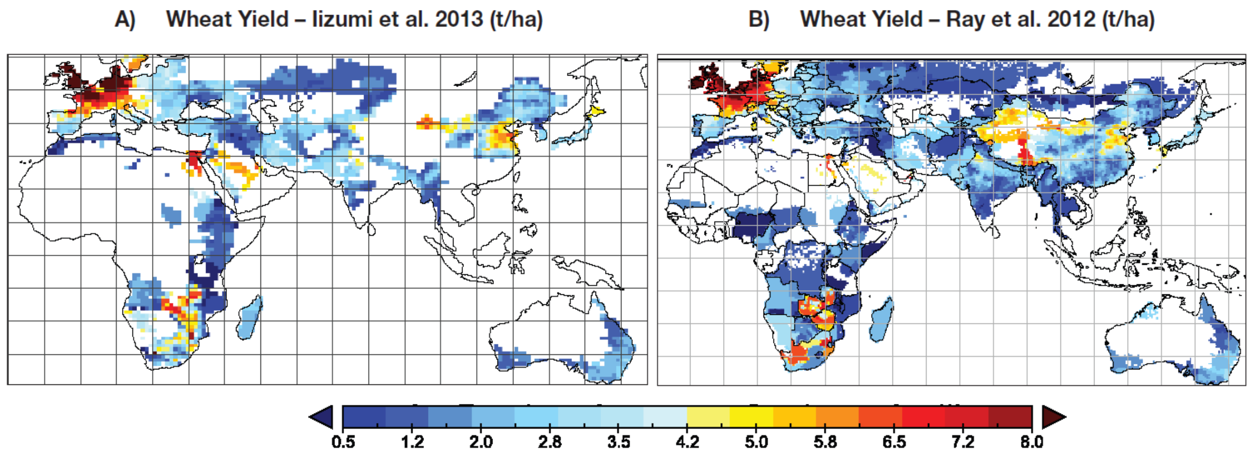


Figure 4: Example of historical ~~validation~~ evaluation data for year 2000 wheat yields from A) Iizumi et al 2013 (at 1.125 degrees spatial resolution) and B) Ray et al 2012 (aggregated from 5 arcminute to 0.5 degree).

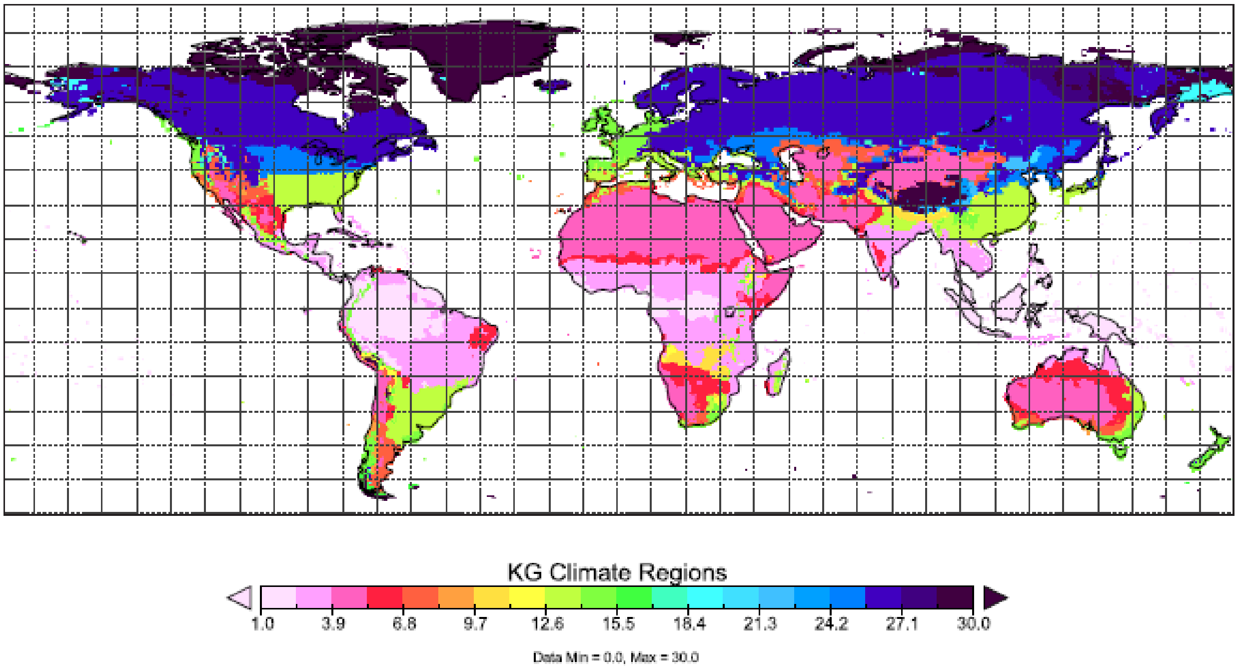


Figure 5: Example of a global Koeppen-Geiger climate classification.

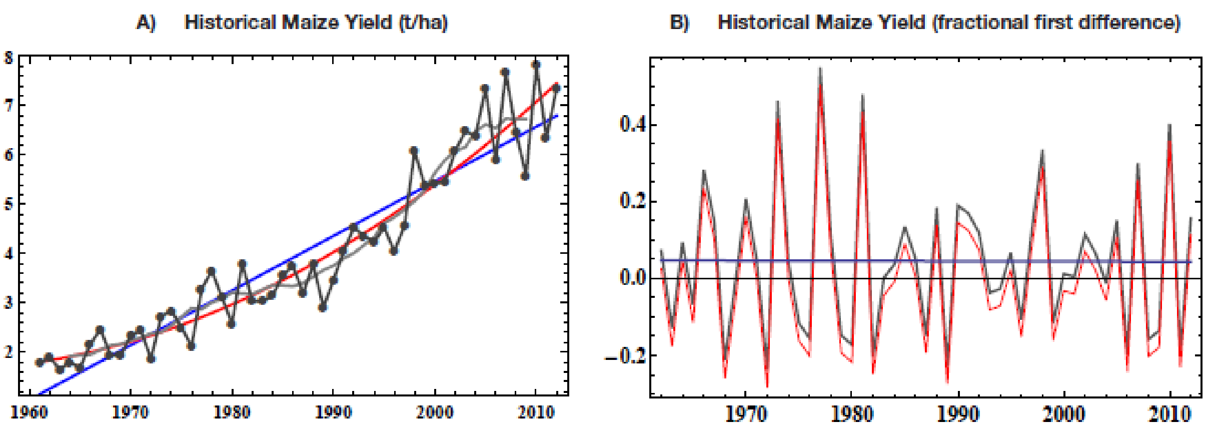


Figure 6: A) FAOSTAT yield for maize in Argentina (solid line and points) with the linear (blue) and quadratic (red) best-fits and 7-year moving average (gray). B) Fractional first difference of maize yields in Argentina (gray), the linear trend (blue line) and the fractional first difference with the trend removed (red).

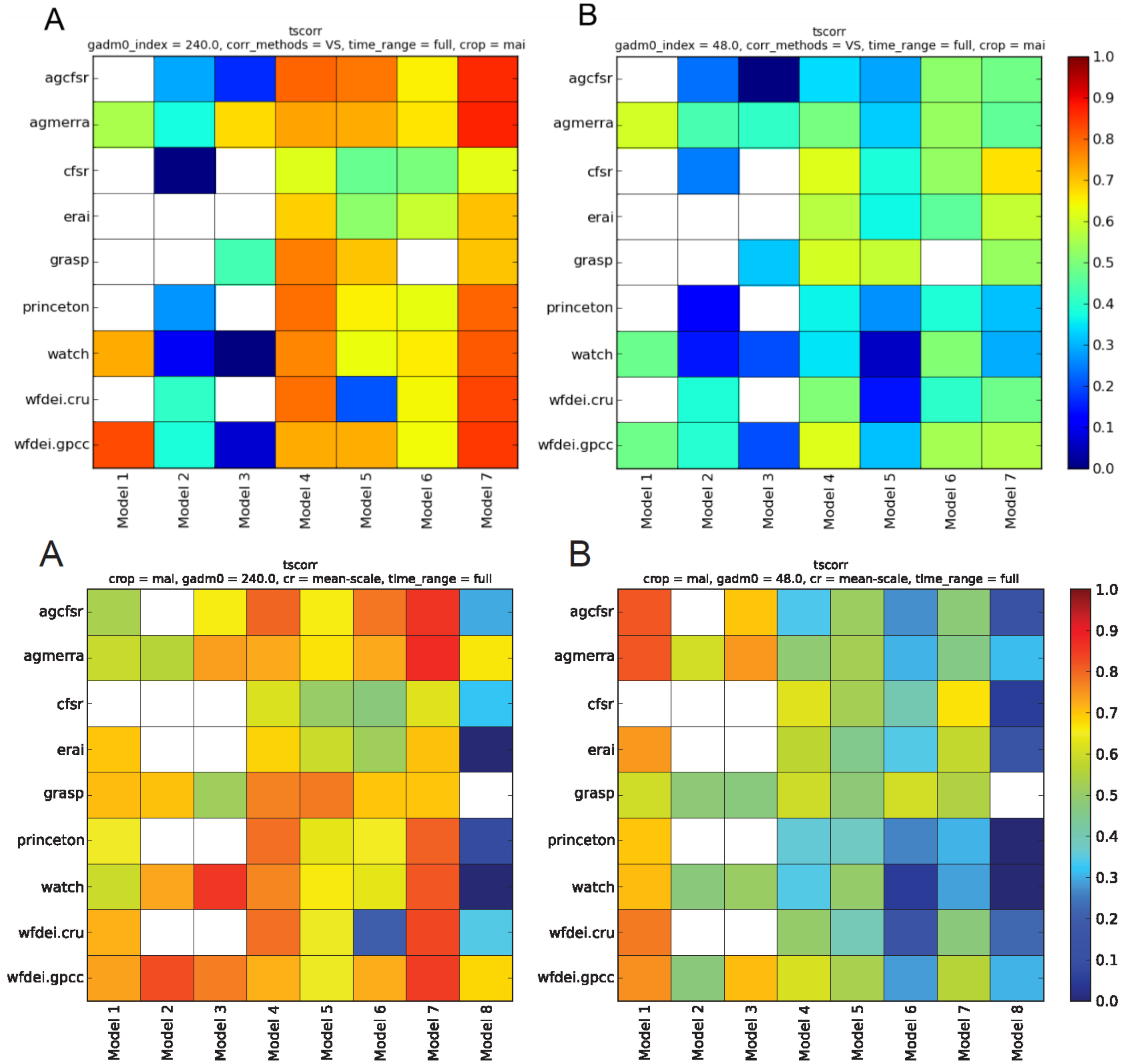


Figure 7: Examples of cross-sections of the multi-metric ~~validation~~ evaluation array for the top two maize-producing countries – the United States (A) and China (B). Plot shows time-series correlations for 8~~7~~ different crop models run (x-axis) with 9 different climate forcing datasets (y-axis). For each model/climate combination the best metric value among the scenarios (default,

fullharm, and harmnon) and detrending methods (linear, quadratic, moving mean, and trend-removed fraction first difference) are shown.