Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-243-AC2, 2017

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "The impacts of data constraints on the predictive performance of a general process-based crop model (PeakN-crop v1.0)" *by* Silvia Caldararu et al.

Silvia Caldararu et al.

Matthew.Smith@Microsoft.com

Received and published: 30 January 2017

Response to Reviewer 2

We thank the reviewer for their review. We believe that addressing these comments will add value to our work. We did find some of the comments unclear but have done our best to reply to them all.

Comment from Reviewer 1: "The extensive linkage to 'food security' is not necessary and the meaning of the own contribution overstated."

Response from Authors 1: Since GMD is not a subject specific journal we find it is helpful to include the link to the bigger picture and clarify why improving agricultural

Printer-friendly version

Discussion paper

models is important. The link to food security was the main reason why we undertook the research described in the manuscript.

Proposed changes to Manuscript 1: We have not removed our original framing of the study. However, in response to comments from the other reviewer, we proposed to include more discussion on the lack of precision of our current model at predicting crop yields and the need for more research with more location-specific data, such as

"Model uncertainty is difficult to compare with previous crop modelling studies, as models with fixed parameter values do not often provide uncertainty estimates. In fact, providing uncertainty values for all model variables and parameters is one of the advantages of a data constrained model. In the current model, uncertainty is highest at the start of the season for all variables but decreases rapidly and final yield uncertainty is much lower. This is due to thresholds: abrupt changes from one growing stage to another when small differences in parameters can lead to large differences in resulting variables. It is, however, important to note that the uncertainty in our yield predictions remains high and the model in its current form is unlikely to provide accurate predictions for practical applications without the addition of new data (Section 7.4). We have however shown that the use of three different data types does reduce prediction uncertainty - pointing to an avenue for future model improvement."

and

"If this model, or any other similar process-based data constrained crop model, is to be used for scientific purposes to understand the response of crops to climate across the globe, the ideal data would be a global data set, such as space-based vegetation observations, combined with high quality field level data that would ideally include growth timeseries, final grain yield and information about agricultural practices. However, if the model is to be used for agricultural purposes, to aid decision making at the local level, high quality field level data would be sufficient. An valuable evaluation in such studies, not conducted here for brevity and due to a lack of location-specific data, would be to

compare the predictive accuracy of the model against the predictive accuracy of a statistical average over the data. Such an analysis would reveal whether and how much benefit is gained by using a data constrained model for predictions"

Comment from Reviewer 2: "The use of categories is not convincing. How can statistical models be considered non-mathematical and process based models mathematical? (line 16-17)."

Response from Authors 2: The phrasing on lines 16-17 is indeed wrong. The process-based and statistical model separation is one that is commonly used not only for crop models but also in the field of earth system models and one that we find useful in explaining how process knowledge and data are used to obtain agricultural predictions.

Proposed changes to Manuscript 2: We will edit the confusing sentence to read

"Predicting and understanding how crops respond to changes in their environment through the use of mathematical models is needed to help address such threats, enabling advanced warning of potential threats and predictions of what alterations to agricultural practices might help prevent or mitigate problems."

We then go on to explain in detail the difference between process-based and statistical models (both are mathematical!)

Comment from Reviewer 3: "The influence of the different data sources on parametrization is not considered. For example, the inclusion of farm yield data would necessarily imply that management effects in-fluence the parametrization. This is similar to the parametrization of statistical models and should have been addressed in a different way on page 2 lines 30-35.

Response from Authors 3: The influence of different data sources on model parametrization is the main topic of our paper. management and field level information is required in process based models but not included explicitly in statistical models, as we discuss in the paragraph mentioned by the reviewer. Unfortunately the meaning of

this comment is not entirely clear.

Proposed changes to Manuscript 3: we now include more details of where we got our data for fertilizer, sowing and harvest dates – an issue also raised by the other Reviewer.

“In addition to the three datasets used for parametrisation, the model also requires input data in the form of sowing and harvest dates and fertiliser inputs. Additional uncertainty is associated with these datasets which is not available nor accounted for in our analyses. For example, the crop calendar (Sacks et al., 2010) and Nitrogen Fertilizer Application (Potter et al., 2010) datasets are global data collections that will imperfectly represent the value for any given location. Alternatives to these global datasets would be to use location-specific data, or to infer the values. Location specific data has the advantage of more accurately reflecting the situation at a given site and would therefore be useful when the model is applied at the field scale, but such data is unlikely to be available for all sites. Successful inference of the values would depend on if there is enough information in the datasets used to infer the model parameters. If there is inadequate data then there would be excessive degrees of freedom for inference, leading to the wrong parameter values begin inferred and the model performing poorly in novel situations. Therefore, the decision whether to obtain more data or infer unknown quantities in future applications of our model and inference framework depends on the data availability and the intended scales of application.”

Comment from Reviewer 4: “The introduction ends with three valid research questions, however, the concrete model that will be used to address these questions remains open.”

Response from Authors 4: We should indicate in the introduction that we intend to introduce and use a new model.

Proposed changes to Manuscript 4: We will clarify in the introduction that we use a new model

“In this paper we present a newly developed general, non-crop specific process based model and use parameter inference to infer the most likely parameters for 15 locations for winter wheat and maize using a combination of space-based vegetation indices, eddy covariance flux data and reported agricultural yields.”

Comment from Reviewer 5: “The claim for a new model (page 5 line 15) adds surprising additional dimensions to the paper.”

Response from Authors 5: As we discussed in section 7.3, we chose to use a new model as it is more general and allows us to perform our analysis for multiple sites and species. We acknowledge that the use of this new model also has certain disadvantages and we mention this in the discussion.

Proposed changes to Manuscript 5: We will partially address this by mentioning that the model is new in the introduction. We also already cover the need to compare our model to others in the discussion

“Here we have chosen a given model structure and extensively tested the way in which constraining the parameters with different datasets in different configurations. The question that arises is to what extent the chosen model itself affects the present results. We have chosen a novel, physiology based model which includes plant optimality concepts, which on one hand has the advantage that it is more general than some of the older models and lacks artificially set thresholds between growth stages, but does have the disadvantage of being less thoroughly tested against field observations. An ideal companion paper to this study would be a comparison of different model structures with a constant data constraining framework, providing greater insights into which parts of the model lead to high errors or uncertainty. However, given the limitations of the current study, we acknowledge this limitation and report most error metrics as relative to prior model runs in an attempt to isolate errors created by the data and model fitting from those caused by the model itself.”

Comment from Reviewer 6: “How was the soil variability parametrized?”

Response from Authors 6: As this is a very simple model at this stage the only soil information needed was nitrogen fertilizer application.

Proposed changes to Manuscript 6: At the suggestion of both reviewers we have added a discussion of any additional soil information needed for a more detailed model.

“The model in the version presented in this paper does not include any water limitation to growth due mainly to a lack of data constraint on any water related parameters, as we found that latent heat data from EC towers is not sufficient. Below-ground measurements of not only root growth but also soil water properties would again provide some of the necessary information. Such belowground data, especially if supplemented by nutrient concentrations can also help constrain a more complex version of the nitrogen uptake scheme, which could be improved to include more explicit soil-plant interactions and additional processes such as biological nitrogen fixation for legumes.”

Comment from Reviewer 7: “The original parameter values are not given and any validation results are missing.”

Response from Authors 7: As we explain in section 4 (Parameter estimation technique) we use a Bayesian fitting method which requires prior intervals for the parameter but not prior parameter values. As explained in section 5, the prior parameter values are randomly sampled from the prior parameter distribution in a manner similar to parameters being sampled from the posterior. The paper contains extensive model validation, in fact it contains little else. Figure 1 shows a comparison of prior and posterior model performance and figures in the appendix contain site level model-data comparison as the results of cross-site validation. Model validation is discussed extensively in both the results and discussion section.

Proposed changes to Manuscript 7: We will adjust the aim statement in the paper to make clear that we are inferring our parameters

“In this paper we present a newly developed general, non-crop specific process based

model and use parameter inference to infer the most likely parameters for 15 locations for winter wheat and maize using a combination of space-based vegetation indices, eddy covariance flux data and reported agricultural yields."

Comment from Reviewer 8: "The assigned uncertainties for the given data sources are difficult to follow. A systematic reasoning for the chosen uncertainty values is missing."

Response from Authors 8: A description of how we include data uncertainty in model fitting can be found in section 4. We acknowledge that this can be difficult to follow for those new to, or unfamiliar with, with Bayesian fitting methods and we will extend this description

Proposed changes to Manuscript 8: We propose to adjust the paragraph on data uncertainty to read.

"We adopt different techniques to estimate the standard deviation σ_(x,D) above, depending on the dataset D at each location. Generally, we assume that the variation in the model predictions about the data is solely due to uncertainty in the data. The GPP data do not have an estimate of uncertainty and so we infer the uncertainty associated with those data as the parameter σ_(x,D). In the case of MODIS fAPAR data we explicitly incorporate a measure of variation in the data within the geographical area used to compute the mean fAPAR as well as inferring a parameter representing additional unexplained variation. We include this parameter to account for known issue in space based remotely sensed data, such as background soil reflectance. The crop yield data already have estimates of observational uncertainty associated with them and so we use those data to define σ_(x,D)."

Comment from Reviewer 9: "The presentation of the results continues the deficits of the M&M section. It does not fulfill the existing standards."

Response from Authors 9: We have presented our results in a manner common to model-data fusion studies.

Proposed changes to Manuscript 9: Without further explanation of the reviewer's existing standards we cannot improve this section to their satisfaction.

Comment from Reviewer 10: "What was the quantitative propagation of the initial parameter setting?"

Response from Authors 10: As explained above, the fitting method does not require initial parameter settings and in any case it is not clear to us what propagation of parameter settings refers to. We have striven to offer a clear explanation of the Bayesian fitting method used in our study but given the length limitations of a scientific paper we found that a detailed explanation of the basics of model fitting methods was not feasible.

Proposed changes to Manuscript 10: As in our reply to the other reviewer, we have expanded on our methods paragraph describing how we propagate parameter uncertainty

"To calculate uncertainty for the model predictions we sample parameter values from their respective posterior distribution and compute predictions with each parameter combination, which results in a corresponding distribution of model predictions. We report this prediction distribution uncertainty using 95th percent confidence intervals. This predicted distribution does not include the prescribed or inferred uncertainty about observations, σx,D, our predicted distributions correspond to the state being predicted and not the observations of that state."

Comment from Reviewer 11: "This leads to my main criticism of the paper: the results given are not reproducible."

Response from Authors 11: In accordance to the GMD publication requirements, the model code and settings are available upon request from the authors. The model fitting algorithm, developed by our group, has been freely available for several years. All the data used is freely available and fully referenced in the text.

Proposed changes to Manuscript 11: We do not propose any changes because we already include statements about the code and data availability in the manuscript.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-243, 2016.