Interactive comment on "The Global Gridded Crop Model intercomparison: data and modeling protocols for Phase 1 (v1.0)" *by* J. Elliott et al.

J. Elliott et al.

jelliott@ci.uchicago.edu

Received and published: 9 October 2014

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 determin-

ing 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.