Manuscript prepared for Geosci. Model Dev. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 27 October 2015

# Response to reviewers: Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts

**Karina Williams**[1] **and Pete Falloon**[1]

[1]Met Office Hadley Centre, FitzRoy Road, Exeter, Devon EX1 3PB, UK

Correspondence to: Karina Williams (karina.williams@metoffice.gov.uk)

**Abstract**

This document contains the authors' responses to the reviews of the GMD discussion paper "Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts". We thank the reviewers for their careful reading of the manuscript and will address each of their points in turn. All figure numbering refers to figures in this document (rather than in the manuscript).

## 1 Response to Reviewer 1

This manuscript evaluates the impact from reduced meteorological forcing and/or initialization data on crop yields with the JULES-Crop model. The goal is to identify the input variables responsible for interannual yield variability in order to find the minimum required data needed to run the model over seasonal timescales. Sub-daily, daily, and daily climatology inputs are compared, and techniques for initialization are also explored. Overall, the methodology can simplify the setup and data needed to run the model, making JULES a useful tool for seasonal crop forecasts. The paper is well written and concise, however the manuscript could benefit from some clarifications and the study should include additional analysis.

I understand the ease of comparisons between the model runs with the control, but there was no attempt to validate the model against observational data. Only once did the manuscript mention that the control run (as performed in another study) correlated well with observations, but correlation with a control run doesn't mean the reduced model correlates well with observations. I would like a stronger attempt to validate against observations; without it, seasonal forecasts by the model won't be very useful.

This is a very good point. In this current study, we are focussed on approximations which remove some of the technical barriers to using JULES-crop with seasonal model forcing. However, any practical use of JULES-crop in a seasonal forecast system would need careful verification against observations during the hindcast period. This work would be very specific to the use case. For example, it is likely that the four generic crop parametrisations that are used here would not be sufficient and a parametrisation that more closely resembled the crop that the stakeholder was interested in would be required, and this tuning would depend on the crop observations available. In addition, the validation would be likely to focus a particular region of interest, bringing in a detailed knowledge of factors affecting yield and meteorology in that region. The metrics used for verification would also depend on the stakeholder requirements e.g. if they were interested in predicting the mean yield, or in predicting when yield falls below a certain threshold. Also, the verification would be specific to the particular seasonal forecast model used to force the crop model. We will add text to emphasize this point.

We have additionally calculated the correlation between time series of model yield and FAO country yield, and compared these across model runs, to check that these support our conclusions (see Figure 1 and Figure 2). However, we felt that the message of the paper was clearer if we focussed on the effects of different approximation on model results, particularly considering that many of the correlations with FAO country data were low.

Suggested change to manuscript:

– Add the text "It is important to note that while this study provides a practical methodology for driving JULES-crop with seasonal forecasts, given commonly available forcing and initialisation data, there are many aspects of the uncertainty chain that remain to be addressed. For example, once an application has been identified (e.g. a decision threshold based on the yield of a particular crop in a particular region), a thorough validation would need to be performed of the relevant model diagnostic against observational data and against hindcast driven runs."

I think the analysis of model runs could be expanded. Using a Pearson Correlation alone is not sufficient to conclude the model is robust enough to run with coarser input data. What is the slope of the relationship between yields from the control and reduced input models? What happens during the growing season - are the growing phases and plant development correlated? How well does the reduced model perform during years with low crop yields? High crop yields? It is important to understand the outliers as well as the mean. The reduced input model must be able to capture the extremes and not just the mean.

Again, we feel the reviewer is making a very important point, particularly as there are many stakeholders of seasonal forecasting applications which would be particularly interested in forecasting extremes. We have chosen to evaluate the model in terms of the Pearson correlation coefficient for comparison with the Osborne et al. (2015) study and because this is often used when assessing the skill of seasonal forecasts e.g. Scaife et al. (2014). While the correlation between maize yield observations (FAO) and the control run is significantly different from zero at the 95% CL, it is not high (Osborne et al. (2015) found it was 0.5 globally). We therefore felt that we would not get any added value from using a more sophisticated metric to compare the control run to the various approximations. (For comparison, for a correlation to be significantly different from zero at the 95% CL with this sample size (using large sample size approx, assuming normally distributed, assuming no autocorrelation), the magnitude would need to be greater than 0.3).

The growing phases and plant development will be very similar in all the runs apart from sens-P, since the temperatures used to drive the model are the same in all the runs apart from sens-P and plant development is calculated directly from temperature (the exception to this is when the model hits an internal threshold e.g. it does not have enough carbon to be considered as established or if it reaches an unphysically high leaf area index).

While we have used the Pearson correlation coefficient to summarise our results, we have also examined the yield time series and scatter plots to check that there are no surprises in the results and make sure that this summary statistic is not misleading. We have used Figures 3, 4 5 to confirm that the approximations we propose in this paper (using the disaggregator, initialising from climatology and setting all driving variables apart from temperature and precipitation to climatology) capture the high and low crop yield years by plotting the annual yields for individual countries as scatter plots and checking that there is no obvious non-linearity or undue influence from outliers. We have done the same for the global time series - Figure 6 and Figure 7 contain a scatter plot for for each combination of runs that we quote a global correlation for in the manuscript. We will add 6 and Figure 7 to the appendix and refer to them in the manuscript text.

Suggested change to manuscript:

- Include the global yield scatter plots (Figure 6 and Figure 7) in the appendix.
- Add the text "The annual `control` yield is plotted against the annual `disagg` yield in Figure 6, and shows no obvious deviations from linearity, even at the extremes"
- Add the text "These results are presented as scatter plots in the appendix." in the caption of Table 2.
- Add the text "The scatter plot of these yield time series (figure 7) shows that the relation between the outputted yield is well approximated by a linear fit."
- Add the text "(see Figure 6 for scatter plots)"

The correlation between the various model runs is quite high; with the exception of precipitation, the model does not seem to have much sensitivity to the input data. Is it possible this could be the result of overtuning of model parameters?

We feel that it is unlikely to be the result of overtuning of the model parameters, as the generic crop parametrisations have not been heavily tuned. However, it could indicate a need for improvements in the soil hydrology, the effect of soil moisture on GPP or the root distributions in JULES (and these are all areas of current research in the JULES community).

Suggested change to manuscript:

– Add the text "It should be noted that the processes and parameters which govern the response of the crop model to the soil moisture distribution, such as the soil water factor $\beta$ and the root distributions in JULES are therefore keys areas for future model development." to the conclusion.

1. The model description seems sufficient for the context of this paper, however the only reference to the original model description in section 2 is a footnote. This should be moved into the text.

We will implement this.

2. The authors mention at least three versions of JULES (4.0, 4.1, and JULES-crop). Please state which version is used in this manuscript (and be consistent when referencing the model).

We thank the reviewer for drawing our attention to this important issue with the current manuscript. JULES-crop is the name of the crop parametrisation within JULES. JULES 4.0 was the first official version of JULES which contained JULES-crop. JULES 4.1 was the first official version of JULES to allow irrigation. The version of JULES used for all the runs except `irrig` was branched from the JULES trunk between version 4.0 and 4.1, and contained some bug fixes and new features that were needed for these runs. These bug fixes and new features have now been added to the JULES trunk, so are now available to the entire JULES community. The version of JULES used for the irrig run was branched from the JULES 4.1 release with an added bug fix, which, again, was subsequently committed to the JULES trunk. We have confirmed that using the control run configuration in both versions of JULES gives the same results (and, also, give the same results as Osborne 2015). While this was the best solution at the time, and the two branches of JULES that we used are both available in a JULES code repository, we have decided to redo this study with JULES 4.2 (i.e. we repeated the 9656 runs used in this paper with JULES 4.2, re-creating the initialisation climatology and redoing the analysis and plots). Being able to refer to an official version of JULES is better for clarity and anyone wishing to reproduce our results will know that all the code has been through the official JULES review process.

Suggested change to manuscript:

- Add the text "All runs were performed with JULES 4.2."
- Add the text "As of JULES version 4.0, it includes a crop parametrisation (JULES-crop) which introduces an additional tile for each crop simulated. "
- All plots and results have been updated. We have confirmed that none of the numbers quoted in the draft for maize have changed (the differences between the results for these runs and the previous runs are in higher sig. fig.).

3. In section 3.3, what is a dump file? It would be clearer if the authors stated the initialization came from the end of the control spin up rather than the beginning of the control main run.

We have simplified this by redoing the run with the same initialisation as the control run, since the long spin up is designed to make the model insensitive to the initialisation.

Suggested change to manuscript:

- Change "The run was initialised with the dump file from the beginning of the control main run and then spun up by repeating the first ten years five times." to "The run was initialised and spun up in the same way as the `control` run."

4. In section 3.4, what data was used to construct the daily climatologies?

We will add this information (for section 3.4 and 3.5).

Suggested change to manuscript:

- Change "In order to investigate the sensitivity to variability in different parts of the driving data, we created daily climatologies of each driving data variable in the full disaggregated run." to "In order to investigate the sensitivity to variability in different parts of the driving data, we created daily climatologies of each driving data variable in the full disaggregated run. For example, for each gridbox, the value used for the 1st January in the precipitation climatology was the mean over the CRU-NCEP precipitation on every 1st January from 1960 to 2009 in that gridbox."

- Change "We created a climatology for the initialisation variables for each day of the year, using daily means outputted from the `control` run." to "We created a climatology for each initialisation variable, for each day of the year, using daily means outputted from the `control` run and averaging over 1960-2009."

5. How were the sens-* runs initialized - was it similar to the disagg run using the "beginning of the control main run"?

The sens-* runs were initialised in the same way as the control run, although the run is not sensitive to this due to the long spin up. We will add a sentence to the paper to clarify this.

Suggested change to manuscript:

- Change "Each run had 50 years of spin up (first 10 years five times) before starting the main run (this was particularly important for the `sens-T` run)." to "The run was initialised and spun up in the same way as the `control` run."

6. The authors state in section 3.4 that the spinup "was particularly important for the sens-T run," but no explanation is given. In fact, the sens-T run isn't mentioned again in the paper, with the exception of one footnote, even though it has the worst correlation with the control run than any other sens-* run. I think this simulation deserves an explanation.

We will remove the reference to the spinup being "particularly important for the sens-T run" as this is not relevant to the rest of the argument (since the init run is a better way to test the effect of initialisation and spinup). We will add further description of the sens-T run.

Suggested change to manuscript:

- Add this paragraph: "If temperature is the only variable allowed to vary between years (i.e. the `sens-T` run), then the global mean maize yield is 10.7 $\mathrm{Mg\,ha^{-1}}$, with standard deviation 0.23 $\mathrm{Mg\,ha^{-1}}$. This reduction in standard deviation compared to the disagg run is consistent with the reduction in standard deviation seen when the effect of soil moisture was removed (the `irrig` run). Unsurprisingly, figure \ref{fig:corrwithdisagg} (bottom left) shows that the sens-T run does not correlate well with the disagg run in areas where the sens-P run had a higher correlation."

- Removed the reference to the spinup being "particularly important for the sens-T run".

- Change "If both daily precipitation and daily mean temperature are allowed to vary (run `sens-PT`), the gridbox correlations with the `disagg` run are much more spatially uniform than for the `sens-P` run: 81%" to "If both daily precipitation and daily mean temperature are allowed to vary (`sens-PT`), the gridbox correlations with the `disagg` run are much more spatially uniform than when either of these variables are varied on their own: in the `sens-PT` run, 81%"

7. It would be interesting to know which variables are included in the initialization with climatology, or if the list is too long to include, perhaps those that the model shows the most sensitivity.

Table 1 gives a list of the initialisation variables used in this study. As suggested, we will include some examples in the manuscript.

| | |
|---|---|
| cropdvi | Development index for each crop pft. |
| croprootc | Root carbon pool for each crop pft (kgC m-2). |
| cropharvc | Carbon in harvest pool for each crop pft (kgC m-2). |
| cropreservec | Carbon in stem reserves pool for each crop pft (kgC m-2). |
| croplai | Leaf area index of each crop pft. |
| cropcanht | Height (m) of each crop pft. |
| seed_rain | random seed (used in disaggregator) Only required if initialising from a dump file. |
| sthzw | Soil wetness in the deep ('water table') layer beneath the standard soil column. This is the mass of soil water (liquid and frozen), expressed as a fraction of the water content at saturation. |
| zw | Depth from the surface to the water table (m). |
| canopy | Amount of intercepted water that is held on each tile (kg m-2). |
| tstar_tile | Temperature of each tile (K). This is the surface or skin temperature. |
| cs | Soil carbon (kg m-2). |
| gs | Stomatal conductance for water vapour (m s-1). |
| rgrain | Snow surface grain size ($\mu$m) on each tile. |
| sthuf | Soil wetness for each soil layer. This is the mass of soil water (liquid and frozen), expressed as a fraction of the water content at saturation. |
| t_soil | Temperature of each soil layer (K). |
| snow_tile | total snow on the tile (and is subsequently put onto the ground at tiles that distinguish between ground and canopy stores). |
| sthu_irr | Unfrozen soil wetness of each layer as a fraction of saturation in irrigated fraction. |

**Table 1.** List of initialisation variables used in the JULES runs performed for this paper. Descriptions taken from the JULES user guide.

Suggested change to manuscript:

– Add the text "The model requires 16 initialisation variables, on multiple model layers or tiles, such as tile surface temperature and moisture in soil layers as a fraction of water content at saturation (see JULES user guide for full list)"

8. Section 4 only include maize results, noting that the other crop types showed similar conclusions. However, it might be useful to include the other crop results, perhaps in Table 2 for comparison, or as supplementary material.

Table 2 and Figures 8, 9, 10, 11, 12, 13 give the results for the other crops. We will include these as supplementary material.

Suggested change to manuscript:

– Add Table 2 and Figures 8, 9, 10, 11, 12, 13 to supplementary material.

– Add the text "Results from the other crops are given in the supplementary material."

1. There are several footnotes in this manuscript, which are discouraged by GMD. I think all of the footnotes can be incorporated into the manuscript.

We will implement this.

2. Page 4605, Line 7: please spell out "respectively"

We will implement this.

3. Page 4606, Line 9: the authors state the run was from 1960-2009, but on line 7, stated the CRU-NCEPv4 was extended to include 2012. Did the model run go to 2012?

No, the model runs were 1960-2009 (which is the same as Osborne 2015). The dataset had been extended to include 2012 for the Global Carbon Project, but we did not make use of this extra data.

Suggested change to manuscript:

- Take out "(extended to include 2012)" as this is not relevant here.

4. Page 4606, Line 21: "Each run" implies there was more than one run; I assume there was only one irrigated run.

Thanks for spotting this mistake. We will correct "each run" to "the run".

5. Figure 2: the caption states the comparisons are with the control run, but the section 4 comparisons refer to the disagg run.

Again, thanks for spotting this mistake. We will correct the caption of figure 2.

## 2 Response to Reviewer 2

1) Unfortunately, the main objective(s) of the paper is (are) not clearly stated in the intro-

duction section. From this part, it seems that there are two main questions followed by this paper, both centered on "data availability" issue:

* (Page 4603, Line 9) "to investigate whether the yield variability can be modelled sufficiently well if only a subset of the forcing variables are taken from the seasonal forecast and the others set to climatology, or if the model is forced with daily meteorological data and disaggregated internally to the model time step"

* (Page 4603, Second paragraph) Due to significant practical challenges to estimate the initial conditions variables, such as soil moisture, on the start date of seasonal forecast runs, "we investigate the loss in predictability of yield if the run is started on the sowing date of the crop in that gridbox and initialised by the climatological values for that date"

I suggest authors provide an explicit statement including the objective of their study and provide a more critical literature review so the readers can clearly understand the gaps/needs in the literature, and the contribution of this paper. While doing this, I also suggest avoiding using confusing sentences while talking about the objectives and challenges addressed by your work (see my detailed comments).

Also, the introduction section is heavily built on using seasonal forecasts as the input to crop models, while, as far as I have understood, this is not the main target of this study, and is in fact an implication of this study.

The motivation for the study is to overcome some of the technical obstacles in using the JULES-crop model with seasonal forecasts. A critical sticking point is the data availability issue, and that is the reason why we investigate the effect of reducing these data requirements. The approximations we suggest are entirely motivated by this need and are not generally applicable to JULES runs - in general it is highly recommended that the full driving data and initialisation requirements are fulfilled when running JULES. We evaluate these approximations entirely with this application (crop yield interannual variability) in mind, and are able to make compromises in model performance which may not be suitable in other contexts (e.g. a study focussing on energy fluxes). We will edit the manuscript to clarify these objectives.

Suggested change to manuscript:

– We have redrafted the introduction. As part of this, we have emphasized the motivation for the paper by explicitly stating: "The aim of this paper is to address those issues that are centred around the availability of data, by investigating to what extent the interannual variability of the modelled yield can be captured if some of these data requirements are relaxed."

* Given that the main difference between model results from set up 3.1 and 3.3 is due to JULES disaggregator, it might be interesting to show to what extent JULES disaggregator can reconstruct the 6-hour climate data, particularly for precipitation and temperature.

We have clarified this by adding more information to the manuscript. A detailed description, the difference between disaggregated and non-disaggregated driving variables at a selection of FLUXNET sites and the impact the disaggregation on evaporation and runoff on global runs is available in Williams 2014. The JULES disaggregator does not directly reconstruct the six hourly data - it disaggregates directly to the internal model timestep, which is an hour in our runs.

Suggested change to manuscript:

– Change "The internal JULES disaggregator (described in Williams and Clark (2014)) was used to disaggregate this forcing data to the model time step." to "The internal JULES disaggregator (described in Williams and Clark (2014)) was used to disaggregate this forcing data to the internal model time step of one hour. For temperature, this involves adding a sinusoidal diurnal cycle. Precipitation in a day is modelled as occurring in one rainfall event of constant intensity, with a duration that depends on the precipitation type."

* Please elaborate more on set-up 3.4 and Table 1; specifically, what exactly do you mean by "we created daily climatologies of each driving data variable in the full disaggregated run"? How did you create "climatologies"? Is it the long-term average of climate variables for each day of the year? Then, did you run your model only for one year? If so, then how are you able to comment on inter-annual variability (this is also the case with set up 3.5, line 23: "Each run lasted 1 year")?

The climatologies are created by averaging the values for one day e.g. 1st January over all the instances of that day in 1960-2009. The inter-annual variability comes from the driving variable(s) which were not set to climatological values e.g. the interannual variability of the sens-P run came from the interannual variability of the precipitation. The sens-* runs lasted from 1960-2009 (excluding spinup). For the init runs, each run lasted one year, then these results were concatenated to get a time series. We have clarified this in the manuscript.

Suggested change to manuscript:

- Add the text "(for 1960 to 2009, as before)"

- Change "In order to investigate the sensitivity to variability in different parts of the driving data, we created daily climatologies of each driving data variable in the full disaggregated run." to "In order to investigate the sensitivity to variability in different parts of the driving data, we created daily climatologies of each driving data variable in the full disaggregated run. For example, for each gridbox, the value used for the 1st January in the precipitation climatology was the mean over the CRU-NCEP precipitation on every 1st January from 1960 to 2009 in that gridbox."

- Change "Each run lasted 1 year and the full 6 hourly driving data was used." to "The full 6 hourly driving data was used. Each run lasted one year and the annual yields were concatenated to get a 48 year time series for each crop in each gridbox.".

* Please note that, you should assume that readers are not familiar with the model, scenarios you have considered, assumptions you have made, methods you have employed, etc. Therefore, I highly recommend to consider revising the writing of important parts of the paper (for example section 3) to ensure readers to a high extent understand important details of your work. I really cannot comment that much on your results and analyses based on the last two scenarios since I cannot understand the underlying assumptions.

We have expanded the model description and experimental set-up sections of the paper. However, our aim is to review the particular features of the model which are directly relevant to this study, rather than a comprehensive description of the JULES land-surface model and its usage (we cite the relevant summary papers and the JULES user guide for these details).

Suggested change to manuscript:

– Add the text "Potential leaf-level photosynthesis (unstressed by water availability and ozone effects) is calculated as the smoothed minimum of three potentially limiting rates, based on Collatz et al. (1991, 1992): (a) the Rubisco-limited rate, which depends on the maximum rate of carboxylation of Rubisco, (b) the light-limited rate and (c) the rate associated with the transport of photosynthetic products for C3 plants or PEP-Carboxylase limitation for C4 plants. The vertical profile of radiation through the canopy can use either the big-leaf approach (following Beer's law) or a multi-layered canopy radiation scheme, which treats the direct and diffuse components of the radiation separately. The latter can optionally include the direct component of the direct beam radiation ('sunflecks')."

– Add the text "Net primary productivity (NPP) is calculated by scaling the leaf-level photosynthesis to the canopy level and subtracting plant maintenance and growth respiration. "

– The descriptions of the set-up of the control, irrig, disagg and sens-* and init runs have been expanded.

3) I have not read the Osborne et al. (2015) or any other earlier JULES and JULES-crop papers. However, I am really concerned about model "calibration" and "validation", and I would appreciate if you can comment on my concerns. While working with large scale models (of any kind), especially if the time interval spans over several decades, relying on simple correlation factors to assess the performance of the model is very misleading. It is particularly important to investigate model performance (especially if it is calibrated using different year types, i.e., wet, dry, normal) during dry seasons, because outliers usually are not paid enough attention when dealing with "average" values over a long time span.

Osborne et al. (2015) presented the JULES-crop crop parameterisation within JULES and presented plots of yield time series (global and country averages) against FAOSTAT data with correlations. While they did not use the categories "wet", "dry" and "normal" to distinguish different years, information on yields for individual years can be obtained from the time series plots. Osborne et al. (2015) also compared the global yield distribution against observations from Monfreda et al. (2008) and also performed site validation looking at sensible and latent heat energy fluxes, crop height, LAI, GPP and yield against FLUXNET observations. A wide variety of methods have been used to evaluate other components of JULES, depending on which is most appropriate for the variables or processes under consideration; for example, Weedon et al. (2014) uses cross-spectral analysis to look at discharge to model an individual basin in the UK, where detailed validation data is available.

4)Do you think it is enough to discuss the performance of the model (and consequently talk about the impacts of different parameters on interannual variability) which simulates 50 years of crop yield globally just by using a correlation coefficient? This is indirectly related to the previous comment, but in that comment, I was just concerned about model validation. Here, I really want to know how each different scenario is compared with the baseline. Is there any difference between different set-ups in dry, normal, wet years?

Please see our earlier comments (to Reviewer 1) on the use of the Pearson correlation coefficient and the new plots for the appendix: Figure 6 and Figure 7.

5)Page 4609, Last paragraph (Also Figure 1): I can see that in many parts of the world where rainfed agriculture is dominant (e.g., sub-Saharan Africa), correlation coefficient between season precipitation and crop yield is very low. How can you justify this result?

In these regions, the index constructed from the total precipitation during the crop season is not able to capture all of the variability in the model yield (note that this figure shows the correlation between season precipitation and crop yield from the control run).

6)Figure 1 (results related to disagg): In this set-up there is no irrigation. How is it possible that the model can simulate maize yield very well even in regions where irrigation is heavily practised (India, Central US)?

The comparison is between the control and disagg model runs, neither of which have irrigation (it is not a correlation with observed yields). We would expect the control and disagg runs to behave similarly in regions where water availability is a limiting factor unless irrigation is applied.

7)Page 4612, Line 19: "since obtaining accurate initialisation data on the timescales needed for seasonal forecast runs is a particularly significant practical challenge". You also mentioned this in the introduction section as a kind of motivation to study the impact of initialization on yield variability. As I also mentioned in another comment below, this is confusing to me. If seasonal forecasts are to be provided say one month before the crop growth season, you can still run your model, say couple of months in advance using observational data to initialize the soil moisture, and then seasonal forecasts data will be fed into the model. Now, if seasonal forecasts are to be provided on the sowing date, you still need to run the model in advance using observation data to initialize the soil moisture. I really cannot understand this argument. I hope in the next version you can clearly explain your point.

Unfortunately, since the observational data is not available soon enough (or at all), this is not practical operationally. For example, NCIC observations for the UK are usually not available for at least a month after recording due to quality control etc. Also, it is very difficult to get the observations of all of variables required to initialise JULES (see Table 1 for the list).

8)Page 4612, last paragraph: Although seasonal forecasts in high temporal resolution (sub-daily) are not available everywhere or provided by all models, but there are many models that provide seasonal forecasts of all climate/weather variables in sub-daily time scale. Moreover, one of your arguments in the beginning that using high resolution data requires high storage, which is correct, does not seem to make much difference in the near future given the emergence of cloud and high performance computing. Given these conditions, where else do you see the contribution and importance of your work?

The data storage problem is very important currently, and we do not see it becoming less important in the next 3-5 years. Seasonal forecast systems are rapidly increasing in resolution and in number of ensemble members. Institutions running the forecast models already need to make careful decisions about exactly what output they can afford to save and often the compromise is on time resolution. The Hadley centre seasonal forecast system (GloSea5), for example, does not currently output the full set of data required to drive JULES offline at subdaily timescales from its operational suite, despite using JULES internally as its land-surface scheme. Secondly, if the seasonal forecast centres were able to output the ideal driving data set for JULES, there would still remain the problem of getting this data to the impacts modelling community. For example, in the EUPORIAS project, seasonal forecasts are stored on a central location and then project partners download the data to their local machines, with bias correction if required. There were therefore limitations imposed by the size of the central storage space, the size of the local storage space and also the bandwidth to get the data from one site to the other. It is not an exaggeration to say that the issues we address in this paper were barriers to the investigation of the use of JULES-crop with seasonal model forcing.

1. Page 4600, Line 24: "However, existing studies of crop model performance focused on seasonal forecast applications show considerable variation depending on the region, scale, processes and crops involved". Isn't this something that you would expect? I mean regardless of seasonal forecasts, crop simulation does depend on region, scale, processes, crops, etc.

We will clarify the text to illustrate that we are specifically talking about variation in performance rather than variation in the methodology used to simulate the crops.
Suggested change to manuscript:

– Change "However, existing studies of crop model performance focused on seasonal forecast applications show considerable variation depending on the region, scale, processes and crops involved" to "However, existing studies of crop model performance focused on seasonal forecast applications show considerable variation in skill depending on the region, scale, processes and crops involved"

2. Could you please elaborate more on this part (Page 4601, Line 11): "Using a statistical approach to assess the reliability of global-scale seasonal crop failure hindcasts, Iizumi et al. (2013) found that within-season hindcasts generally reproduced inter-annual variability in observed yields in major wheat exporting countries (r2 = 0.56-0.61) better than pre-season hindcasts (r2 = 0.43-0.59)."

* What exactly do you mean by seasonal crop failure hindcast?

* Also what exactly do you mean by with-in season hindcast and pre-season hindcast?

We will add this to the manuscript.
Suggested change to manuscript:

– Add the text "Using a statistical approach to assess the reliability of hindcasts of global-scale yield decreases of at least 5%, Iizumi et al. (2014) found that within-season hindcasts with lead times of 1-3 months generally reproduced inter-annual variability in observed yields in major wheat exporting countries better than pre-season hindcasts with lead times of 3-5 months."

3. What is the baseline when you phrase (page 4601, Line 23) "Palmer et al. (2004) and Cantelaube and Terres (2005) also found that downscaling seasonal hindcasts improved crop model performance"? Do you mean comparing to not-downscaled seasonal hindcasts?

We will add "when downscaled seasonal forecasts were used instead of the original, pre-downscaled versions" to the manuscript.

Suggested change to manuscript:

- Change "Palmer et al. (2004) and Cantelaube and Terres (2005) also found that downscaling seasonal hindcasts improved crop model performance - the $r^2$ value of simulated biomass for the whole of Europe increased from 0.62 to 0.69 with greater regional improvements." to "Palmer et al. (2004) and Cantelaube and Terres (2005) also found that downscaling seasonal hindcasts improved crop model performance - the $r^2$ value of simulated biomass for the whole of Europe increased from 0.62 to 0.69 with greater regional improvements when downscaled seasonal forecasts were used instead of the original, pre-downscaled versions."

4. Page 4602, Line 14: Can you provide one or more examples of initial conditions here?

We will add this to the manuscript.

Suggested change to manuscript:

- "such as soil moisture"

5. Page 4602, Line 25: I suggest you briefly define "crop failure" at some point in the introduction so that readers better follow this section.

We thank the reviewer for drawing our attention to this issue and we have added an explanation of crop failure where it occurs.

Suggested change to manuscript:

– Change "reliability of global-scale seasonal crop failure hindcasts, Iizumi et al. (2014)" to "reliability of hindcasts of global-scale yield decreases of at least 5%, Iizumi et al. (2014)"

– Change "Nicklin et al. (2011) found some positive skill in reproducing crop failure of groundnut in West Africa with GLAM driven by seasonal forecast data" to "Nicklin et al. (2011) found some positive skill in reproducing both severe crop failure (yields below 10th percentile of climatology) and less severe crop failure (yields below the 25th percentile of climatology) of groundnut in West Africa with GLAM driven by seasonal forecast data".

– Change "As part of the EU FP7 project EUPORIAS (Hewitt et al., 2013), JULES-crop will be driven by seasonal forecasts and its ability to produce probabilistic forecasts of crop failure will be investigated." to "As part of the EU FP7 project EUPORIAS (Hewitt et al., 2013), JULES-crop will be driven by seasonal forecasts and its ability to produce probabilistic forecasts of crop yield will be investigated."

6. There are several parts in the one to the last paragraph of the introduction section that, in my view, requires serious clarification, specifically because, as stated earlier, this is one of the important questions addressed by the paper:

* Is there any difference between start date of the seasonal forecast runs and JULES-crop runs? And how are they different from sowing date (is sowing date assumed to be equivalent to the beginning of the crop growth season)?

* What does "run" exactly refer to in "if the run is started on the sowing date"?

* What does it mean by "Starting the run before or on the sowing date means that the initialisation of crop variables (e.g. height) is trivial"?

Data from seasonal models is usually available at various start dates throughout the year. The advantage of starting a JULES-crop model run (forced by seasonal model output) on or before the crop sowing date is this makes the initialisation considerably easier. Once the crops have been sown, the initialisation will need to include the state of the crops e.g. their development index, height, leaf area index, and carbon in the root, harvest and stem reserve pools (see table 1). It is therefore easier to initialise the crops before they are sown, so that the development index is -2 and the height, leaf area index, and carbon in the root, harvest and stem reserve pools are set negligibly small to represent the fact that the crop does not exist yet. "run" refers to a JULES-crop model run. We will amend this in the text.

Suggested change to manuscript:

- Add the text "since the crop either does not yet exist, or only exists as a seed."
- "if the run is started on the sowing date" to "if the JULES-crop model run is started on the sowing date"

7. Page 4604, Line 5: I suggest not using "Sect." as a short from of "section".

We will implement this.

8. Page 4606, Line 9: You never identified the "main run" in "The run was from 1960 to 2009 and spun up by repeating the first 10 years five times, before starting the 10 main run". So does it the main run is from 1970 to 2009? Also, is it necessary to run the first 10 years five times? If so, please elaborate.

The long spin up removes the dependence on the initialisation variables (this spinup would need to be longer if we were interested in e.g. soil carbon pools, but this is not relevant to this study.). We will explicitly give the duration of the main run in the text.

Suggested change to manuscript:

– change "The run was from 1960 to 2009 and spun up by repeating the first 10 years five times, before starting the main run". to "The main run was from 1960 to 2009. The initialisation variables were taken from CRU-NCEPv-forced run with the crop model switched off and the model was spun up by repeating the first 10 years five times, before starting the main run, in order to remove the sensitivity to this initialisation."

9. Section 3.2: Is there any implicit assumption that "there is no constraint on water availability"?

Yes, irrigation in JULES is currently unconstrained by water availability. Thank you for pointing out this omission - we will clarify this in the text.

Suggested change to manuscript:

– Add the text "with no constraint on water availability"

10. Page 4608, Paragraph 2: What are the correlations between different scenarios and the baseline (I think you have used FAO reported yield values)?

The correlations between runs with and without approximations are given in table 2. As described above, we have additionally calculated the correlation between time series of model yield and FAO country yield, to check that these support our conclusions (see Figure 1 and Figure 2).

11. Page 4608, Paragraph 2: Related to your discussion about Brazil, so are you saying that using daily data followed by model disaggregation in the model outperforms using 6-hour data in the model in Brazil?

Yes, although (as we state in the manuscript) we expect this to be the result of additional degrees of freedom in the disaggregator compensating for a bias in another part of the model (Williams 2014 has a more detailed description of this issue). The disaggregator in

JULES is an area where further development is needed (this is currently in progress in the JULES community).

12. Page 4609, Paragraph 1: In the second set-up, based on your argument here, can we say that removing soil moisture stress from the model is equivalent to no water availability (either green or blue water) constraint? Then, I am wondering what the relevance and importance of this set-up would be. Does it provide the readers with any interesting insights? To what extent did you used the insights in you discussion/conclusion?

This configuration is useful for understanding the results shown by the other set-ups, by definitively showing the regions where the interannual variability in yield is predominantly due to the response to soil moisture variability and where other factors are more influential. It demonstrates the most important mechanism by which precipition can influence the crops, which is vital for understanding the success of the other approximations proposed.

Suggested change to manuscript:

- Add the text "which we have confirmed with a fully irrigated run" to the conclusion section to emphasize this point.

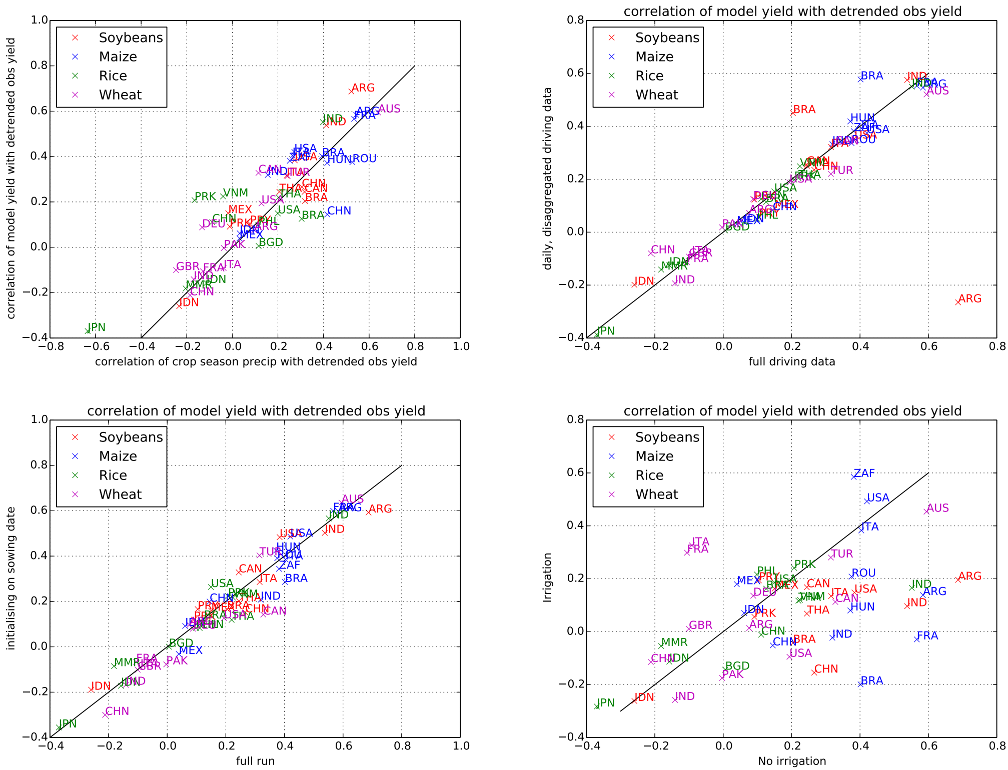

Soybeans
Maize
Rice
Wheat
correlation of model yield with detrended obs yield
correlation of crop season precip with detrended obs yield
daily, disaggregated driving data
full driving data
initialising on sowing date
full run
Irrigation
No irrigation

**Figure 1.** X-axis of all plots shows the correlation coefficient for FAO country yields and yields from the `control` run. Y-axis shows the correlation coefficient for FAO country yields and crop season precipitation (top left), and yield in the `disagg` (top right), `init` (bottom left) and `irrig` runs (bottom right). Red, blue, green and purple is soybeans, maize, rice and wheat respectively and each country is labelled by its 3 letter ISO abbreviations. The result for Argentina is sensitive to the yield threshold imposed on the model yield, which changes the number of gridboxes included in the calculation (Argentina is one of the region where the representation of soybeans in the model needs to be improved, particularly since it is one of the mayor producers of soybeans).

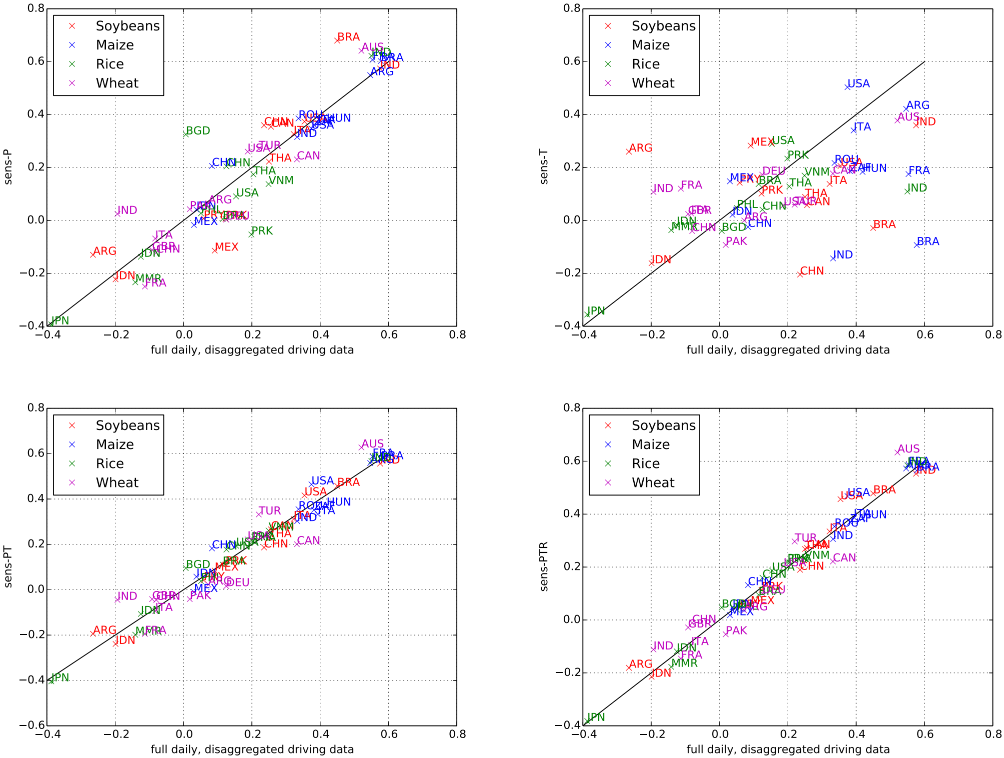



**Figure 2.** X-axis of all plots shows the correlation coefficient for FAO country yields and yields from the `disagg` run. Y-axis shows the correlation coefficient for FAO country yields and yield in the `sens−P` (top left), `sens−TP` (top right), `sens−T` (bottom left) and `sens−TPR` runs (bottom right). Red, blue, green and purple is soybeans, maize, rice and wheat respectively and each country is labelled by its 3 letter ISO abbreviations.

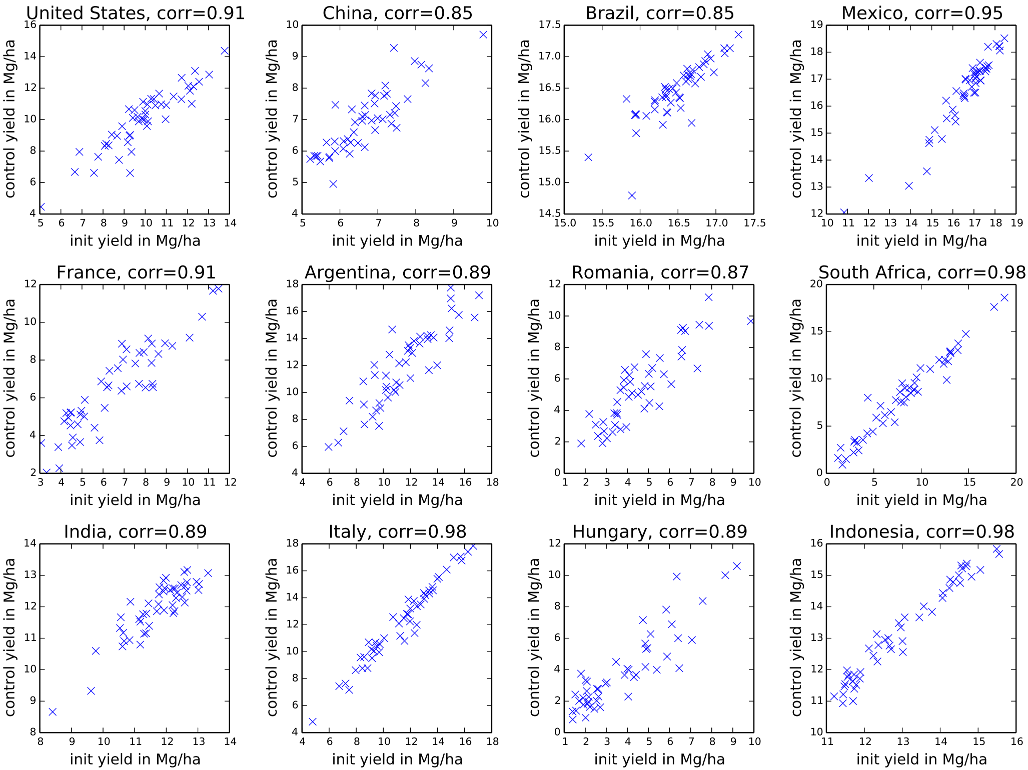



**Figure 3.** Scatter plot of yield from the control run against yield from the init run.

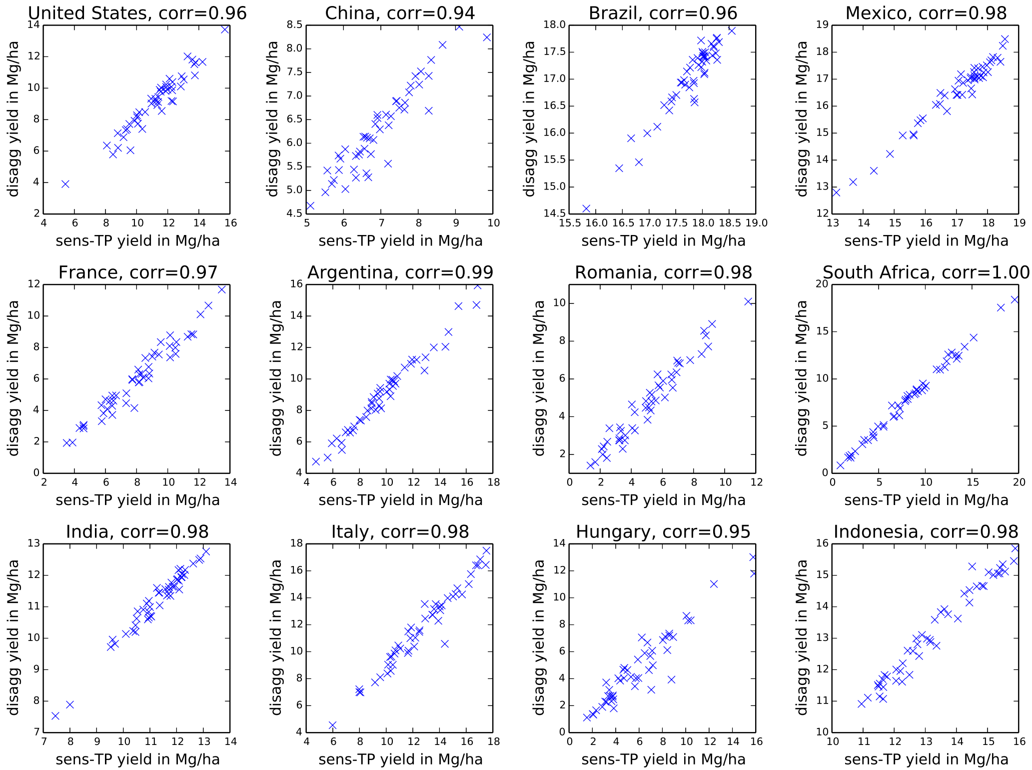



**Figure 4.** Scatter plot of yield from the disagg run against yield from the sens-TP run.

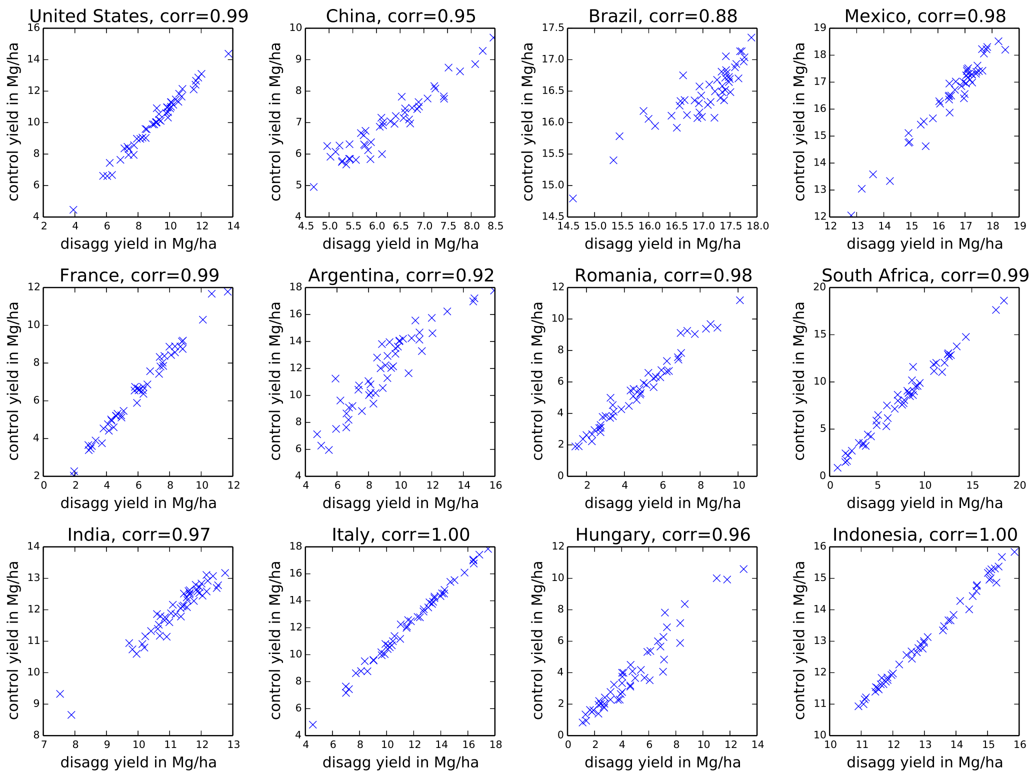



**Figure 5.** Scatter plot of yield from the control run against yield from the disagg run.

**Figure 6.** Scatter plots comparing the global mean maize yield from different model runs.

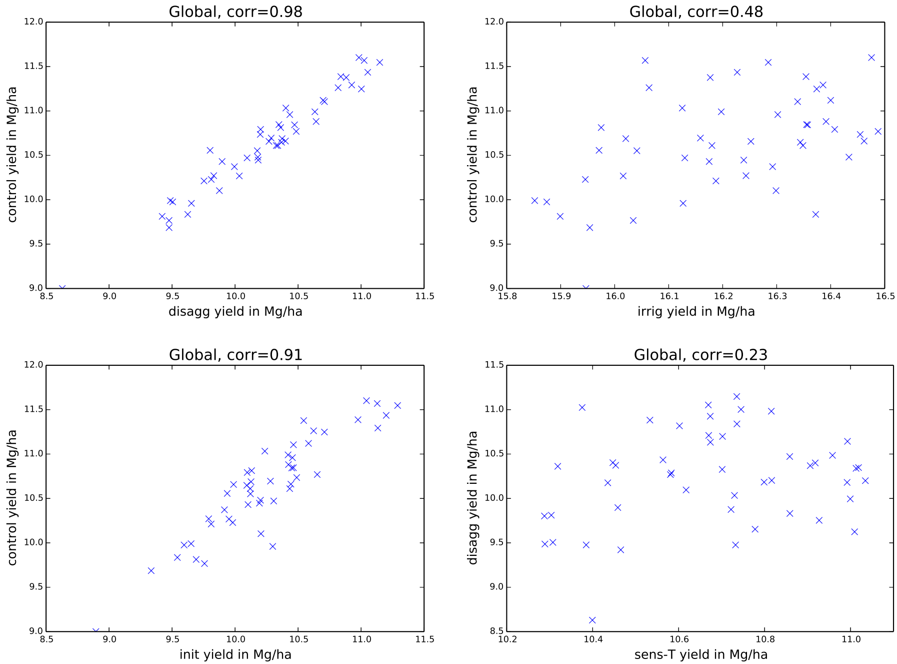
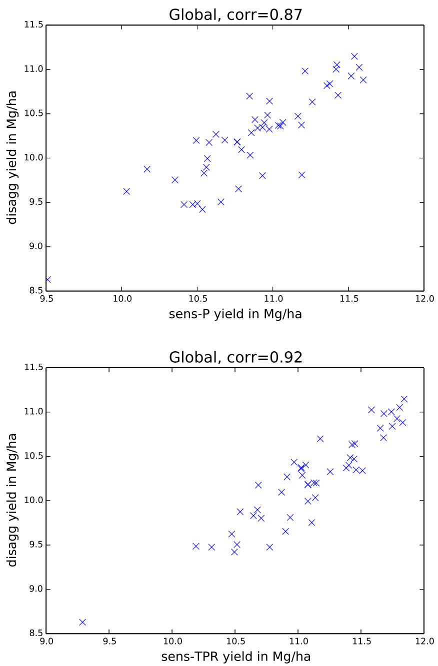


**Figure 7.** Scatter plots comparing the global mean maize yield from different model runs.



## 3 To be included as Supplementary Material

| name | mean | standard deviation | global corr with control | global corr with disagg |
|---|---|---|---|---|
| | | maize | | |
| control | 10.6 | 0.55 | | |
| irrig | 16.2 | 0.18 | 0.48 | |
| init | 10.3 | 0.48 | 0.91 | |
| disagg | 10.2 | 0.53 | 0.98 | |
| sens-T | 10.7 | 0.23 | | 0.23 |
| sens-P | 10.9 | 0.42 | | 0.87 |
| sens-TP | 11.1 | 0.51 | | 0.92 |
| sens-TPR | 11.1 | 0.50 | | 0.92 |
| sens-TPW | 10.3 | 0.52 | | 0.96 |
| | | soybean | | |
| control | 6.23 | 0.38 | | |
| irrig | 8.22 | 0.31 | 0.44 | |
| init | 6.06 | 0.44 | 0.86 | |
| disagg | 6.22 | 0.44 | 0.82 | |
| sens-T | 6.61 | 0.30 | | 0.56 |
| sens-P | 6.36 | 0.42 | | 0.54 |
| sens-TP | 6.54 | 0.43 | | 0.90 |
| sens-TPR | 6.48 | 0.43 | | 0.89 |
| sens-TPW | 6.32 | 0.42 | | 0.97 |
| | | rice | | |
| control | 7.14 | 0.23 | | |
| irrig | 7.91 | 0.23 | 0.80 | |
| init | 7.14 | 0.21 | 0.98 | |
| disagg | 6.52 | 0.24 | 0.98 | |
| sens-T | 6.72 | 0.17 | | 0.75 |
| sens-P | 6.62 | 0.15 | | 0.48 |
| sens-TP | 6.64 | 0.20 | | 0.98 |
| sens-TPR | 6.59 | 0.21 | | 0.98 |
| sens-TPW | 6.62 | 0.21 | | 0.99 |
| | | spring wheat | | |
| control | 4.75 | 0.34 | | |
| irrig | 8.29 | 0.10 | 0.19 | |
| init | 4.01 | 0.27 | 0.86 | |
| disagg | 4.91 | 0.43 | 0.89 | |
| sens-T | 4.06 | 0.12 | | 0.56 |
| sens-P | 5.25 | 0.26 | | 0.81 |
| sens-TP | 5.30 | 0.36 | | 0.95 |
| sens-TPR | 5.33 | 0.38 | | 0.96 |
| sens-TPW | 4.99 | 0.40 | | 0.96 |

**Table 2.** Results from the global runs described in Section 3 of the manuscript. First column is the run name, second is the mean crop yield in $\mathrm{Mg\,ha^{-1}}$, third is the standard deviation of the annual global yield time series in $\mathrm{Mg\,ha^{-1}}$. The fourth column gives the Pearson correlation coefficient with the global yield in the `control` run and the fifth column gives the Pearson correlation coefficient with the global yield in the `disagg` run. All results have been weighted as described in Section 4 of the manuscript.

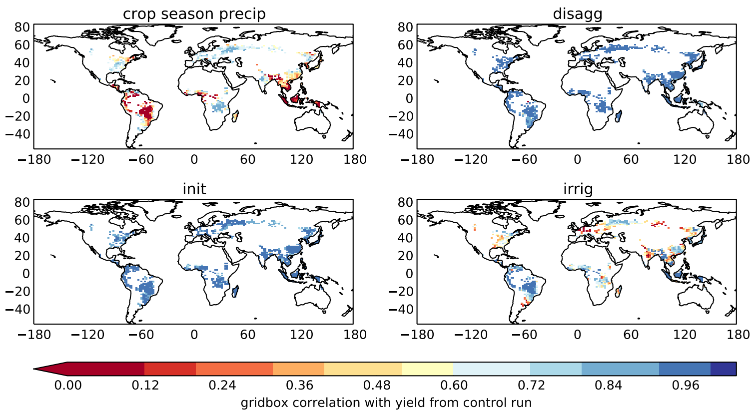



**Figure 8.** All plots show the correlations with the annual soybean yield in the `control` run for each gridbox. Top left: the correlation between yield in `control` run and crop season precipitation. Top right, bottom left and bottom right: the correlation between yield in `control` run and yield in the `disagg`, `init` and `irrig` runs, respectively.

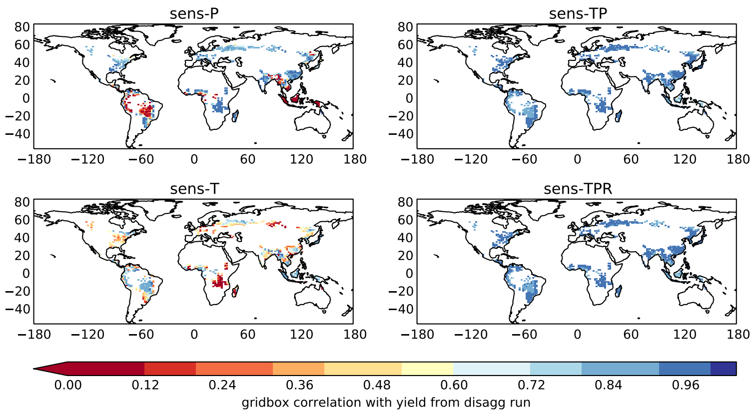



**Figure 9.** The correlations between the annual soybean yield in the `control` run and the annual soybean yield from the `sens−P` (top left), `sens−TP` (top right), `sens−T` (bottom left) and `sens−TPR` (bottom right) runs for each gridbox.

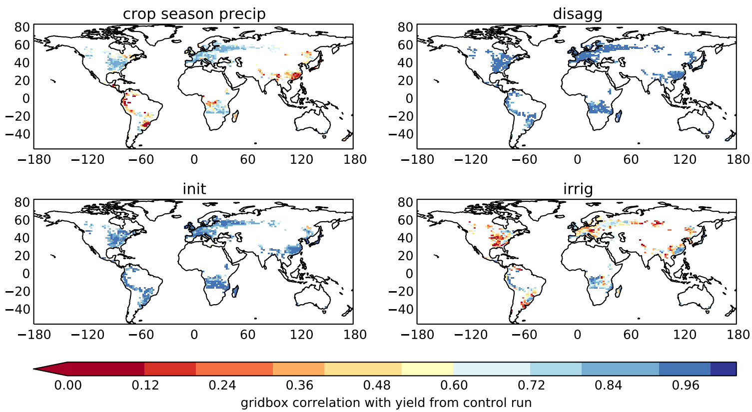



**Figure 10.** All plots show the correlations with the annual spring wheat yield in the `control` run for each gridbox. Top left: the correlation between yield in `control` run and crop season precipitation. Top right, bottom left and bottom right: the correlation between yield in `control` run and yield in the `disagg`, `init` and `irrig` runs, respectively.

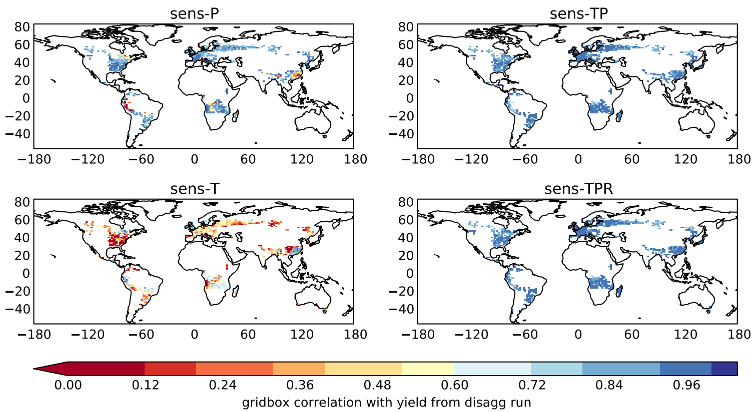



**Figure 11.** The correlations between the annual spring wheat yield in the `control` run and the annual spring wheat yield from the `sens-P` (top left), `sens-TP` (top right), `sens-T` (bottom left) and `sens-TPR` (bottom right) runs for each gridbox.

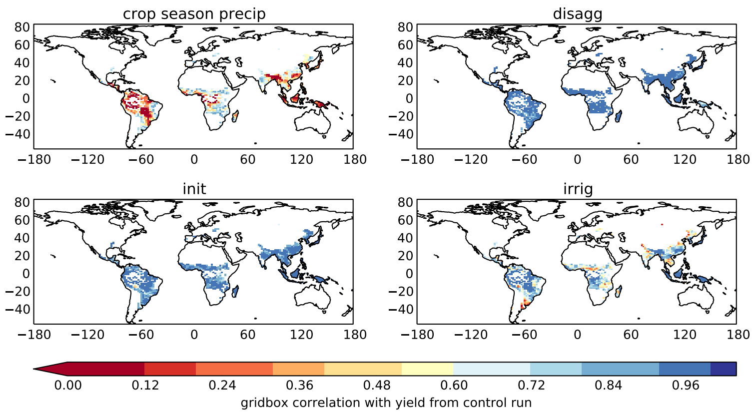



**Figure 12.** All plots show the correlations with the annual rice yield in the `control` run for each gridbox. Top left: the correlation between yield in `control` run and crop season precipitation. Top right, bottom left and bottom right: the correlation between yield in `control` run and yield in the `disagg`, `init` and `irrig` runs, respectively.

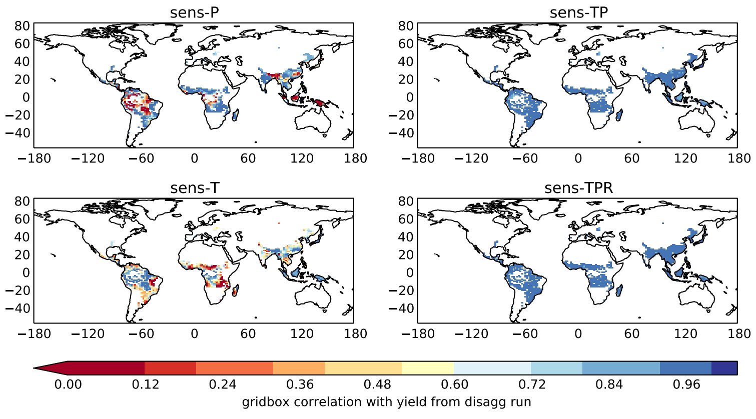

sens-P
sens-TP
sens-T
sens-TPR
80
60
40
20
0
−20
−40
−180
−120
−60
0
60
120
180
0.00
0.12
0.24
0.36
0.48
0.60
0.72
0.84
0.96
gridbox correlation with yield from disagg run

**Figure 13.** The correlations between the annual rice yield in the `control` run and the annual rice yield from the `sens-P` (top left), `sens-TP` (top right), `sens-T` (bottom left) and `sens-TPR` (bottom right) runs for each gridbox.

## References


Cantelaube, P. and Terres, j.-M.: Seasonal weather forecasts for crop yield modelling in Europe, Tellus A, 57, 476–487, doi:10.1111/j.1600-0870.2005.00125.x, http://dx.doi.org/10.1111/j.1600-0870.2005.00125.x, 2005.

Collatz, Ball, Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agricultural and Forest Meteorology, 54, 107–136, doi:10.1016/0168-1923(91)90002-8, http://dx.doi.org/10.1016/0168-1923(91)90002-8, 1991.

Collatz, G. J., Ribas-Carbo, M., and Berry, J. A.: Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants, Australian Journal of Plant Physiology, 19, 519–538, http://www.publish.csiro.au/?paper=PP9920519, 1992.

Hewitt, C., Buontempo, C., and Newton, P.: Using Climate Predictions to Better Serve Society's Needs, Eos Trans. AGU, 94, 105–107, doi:10.1002/2013eo110002, http://dx.doi.org/10.1002/2013eo110002, 2013.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanenkov, V., Oettli, P., Newby, T., Ishigooka, Y., and Furuya, J.: Historical changes in global yields: major cereal and legume crops from 1982 to 2006, Global Ecology and Biogeography, 23, 346–357, doi:10.1111/geb.12120, http://dx.doi.org/10.1111/geb.12120, 2014.

Monfreda, C., Ramankutty, N., and Foley, J. A.: Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochem. Cycles, 22, GB1022+, doi:10.1029/2007gb002947, http://dx.doi.org/10.1029/2007gb002947, 2008.

Nicklin, K. J., Challinor, A., and Tompkins, A.: The use of seasonal forecasts in a crop failure early warning system for West Africa, AGU Fall Meeting Abstracts, pp. A1007+, http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2011AGUFMGC31A1007N, 2011.

Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., and Wheeler, T.: JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator, Geoscientific Model Development, 8, 1139–1155, doi:10.5194/gmd-8-1139-2015, http://dx.doi.org/10.5194/gmd-8-1139-2015, 2015.

Palmer, T. N., Doblas-Reyes, F. J., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., Feddersen, H., Cantelaube, P., Terres, J. M., Davey, M., Graham, R., Délécluse, P., Lazar, A., Déqué, M., Guérémy, J. F., Díez, E., Orfila, B., Hoshen, M., Morse, A. P., Keenlyside, N., Latif, M., Maisonnave, E., Rogel, P., Marletto, V., and Thomson, M. C.: Development Of A European Multimodel Ensemble System For Seasonal-to-interannual Prediction (DEMETER), Bull. Amer. Meteor. Soc., 85, 853–872, doi:10.1175/bams-85-6-853, http://dx.doi.org/10.1175/bams-85-6-853, 2004.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41, 2014GL059 637–2519, doi:10.1002/2014gl059637, http://dx.doi.org/10.1002/2014gl059637, 2014.

Weedon, G. P., Prudhomme, C., Crooks, S., Ellis, R. J., Folwell, S. S., and Best, M. J.: Evaluating the Performance of Hydrological Models via Cross-Spectral Analysis: Case Study of the Thames Basin, United Kingdom, J. Hydrometeor, 16, 214–231, doi:10.1175/jhm-d-14-0021.1, http://dx.doi.org/10.1175/jhm-d-14-0021.1, 2014.

Williams, K. and Clark, D.: Disaggregation of daily data in JULES, Tech. Rep. 96, Met Office Hadley Centre, http://www.metoffice.gov.uk/media/pdf/2/j/HCTN96.pdf, 2014.

Manuscript prepared for Geosci. Model Dev. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 26 October 2015

# Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts

**Karina Williams**[1] **and Pete Falloon**[1]

[1]Met Office Hadley Centre, FitzRoy Road, Exeter, Devon EX1 3PB, UK

Correspondence to: Karina Williams (karina.williams@metoffice.gov.uk)

## Abstract

JULES-crop is a parametrisation of crops in the Joint UK Land Environment Simulator (JULES). We investigate the sources of the interannual variability in the modelled maize yield, using global runs driven by reanalysis data, with a view to understanding the impact of various approximations in the driving data and initialisation. The standard forcing dataset for JULES consists of a combination of meteorological variables describing precipitation, radiation, temperature, pressure, specific humidity and wind, at subdaily time resolution. We find that the main characteristics of the modelled yield can be reproduced with a subset of these variables and using daily forcing, with internal disaggregation to the model timestep. This has implications in particular for the use of the model with seasonal forcing data, which may not have been provided at subdaily resolution for all required driving variables. We also investigate the effect on annual yield of initialising the model with climatology on the sowing date. This approximation has the potential to considerably simplify the use of the model with seasonal forecasts, since obtaining observations or reanalysis output for all the initialisation variables required by JULES for the start date of the seasonal forecast would present significant practical challenges.


## 1 Introduction

The ability to forecast crop yield on a seasonal timescale has significant economic and humanitarian benefits (Hansen et al., 2006; Iizumi et al., 2014; Mishra et al., 2008). Climate variability and extremes can have significant impacts on crops (e.g. Challinor et al. (2014)) and improvements in the seasonal forecast of meteorological ~~meteological~~ variables such as temperature and rainfall (Molteni et al., 2011; MacLachlan et al., 2015; Manzanas et al., 2014) therefore have the potential to improve yield forecasts. However, existing studies of crop model performance focused on seasonal forecast applications show considerable variation in skill depending on the region, scale, processes and crops involved (Hansen et al., 2011; Dessai and Bruno Soares, 2013; Falloon et al., 2013). Crop model simulations driven by statistically downscaled seasonal hindcasts for European wheat (Palmer et al., 2004; Cantelaube and Terres, 2005), and specifically for wheat in Italy (Marletto et al., 2007) showed that reliable crop yield predictions could be produced using an ensemble multi-model approach and the JRC crop model, for instance, estimating a high probability of a positive yield anomaly in 1996 and a negative yield anomaly 1998 in the UK, consistent with observations. Similarly, Coelho and Costa (2010) used an ensemble of bias-corrected and disaggregated seasonal forecasts to simulate maize yields over Southern Brazil, with the GLAM crop model. The model showed generally good agreement with observations, with observed yields within the 95% forecast interval for most years. Using a statistical approach to assess the reliability of hindcasts of global-scale yield decreases of at least 5%, Iizumi et al. (2013) ~~seasonal crop failure hindcasts ,~~ found that within-season hindcasts with lead times of 1-3 months generally reproduced inter-annual variability in observed yields in major wheat exporting countries ~~($r^2$=0.56-0.61)~~ better than pre-season hindcasts with lead times of 3-5 months~~($r^2$=0.43-0.59)~~. Iizumi et al. (2014) modelled global yields of major crops by combining satellite derived NPP data and global agricultural datasets for crop calendar, harvested area and country yield statistics. This statistical model mostly performed well compared to observations, with modelled yields explaining 45-81% of the spatial variation of observed yields in 2000, and correlation coefficients between modelled yield time series and sub-national yield statistics for 1982-2006 in major crop-producing regions generally greater than 0.8. Nicklin et al. (2011) found some positive skill in reproducing both severe crop failure (yields below 10th percentile of climatology) and less severe crop failure (yields below the 25th percentile of climatology) of ~~crop failure of~~ groundnut in West Africa with GLAM driven by seasonal forecast data, and found that these results were relatively independent of assumptions on the varieties of groundnut modelled. Mishra et al. (2008) ran the SARRA-H crop model at five locations in Burkina Faso, showing ~~and found~~ that, in most cases, incorporating seasonal rainfall forecasts improved sorghum yield predictions made early in the season.

Palmer et al. (2004) and Cantelaube and Terres (2005) also found that downscaling seasonal hindcasts improved crop model performance - the $r^2$ value of simulated biomass for the whole of Europe increased from 0.62 to 0.69 with greater regional improvements when downscaled seasonal forecasts were used instead of the original, pre-downscaled versions. On the other hand Challinor et al. (2005) found that bias correction of GCM-derived seasonal hindcasts data had generally small effects for simulation of groundnut yields in India. Watson and Challinor (2013) found that errors in rainfall data had the largest impact on crop model skill for groundnut in India, mainly because the study region was rainfall limited, while generally the largest yield errors were caused by errors in inter-annual variability in temperature and precipitation. In contrast, for French maize, temperature errors had a stronger influence on yield estimates from both a statistical model and a process-based model than precipitation (Watson et al., 2014).

The ability of crop models to represent inter-annual effects of climate variables also varies depending on the processes represented in the models (Falloon et al., 2014b). For example, high temperature stress around anthesis (the onset of flowering) can have strong impacts on crop yields but not all models include this effect, and responses vary across models that do (Asseng et al., 2013). In general, there is little information of the role of initial conditions such as soil moisture on ~~in~~ crop model performance on seasonal timescales (Falloon et al., 2013), although hydrological studies have shown that different spin-up approaches may be needed for different impacts (Cosgrove et al., 2003) and different regions.

The JULES-crop model (Osborne et al., 2015) was developed with the dual aim of being able to simulate the impact of weather and climate on crop productivity and the impact that crop-lands have on weather and climate. It is a component of the Joint UK Land Environment Simulator (JULES) (Best et al., 2011; Clark et al., 2011), which is a community land surface model that can be used both online as part of the Met Office Unified Modelling system and offline for impacts studies. As part of the EU FP7 project EUPORIAS (Hewitt et al., 2013), JULES-crop will be driven by seasonal forecasts and its ability to produce probabilistic forecasts of crop yield ~~failure~~ will be investigated. EUPORIAS (European Provision Of Regional Impacts Assessments on Seasonal and Decadal Timescales) aims to maximise the societal benefit of seasonal and decadal forecasts by making the predictions directly relevant to decision-makers. As part of this project, a multi-model ensemble of seasonal meteorological forecasts will be used to drive an ensemble of impacts models, including JULES-crop.

However, using JULES-crop on a seasonal timescale introduces a number of technical and scientific issues. The aim of this paper is to address those issues that ~~Many of these~~ are centred around the availability of data, by investigating to what extent the interannual variability of the modelled yield can be captured if some of these data requirements are relaxed.

The first data availability issue concerns the driving data. ~~.~~ JULES is driven by a combination of meteorological variables describing air temperature, precipitation, radiation, wind speed, humidity and pressure (for a full description, see the JULES User Guide, available at https://jules.jchmr.org/[1]) for each grid box in the model domain, ideally at sub-daily resolution. Output in this format for each ensemble member requires a large amount of storage space and is typically not made externally available by seasonal forecast centres. It is therefore useful to investigate whether the yield variability can be modelled sufficiently well if only a subset of the forcing variables are taken from the seasonal forecast and the others set to climatology, or if the model is forced with daily meteorological data and disaggregated internally to the model timestep. To gain a better understanding of the dependence of the yield on the different forcing variables, we look at the effect of removing water stress and the correlation of the yield with the total gridbox precipitation during the crop growing season.

The second data availability issue concerns the variables required to initialise the JULES-crop runs, such as the moisture content of each soil layer (as a fraction of the water content at saturation). Obtaining accurate values for these variables on the start date of the seasonal forecast runs would present a significant practical challenge, as recent observations would be required to estimate these values directly or as input to a reanalysis run. Therefore, we investigate the loss in predictability of yield if the JULES-crop model run is started on the sowing date of the crop in that gridbox and initialised by the climatological values for that date. This set-up would be simple to reproduce with seasonal forecast forcing that has

[1] ~~available at https://jules.jchmr.org/~~

been bias corrected to a reanalysis dataset, such as those available as part of EUPORIAS, since JULES-crop can be run with this reanalysis dataset to produce a climatology of the initialisation variables. Starting the run before or on the sowing date means that the initialisation of crop variables (e.g. height) is trivial since the crop either does not yet exist, or only exists as a seed. It has also been suggested that the initialisation of impact model runs driven by seasonal forecasts is more critical for some impacts and regions than others, for example, it may be more critical for water resources in cold regions where snow stores are important than for ~~to~~ dry land cropping (Falloon et al., 2014a).

It is important to note that while this study provides a practical methodology for driving JULES-crop with seasonal forecasts, given commonly available forcing and initialisation data, there are many aspects of the uncertainty chain that remain to be addressed. For example, once an application has been identified (e.g. a decision threshold based on the yield of a particular crop in a particular region), a thorough validation would need to be performed of the relevant model diagnostic against observational data and against hindcast driven runs.

This paper is organised as follows. Section 2 describes the JULES-crop model and how it interacts with the other JULES components, Section 3 describes the model set-up used for the runs presented in this paper, Section 4 presents the results and Section 5 draws conclusions from these runs about the model behaviour and sensitivities and how these can inform the design of JULES-crop runs forced with seasonal forecasts.

## 2 Model description

JULES is a process-based model that simulates fluxes of carbon, water, energy and momentum between the land surface and the atmosphere. Sub-grid heterogeneity ~~hetrogeneity~~ is represented through tiles of various surface types, such as broadleaf trees, bare soil and C3 grass. As of JULES version 4.0, it includes a crop parametrisation (JULES-crop) which introduces an additional tile for each crop simulated. We refer the reader to Best et al. (2011); Clark et al. (2011) for a fuller description of JULES and Osborne et al. (2015) for a description of JULES-crop in particular, and focus here on features that are particularly relevant to this article, such as the influence of temperature on crop growth stage, soil moisture on photosynthesis and the partitioning of carbon into different parts of the plant.

~~is possible to create additional tiles to represent crops~~[1] ~~.~~ The status of development of the crop on each tile is parametrised by the crop development index (DVI), which is -2 before sowing, -1 at sowing, 0 at emergence and 1 at flowering. Under normal conditions, harvest occurs at a DVI of 2. The progression between the development stages is determined by crop-specific thermal time parameters, set by the user. For the purposes of this paper, thermal time is an accumulation of effective temperature between one development stage and the next (since we do not include a photoperiod dependence). Effective temperature is defined by

$$T_{eff} = \begin{cases} 0 & \text{for} \quad T < T_b \\ T - T_b & \text{for} \quad T_b \leq T \leq T_o \\ (T_o - T_b)(1 - \dfrac{T - T_o}{T_m - T_o}) & \text{for} \quad T_o < T < T_m \\ 0 & \text{for} \quad T \geq T_m \end{cases}, \tag{1}$$

where $T$ is the air temperature of the tile at that timestep and $T_b$, $T_o$ and $T_m$ are crop-specific cardinal temperatures.

Potential leaf-level photosynthesis (unstressed by water availability and ozone effects) is calculated as the smoothed minimum of three potentially limiting rates, based on Collatz et al. (1991, 1992): (a) the Rubisco-limited rate, which depends on the maximum rate of carboxylation of Rubisco, (b) the light-limited rate and (c) the rate associated with the transport of photosynthetic products for C3 plants or PEP-Carboxylase limitation for C4 plants. The vertical profile of radiation through the canopy can use either the big-leaf

[1] ~~See for a fuller description of JULES and for JULES-crop in particular. We focus here on features that are particularly relevant to this article, such as the influence of temperature and soil moisture on crop growth.~~

approach (following Beer's law) or a multi-layered canopy radiation scheme, which treats the direct and diffuse components of the radiation separately. The latter can optionally include the direct component of the direct beam radiation ('sunflecks'). The potential leaf-level photosynthesis is scaled by a soil water factor $\beta$, to account for soil moisture stress. This factor is zero when the mean soil moisture content in the root zone $\theta$ is less than or equal to a wilting point concentration $\theta_w$, 1 when $\theta$ is greater than the critical concentration $\theta_c$ and linearly increasing in between (i.e. a slant step function). As of JULES version 4.1, it is possible to irrigate part of each gridbox, which involves adding water to the soil until $\beta = 1$ during certain times of the year.

Net ~~Plant growth is modelled by integrating the net~~ primary productivity (NPP) is calculated by scaling the leaf-level photosynthesis to the canopy level and subtracting plant maintenance and growth respiration. Crop growth is modelled by integrating NPP over the course of a day and splitting this carbon between the crop root, stem, leaf, harvest and stem reserve carbon pools for that tile ($C_{root}$, $C_{leaf}$, $C_{stem}$, $C_{harv}$, $C_{resv}$ respectively~~resp.~~). The proportion of carbon given to each pool depends on the DVI of the crop and the crop type.

Once the proportion of carbon given to the stem pool drops below 0.01, carbon from the stem reserve pool is mobilised to the harvest pool, by reducing $C_{resv}$ by 10% each day and adding this carbon to the harvest pool. Similarly, once the DVI is above 1.5, carbon from the leaf pool is mobilised to the harvest pool, by reducing $C_{leaf}$ by 5% each day and adding this carbon to $C_{harv}$, to simulate leaf senescence. At harvest, the carbon in the harvest pool becomes yield and each crop carbon pool is reset.

~~Within JULES, the leaf-level photosynthesis is scaled by a soil water factor $\beta$, to account for soil moisture stress. This factor is zero when the mean soil moisture content in the root zone $\theta$ is less than or equal to a wilting point concentration $\theta_w$, 1 when $\theta$ is greater than the critical concentration $\theta_c$ and linearly increasing in between (i.e. a slant step function). As of JULES version 4.1, it is possible to irrigate part of each gridbox, which involves adding water to the soil until $\beta = 1$ during certain times of the year.~~

The model does not include a way of calibrating against yield observations (e.g. a yield gap parameter which accounts for the impact of pests, diseases and non-optimal management on the crop yield). Therefore the outputted yield is the water-limited potential yield when irrigation is switched off and the potential yield when the crop is fully irrigated.

## 3 Experimental Set-up

All runs were performed with JULES 4.2.

### 3.1 Control run (`control`)

The experimental set-up for the `control` run follows the global set-up in Osborne et al. (2015). The `control` run was forced by 6 hourly CRU-NCEPv4 climate data ~~(extended to include 2012)~~ as used by the Global Carbon Project (Le Quéré et al., 2014), regridded to a n96 grid (i.e. gridboxes are 1.875 degrees by 1.25 degrees). The main run was from 1960 to 2009. The initialisation variables were taken from a CRU-NCEPv-forced run with the crop model switched off and the model was ~~2009 and~~ spun up by repeating the first 10 years five times, before starting the main run, in order to remove the sensitivity to this initialisation. Wheat, soybean, maize and rice were modelled, with the crop parameters listed in Osborne et al. (2015). A multi-layer canopy radiation scheme was used, which accounts for direct/diffuse radiation components including sun-flecks (`can_ran_mod`=5). The crop sowing dates were taken from Sacks et al. (2010) and extended using nearest neighbour interpolation. The crop tile fractions were taken from Monfreda et al. (2008) and other ancillaries taken from HadGEM2-ES (Collins et al., 2011; Jones et al., 2011). Irrigation was not switched on.

### 3.2 Fully irrigated run (`irrig`)

We repeated the `control` run with irrigation demand switched on, such that, when one of the crops on the gridbox had DVI$>$-1, water was added to the top two soil levels until the critical soil moisture content $\theta_c$ was reached, so that the soil water factor $\beta$ was 1, with no constraint on water availability. The run was initialised and spun up in the same way as the `control` ~~1. Each run repeated the first 10 years five times, to spin up , before starting the main~~ run.

### 3.3 Full disaggregated run (`disagg`)

We created daily means and daily temperature ranges from the CRU-NCEPv4 driving data, and used this to drive a JULES run. The internal JULES disaggregator (described in Williams and Clark (2014)) was used to disaggregate this forcing data to the internal model time step of one hour. For temperature, this involves adding a sinusoidal diurnal cycle. Precipitation in a day is modelled as occurring in one rainfall event of constant intensity, with a duration that depends on the precipitation type. ~~.~~ The run was initialised and spun up in the same way as the `control` run~~with the dump file from the beginning of the control main run and then spun up by repeating the first ten years five times~~. All other settings were the same as the `control` run.

### 3.4 Disaggregated runs with some forcing from climatology (`sens-*`)

In order to investigate the sensitivity to variability in different parts of the driving data, we created daily climatologies of each driving data variable in the full disaggregated run. For example, for each gridbox, the value used for the 1st January in the precipitation climatology was the mean over the CRU-NCEP precipitation on every 1st January from 1960 to 2009 in that gridbox. We then repeated the runs (for 1960 to 2009, as before) with climatological driving data for all variables apart from certain combinations. The combinations we refer to in this paper are shown in Table 1. The run was initialised and spun up in the same way as the

~~Each run had 50 years of spin up (first 10 years five times) before starting the main run (this was particularly important for the~~ `control`~~`sens-T`~~ run. ~~).~~

### 3.5 Runs initialised from climatology (`init`)

We created a climatology for each initialisation variable, ~~the initialisation variables~~ for each day of the year, using daily means outputted from the `control` run and averaging over 1960-2009. The model requires 16 initialisation variables, on multiple model layers or tiles, such as tile surface temperature and moisture in soil layers as a fraction of water content at saturation (see JULES user guide for full list). The model domain was split by sowing date and we performed a separate run for each sowing date for each crop for each year, initialised by the climatology for that sowing date, without spin up. For example, for Maize, we modelled 77 different sowing dates across the globe for 48 years, which involved $77\times48$ individual JULES runs. The ~~Each run lasted 1 year and the~~ full 6 hourly driving data was used. Each run lasted one year and the annual yields were concatenated to get a 48 year time series for each crop in each gridbox.

## 4 Results

Global time series for each crop were constructed from the model output by first masking any gridboxes which had one or more years in which the the crop did not reach a DVI of 1.5 or greater or had a yield less than the seed carbon 0.01 $\mathrm{kg\,C\,m^{-2}}$ (which we assumed was due to a failure on the part of the model or model settings to represent the crops in this gridbox) and then weighting according to grid box size and crop tile fraction. We define a year as January 1st to December 31st (i.e. the model year). In a small fraction of the gridboxes with harvest dates around the end of December/beginning of January, this definition caused issues, as two harvests could fall in one year and none in the next. These points were masked out, as the zero yield appears as a model failure. Osborne et al. (2015) [1]~~.~~ found that maize yield in the `control` run had the highest correlation with detrended global FAO yield observations out of the four crop types modelled (maize, soybean, rice

[1] ~~A year is defined as January 1st to December 31st (i.e. the model year). In a small fraction of the gridboxes with harvest dates around the end of December/beginning of January, this definition~~

and wheat); therefore we will explicitly discuss ~~present~~ the results for maize only, although we have confirmed that our overall conclusions apply to each of the four crops individually. Results from the other crops are given in the supplementary material.

~~caused issues, as two harvests could fall in one year and none in the next. These points were masked out, as the zero yield appears as a model failure.~~

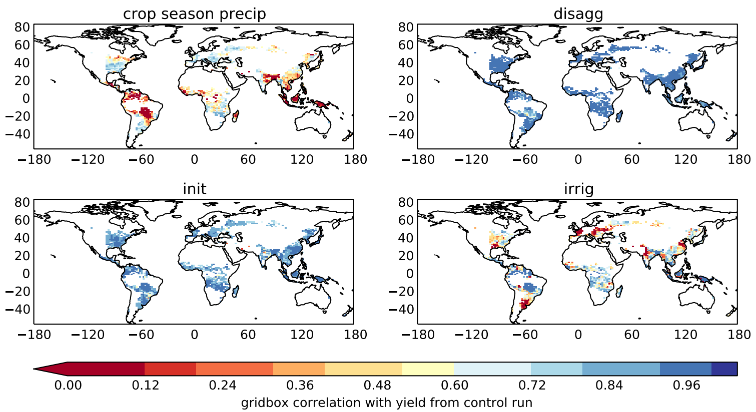



**Figure 1.** All plots show the correlations with the annual maize yield in the `control` run for each gridbox. Top left: the correlation between yield in `control` run and crop season precipitation. Top right, bottom left and bottom right: the correlation between yield in `control` run and yield in the `disagg`, `init` and `irrig` runs, respectively.

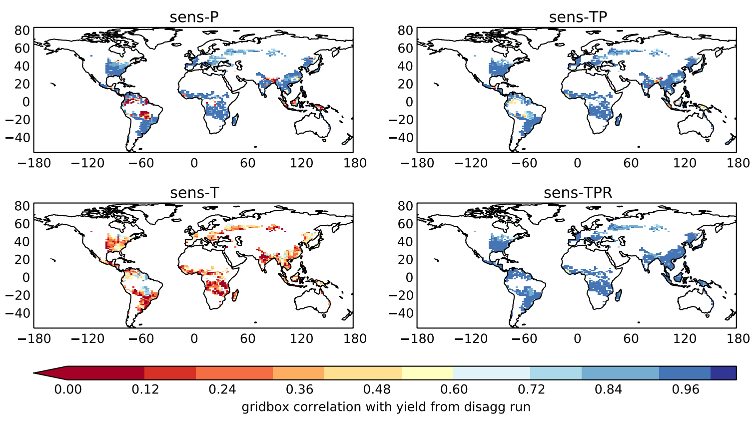

sens-P
sens-TP
sens-T
sens-TPR
80
60
40
20
0
−20
−40
−180 −120 −60 0 60 120 180
0.00 0.12 0.24 0.36 0.48 0.60 0.72 0.84 0.96
gridbox correlation with yield from disagg run

**Figure 2.** The correlations between the annual maize yield in the disagg ~~control~~ run and the annual maize yield from the sens−P (top left), sens−TP (top right), sens−T (bottom left) and sens−TPR (bottom right) runs for each gridbox.

Using daily forcing data and disaggregating rather than using the full six hourly data results in a slightly lower mean global yield (10.2 $\mathrm{Mg\,ha^{-1}}$ for the disaggregated run, compared to 10.6 $\mathrm{Mg\,ha^{-1}}$, see Table 2). The global yield time series from the disaggregated run correlates very well with the global yield time series from the control run: the Pearson correlation coefficient is 0.98. The annual control yield is plotted against the annual disagg yield in Figure 3, and shows no obvious deviations from linearity, even at the extremes. Figure ~~Figure~~ 1 (top right) shows the correlation for each grid box, 94% of which are greater than 0.85 (note that there will be spatial correlation between gridboxes and auto-correlation in the time series for each gridbox. Also the Pearson correlation coefficient is not resistant to outliers). It is interesting to note that many of the gridboxes with low correlations are in Brazil, a region where the disaggregator has been seen previously to reproduce the climatology of key variables such as evaporation better than runs driven with three hourly data (Williams and Clark, 2014). As discussed in Williams and Clark (2014), since the three hourly data is more representative of the underlying driving data than the disaggregated data, this apparent 'improvement' with the disaggregator is likely to be result of the extra parameters involved in the disaggregation being tuned to compensate for a bias elsewhere in the model. As a result, the maize yield from the disaggregated run can actually have a higher correlation with FAO country yield data than the control run for Brazil (not shown here). We can therefore conclude that using daily forcing data and disaggregating is a very good approximation to the control run, for the purposes of looking at variability in the maize yield.

Comparing the control run with the fully irrigated run allows us to determine how much of the modelled yield variability is driven by soil moisture variability. Removing the effect of soil moisture stress increases global NPP as expected, which results in considerably higher global mean yields: maize yield rises from 10.6 to 16.2 $\mathrm{Mg\,ha^{-1}}$ (Table 2). This increase in NPP also has the effect of increasing the number of gridboxes which contribute to the global yield time series, since fewer gridboxes have crops that are harvested prematurely in the model due to lack of growth. Removing soil moisture stress also significantly decreases the (year-to-year) standard deviation for maize yield, which has a global standard deviation of 0.55 $\mathrm{Mg\,ha^{-1}}$ in the control run and 0.18 $\mathrm{Mg\,ha^{-1}}$ in the irrigated run.

We also calculated the Pearson correlation coefficient between the global control run yield and irrigated run yield for each gridbox, as shown in ~~(~~Figure 1, bottom right~~)~~. There was a high correlation coefficient between the two runs in areas with high rainfall during the model maize growing season such as South-east Asia, Central Brazil, the northern part of the Amazon basin and Bangladesh/East India, where we would not expect soil moisture to be a limiting factor in crop growth, even with no irrigation. However, in drier regions, these correlations were much lower, as expected. The percentage of unmasked gridboxes with correlations above 0.85 was just 20% for maize, showing that in most regions, soil moisture variability is an import contribution to the yield variability in the control run.

Moving on from soil moisture to precipitation, we constructed a time series for the crop season precipitation by integrating the rainfall between the sowing and harvesting dates for each crop in each gridbox. In many regions, this crop season precipitation index correlates reasonably well with the crop yield for the unmasked gridboxes, particularly outside of South-east Asia, Central Brazil, the northern part of the Amazon basin and Bangladesh/East India, where, as we have already identified, the modelled yield variability does not follow soil moisture variability.

It is therefore interesting to look at how much of the modelled yield variability can be reproduced if the daily precipitation is used to drive the model, while keeping all other variables at their climatological value for each day of the year (sens-P). *A priori* we can not assume this will be a good approximation to using the full daily driving data (disagg) from the result for the crop season precipitation index above, since, in the control run, the precipitation is not independent of the other driving data. However, Figure 2 (top left) shows that the sens-P run does indeed correlate well with the disagg ~~disagg~~ run in areas outside of South-east Asia, Central Brazil, the northern part of the Amazon basin and Bangladesh/East India. 74% of the gridboxes shown have a correlation of 0.85 or more. The correlation between the global yield timeseries from the sens-P run and the disagg run is 0.87. The sens−P run does have a slightly higher mean global maize yield than the disagg run: 10.9 $\mathrm{Mg\,ha^{-1}}$ as compared to 10.2 $\mathrm{Mg\,ha^{-1}}$.

If temperature is the only variable allowed to vary between years (i.e. the sens−T run), then the global mean maize yield is 10.7 $\mathrm{Mg\,ha^{-1}}$, with standard deviation 0.23 $\mathrm{Mg\,ha^{-1}}$. This reduction in standard deviation compared to the disagg run is consistent with the reduction in standard deviation seen when the effect of soil moisture was removed (the irrig run). Unsurprisingly, figure 2 (bottom left) shows that the sens-T run does not correlate well with the disagg run in areas where the sens-P run had a higher correlation.

If both daily precipitation and daily mean temperature are allowed to vary (~~run~~ sens−PT), the gridbox correlations with the disagg run are much more spatially uniform than when either of these variables are varied on their own: in the ~~for the~~ sens−PT~~sens−P~~ run, ~~:~~ 81% of the gridboxes have a correlation higher than 0.85 (Figure 2, top right)[1]. Many of the areas with low correlations in the sens−P run are much higher in the sens−PT run, such as parts of Brazil, Columbia, Bangladesh and Southeast Asia, although these still remain lower than surrounding regions. The correlation between the global maize yield time series in the sens−TP run and the disagg run is 0.92. The scatter plot of these yield time series (Figure 4) shows that the relation between the outputted yield is well approximated by a linear fit. In general, therefore, driving the model with daily precipitation and mean temperature and using climatology for all other driving variables is a good approximation to make when looking at the interannual yield variability across the majority of global maize-growing regions.

In order to improve the approximation further, it may be desirable to additionally allow downward shortwave radiation to vary (sens−PTR) or additionally allow wind speed to vary (sens−PTW). Allowing downward shortwave radiation to vary improves performance (i.e. gridbox correlations with the disagg run) in the areas which still have relatively low performance in the sens−PT run i.e. Brazil, Columbia, Bangladesh and Southeast Asia

[1] ~~We can see in Figure 2 (bottom left) that, as expected, the sens−T run does not correlate well with the disagg run in the areas where the sens−P run had a higher correlation.~~

(Figure 2, bottom right). Alternatively, allowing wind speed to vary results in a mean global yield that is closer to the mean global of the `disagg` run (Table 2).

The final remaining question concerns the model initialisation. The set of runs that are initialised on each sowing date with climatology (`init`) in general reproduce the spatial distribution of yield from the `control` run. The global yield is generally lower than in the `control` run in each year, which results in slightly lower mean global yield (10.3 $\mathrm{Mg\,ha^{-1}}$) compared to the `control` run (10.6 $\mathrm{Mg\,ha^{-1}}$). The correlation between the global maize yield in the `init` run and the `control` run is 0.91 (see Figure 3 for scatter plots) and 70% of individual gridboxes have a correlation above 0.85 (Figure 1, bottom left). The correlations are relatively poor in some parts of India, the Congo basin and South/Southeastern Brazil. However, outside these areas, initialising on the sowing date has the potential to be a very useful approximation.

## 5 Conclusions

In this article, we have investigated a number of possible approximations that could be made when running JULES-crop:

- Use driving data at daily rather than subdaily resolution, and disaggregate internally to the model timestep
- Use a subset of daily driving data and set the rest to a daily climatology
- Initialise with climatology on the crop sowing date

Each of these approximations significantly simplify the use of JULES-crop for seasonal crop yield forecasts, due to the reduction in required driving and initialisation data. With this usage in mind, we have concentrated on the effect of these approximations on the interannual variability of the modelled yield.

Using daily forcing data and disaggregating performs the best out of these approximations, although care should be taken if modelling the Amazon basin, where the precipitation disaggregation parameters may have been tuned to compensate for biases in JULES.

We have shown that, in most regions outside South-east Asia, Central Brazil, the northern part of the Amazon basin and Bangladesh/East India, the interannual variability of the yield from a JULES-crop run in the control configuration is mainly driven by precipitation, which affects the crop via water availability from the soil, which we have confirmed with a fully irrigated run. As a result, in these regions, it is a good approximation to drive the model with forecast precipitation and leave the other driving data at their climatological values for each day of year. It should be noted that the processes and parameters which govern the response of the crop model to the soil moisture distribution, such as the soil water factor $\beta$ and the root distributions in JULES, are therefore keys areas for future model development. Driving the model with both precipitation and temperature improves the performance in areas with high soil moisture and some further improvement in these areas can be obtained from the addition of downward shortwave radiation.

Perhaps the most important approximation considered here is initialising with climatology on the sowing date, since obtaining accurate initialisation data on the timescales needed for seasonal forecast runs is a particularly significant practical challenge. We have confirmed that this approximation performs well across the majority of maize-growing regions and identified areas where the approximation breaks down.

Taken together, these approximations allow JULES-crop to be driven by seasonal meteorological forecasts where ensembles of bias corrected daily precipitation and daily temperature (and possibly downward short-wave radiation) are available. The reference dataset used for the bias correction can be used to generate the climatology of the initialisation variables and the other driving variables. Since this data is widely available, this provides a practical methodology by which to obtain seasonal crop forecasts with JULES-crop.

*Acknowledgements.* The authors are grateful to Jemma Gornall, Tom Osborne and Camilla Mathison for useful discussions. This article was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101) and the EUPORIAS project, funded by the European Commission 7th Framework Programme for Research, grant agreement 308291. This work contributed to EUPORIAS D23.5.

| name | sens-T | sens-P | sens-TP | sens-TPR | sens-TPW |
|---|---|---|---|---|---|
| mean temperature (T) | x | | x | x | x |
| precipitation (P) | | x | x | x | x |
| downward short-wave radiation (R) | | | | x | |
| wind speed (W) | | | | | x |

**Table 1.** Combinations of driving variables that are allowed to vary in the sens-* runs. Each column is a separate run. All driving variables not marked with an ‘x’ are set to their daily climatology.

| name | mean | standard deviation | global corr with `control` | global corr with `disagg` |
|---|---|---|---|---|
| `control` | 10.6 | 0.55 | | |
| `irrig` | 16.2 | 0.18 | 0.48 | |
| `init` | ~~10.3~~ | ~~0.48~~ | ~~0.91~~ | |
| `disagg` | 10.2 | 0.53 | 0.98 | |
| `sens-T` | 10.7 | 0.23 | | 0.23 |
| `sens-P` | 10.9 | 0.42 | | 0.87 |
| `sens-TP` | 11.1 | 0.51 | | 0.92 |
| `sens-TPR` | 11.1 | 0.50 | | 0.92 |
| `sens-TPW` | 10.3 | 0.52 | | 0.96 |

**Table 2.** Results from the global runs described in Section 3. First column is the run name, second is the mean maize yield in $\mathrm{Mg\,ha^{-1}}$, third is the standard deviation of the annual global yield time series in $\mathrm{Mg\,ha^{-1}}$. The fourth column gives the Pearson correlation coefficient with the global yield in the `control` run and the fifth column gives the Pearson correlation coefficient with the global yield in the `disagg` run. All results have been weighted as described in Section 4. These results are presented as scatter plots in the appendix.

## Appendix A: Scatter plots of global yield from model runs

**Figure 3.** Scatter plots comparing the global mean maize yield from different model runs.

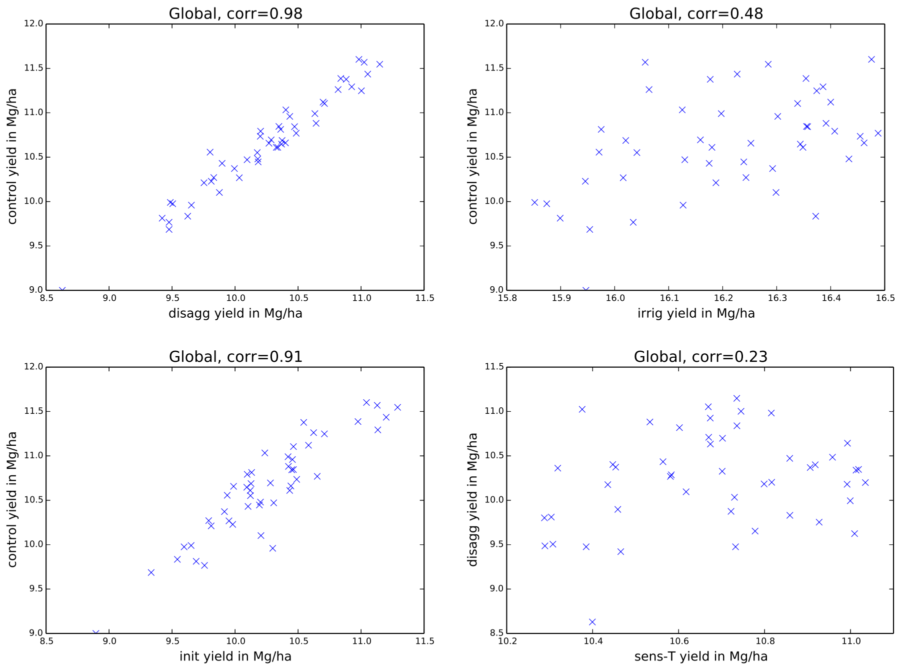
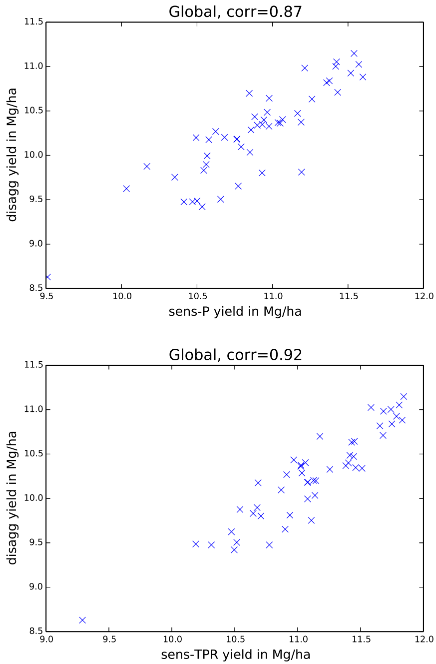


**Figure 4.** Scatter plots comparing the global mean maize yield from different model runs.



## 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., Muller, 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., White, J. W., Williams, J. R., and Wolf, J.: Uncertainty in simulating wheat yields under climate change, Nature Clim. Change, 3, 827–832, doi:10.1038/nclimate1916, http://dx.doi.org/10.1038/nclimate1916, 2013.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geoscientific Model Development, 4, 677–699, doi:10.5194/gmd-4-677-2011, http://dx.doi.org/10.5194/gmd-4-677-2011, 2011.

Cantelaube, P. and Terres, j.-M.: Seasonal weather forecasts for crop yield modelling in Europe, Tellus A, 57, 476–487, doi:10.1111/j.1600-0870.2005.00125.x, http://dx.doi.org/10.1111/j.1600-0870.2005.00125.x, 2005.

Challinor, A. J., Slingo, J. M., Wheeler, T. R., and Doblas-Reyes, F. J.: Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles, Tellus A, 57, 498–512, doi:10.1111/j.1600-0870.2005.00126.x, http://dx.doi.org/10.1111/j.1600-0870.2005.00126.x, 2005.

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, doi:10.1038/nclimate2153, http://dx.doi.org/10.1038/nclimate2153, 2014.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701–722, doi:10.5194/gmd-4-701-2011, http://dx.doi.org/10.5194/gmd-4-701-2011, 2011.

Coelho, C. A. S. and Costa, S. M. S.: Challenges for integrating seasonal climate forecasts in user applications, Current Opinion in Environmental Sustainability, 2, 317–325, doi:10.1016/j.cosust.2010.09.002, http://dx.doi.org/10.1016/j.cosust.2010.09.002, 2010.

Collatz, Ball, Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agricultural and Forest Meteorology, 54, 107–136, doi:10.1016/0168-1923(91)90002-8, http://dx.doi.org/10.1016/0168-1923(91)90002-8, 1991.

Collatz, G. J., Ribas-Carbo, M., and Berry, J. A.: Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants, Australian Journal of Plant Physiology, 19, 519–538, http://www.publish.csiro.au/?paper=PP9920519, 1992.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geoscientific Model Development, 4, 1051–1075, doi:10.5194/gmd-4-1051-2011, http://dx.doi.org/10.5194/gmd-4-1051-2011, 2011.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., and Tarpley, J. D.: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), J. Geophys. Res., 108, 8845+, doi:10.1029/2002jd003316, http://dx.doi.org/10.1029/2002jd003316, 2003.

Dessai, S. and Bruno Soares, M.: Systematic literature review on the use of seasonal to decadal climate and climate impacts predictions across European sectors, Tech. Rep. D12.1, University of Leeds, http://www.euporias.eu/system/files/D12.1_Final.pdf, 2013.

Falloon, P., Fereday, D., Stringer, N., Williams, K., Gornall, J., Wallace, E., Eade, R., Brookshaw, A., Camp, J., Betts, R., Dankers, R., Nicklin, K., Vellinga, M., Graham, R., Arribas, A., and MacLachlan, C.: Assessing Skill for Impacts in Seasonal to Decadal Climate Forecasts, Geology & Geosciences, 2, http://www.omicsgroup.org/journals/assessing-skill-for-impacts-in-seasonal-to-decadal-climate-forecasts-2329-6755-2-1000e111.php?aid=18791, 2013.

Falloon, P., Buontempo, C., Ceron, J. P., Dubus, L., Greuell, W., Hutjes, R., Jönsson, A. M., Lowe, R., Manzanas, R. G., San Martin, M. T., Newton, P., Nicklin, K., Pechlivanidis, I., Pouget, L., Supit, I., and Williams, K.: Report on initialisation of impacts models for seasonal predictions, Tech. Rep. D23.3, EUPORIAS, http://www.euporias.eu/system/files/D23.3_Final.pdf, 2014a.

Falloon, P., Challinor, A., Dessai, S., Hoang, L., Johnson, J., and Koehler, A.-K.: Ensembles and uncertainty in climate change impacts, Frontiers in Environmental Science, 2, 33+, doi:10.3389/fenvs.2014.00033, http://dx.doi.org/10.3389/fenvs.2014.00033, 2014b.

Hansen, J. W., Challinor, A., Ines, A., Wheeler, T., and Moron, V.: Translating climate forecasts into agricultural terms: advances and challenges, Climate Research, 33, 27–41, doi:10.3354/cr033027, http://dx.doi.org/10.3354/cr033027, 2006.

Hansen, J. W., Mason, S. J., Sun, L., and Tall, A.: Review Of Seasonal Climate Forecasting For Agriculture In Sub-saharan Africa, Experimental Agriculture, 47, 205–240, doi:10.1017/s0014479710000876, http://dx.doi.org/10.1017/s0014479710000876, 2011.

Hewitt, C., Buontempo, C., and Newton, P.: Using Climate Predictions to Better Serve Society's Needs, Eos Trans. AGU, 94, 105–107, doi:10.1002/2013eo110002, http://dx.doi.org/10.1002/2013eo110002, 2013.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanenkov, V., Oettli, P., Newby, T., Ishigooka, Y., and Furuya, J.: Historical changes in global yields: major cereal and legume crops from 1982 to 2006, Global Ecology and Biogeography, 23, 346–357, doi:10.1111/geb.12120, http://dx.doi.org/10.1111/geb.12120, 2014.

Jones, C. D., Hughes, J. K., Bellouin, N., Hardiman, S. C., Jones, G. S., Knight, J., Liddicoat, S., O'Connor, F. M., Andres, R. J., Bell, C., Boo, K. O., Bozzo, A., Butchart, N., Cadule, P., Corbin, K. D., Doutriaux-Boucher, M., Friedlingstein, P., Gornall, J., Gray, L., Halloran, P. R., Hurtt, G., Ingram, W. J., Lamarque, J. F., Law, R. M., Meinshausen, M., Osprey, S., Palin, E. J., Chini, L. P., Raddatz, T., Sanderson, M. G., Sellar, A. A., Schurer, A., Valdes, P., Wood, N., Woodward, S., Yoshioka, M., and Zerroukat, M.: The HadGEM2-ES implementation of CMIP5 centennial simulations, Geoscientific Model Development, 4, 543–570, doi:10.5194/gmd-4-543-2011, http://dx.doi.org/10.5194/gmd-4-543-2011, 2011.

Le Quéré, C., Peters, G. P., Andres, R. J., Andrew, R. M., Boden, T. A., Ciais, P., Friedlingstein, P., Houghton, R. A., Marland, G., Moriarty, R., Sitch, S., Tans, P., Arneth, A., Arvanitis, A., Bakker, D. C. E., Bopp, L., Canadell, J. G., Chini, L. P., Doney, S. C., Harper, A., Harris, I., House, J. I., Jain, A. K., Jones, S. D., Kato, E., Keeling, R. F., Klein Goldewijk, K., Körtzinger, A., Koven, C., Lefèvre, N., Maignan, F., Omar, A., Ono, T., Park, G. H., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Schwinger, J., Segschneider, J., Stocker, B. D., Takahashi, T., Tilbrook, B., van Heuven, S., Viovy, N., Wanninkhof, R., Wiltshire, A., and Zaehle, S.: Global carbon budget 2013, Earth System Science Data, 6, 235–263, doi:doi:10.5194/essd-6-235-2014, http://www.earth-syst-sci-data.net/6/235/2014/essd-6-235-2014.html, 2014.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, Q.J.R. Meteorol. Soc., 141, 1072–1084, doi:10.1002/qj.2396, http://dx.doi.org/10.1002/qj.2396, 2015.

Manzanas, R., Frías, M. D., Cofiño, A. S., and Gutiérrez, J. M.: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill, J. Geophys. Res. Atmos., 119, 1708–1719, doi:10.1002/2013jd020680, http://dx.doi.org/10.1002/2013jd020680, 2014.

Marletto, V., Ventura, F., Fontana, G., and Tomei, F.: Wheat growth simulation and yield prediction with seasonal forecasts and a numerical model, Agricultural and Forest Meteorology, 147, 71–79, doi:10.1016/j.agrformet.2007.07.003, http://dx.doi.org/10.1016/j.agrformet.2007.07.003, 2007.

Mishra, A., Hansen, J. W., Dingkuhn, M., Baron, C., Traoré, S. B., Ndiaye, O., and Ward, M. N.: Sorghum yield prediction from seasonal rainfall forecasts in Burkina Faso, Agricultural and Forest Meteorology, 148, 1798–1814, doi:10.1016/j.agrformet.2008.06.007, http://dx.doi.org/10.1016/j.agrformet.2008.06.007, 2008.

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., and Vitart, F.: The new ECMWF seasonalforecast system (System 4), Tech. Rep. 656, ECMWF, http://old.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm656.pdf, 2011.

Monfreda, C., Ramankutty, N., and Foley, J. A.: Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochem. Cycles, 22, GB1022+, doi:10.1029/2007gb002947, http://dx.doi.org/10.1029/2007gb002947, 2008.

Nicklin, K. J., Challinor, A., and Tompkins, A.: The use of seasonal forecasts in a crop failure early warning system for West Africa, AGU Fall Meeting Abstracts, pp. A1007+, http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2011AGUFMGC31A1007N, 2011.

Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., and Wheeler, T.: JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator, Geoscientific Model Development, 8, 1139–1155, doi:10.5194/gmd-8-1139-2015, http://dx.doi.org/10.5194/gmd-8-1139-2015, 2015.

Palmer, T. N., Doblas-Reyes, F. J., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., Feddersen, H., Cantelaube, P., Terres, J. M., Davey, M., Graham, R., Délécluse, P., Lazar, A., Déqué, M., Guérémy, J. F., Díez, E., Orfila, B., Hoshen, M., Morse, A. P., Keenlyside, N., Latif, M., Maisonnave, E., Rogel, P., Marletto, V., and Thomson, M. C.: Development Of A European Multimodel

Ensemble System For Seasonal-to-interannual Prediction (DEMETER), Bull. Amer. Meteor. Soc., 85, 853–872, doi:10.1175/bams-85-6-853, http://dx.doi.org/10.1175/bams-85-6-853, 2004.

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, doi:10.1111/j.1466-8238.2010.00551.x, http://dx.doi.org/10.1111/j.1466-8238.2010.00551.x, 2010.

Watson, J. and Challinor, A.: The relative importance of rainfall, temperature and yield data for a regional-scale crop model, Agricultural and Forest Meteorology, 170, 47–57, doi:10.1016/j.agrformet.2012.08.001, http://dx.doi.org/10.1016/j.agrformet.2012.08.001, 2013.

Watson, J., Challinor, A., Fricker, T., and Ferro, C.: Comparing the effects of calibration and climate errors on a statistical crop model and a process-based crop model, Climatic Change, pp. 1–17, doi:10.1007/s10584-014-1264-3, http://dx.doi.org/10.1007/s10584-014-1264-3, 2014.

Williams, K. and Clark, D.: Disaggregation of daily data in JULES, Tech. Rep. 96, Met Office Hadley Centre, http://www.metoffice.gov.uk/media/pdf/2/j/HCTN96.pdf, 2014.