Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-304-AC2, 2017

Hydrology and Earth System Sciences
Discussions

Interactive comment on "A coupled human-natural system to assess the operational value of weather and climate services for irrigated agriculture" *by* Yu Li et al.

Yu Li et al.

andrea.castelletti@polimi.it

Received and published: 14 July 2017

We thank the reviewer for the positive comments and thorough review that will surely help us improving the manuscript. We will take all of them into consideration while revising the paper. Below, our point-to-point response.

Referee comment n. 2

The presented manuscript describes and applies a methodological framework to assess the operational value of weather and climate forecast products on irrigated agriculture. It combines a set of forecast products with an agronomic model that simulates

the crop yield based on meteorological inputs and an agent-based model that establishes the optimal cropping pattern depending on the forecasts available and the risk profile of the farmers. The novelty of the paper consists in the joint assessment of the forecast quality and its impact on management decisions and farmers risk profile. The methodology is well described and the structure and organization of the paper is coherent and adequate. The results point at the fact that the forecast quality is not necessary correlated with its impact on management decisions. The paper fits the scope of the journal and has a clear potential for publication, given the increasing momentum of weather and climate services and how its "real" impact can be measured. I have no major concerns about the manuscript, although some improvements would further increase its quality. Therefore, I would consider it ready for publication after fixing the minor concerns I point at below.

We thank the referee for the positive comment.

TECHNICAL COMMENTS

1. Page 2, lines 28-31: In my opinion, the first sentence of this paragraph is just a summary of the previous one. I would delete it and reflect in the previous paragraph that an alternative promising metric would be the quality obtained on predicting decision- relevant variables.

   Following the reviewer comment, we will delete this sentence.

2. Page 3, line 23: Although it becomes clear when moving forward that "post-processing" means "downscaling and bias-correction", I would add a remark here just to clarify it.

   We thank the reviewer for the suggestion and we will clarify from the beginning the meaning of post-processing.

3. Page 4, line 6: what do you mean when you state "pilot"? I think it is a synonym of "case study", but sometimes the term "pilot" implies you run field experiments

to apply the method developed. Please clarify the term.

We agree with the reviewer that the term pilot might be misinterpreted. Since we used it a synonym for case study, we will remove this term in the revised version.

4. Page 5, lines 33-34: can you provide information to support the assumption of using crop yield as main driver of the cropping pattern decisions? Sometimes other variables like management complexity or profit predictability is more important than crop yield. In my opinion, you should clarify, if it is the case, that you make this assumption in the absence of more detailed information about the farmers' decision-making process.

   We agree with the reviewer that this assumption should be clarified. In the revised version of the paper, we will mention that, in the absence of more detailed information about the farmers' decision-making process, we introduced this assumption on the basis of other similar studies (Hansen, 2004; Baigorria et al., 2008). *Hansen, J. (2004). "Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya". In: Agricultural and Forest Meteorology 125.1-2, pp. 143–157. Baigorria, G. a., J. W. Jones, and J. J. O'Brien (2008). "Potential predictability of crop yield using an ensemble climate forecast by a regional circulation model". In: Agricultural and Forest Meteorology 148.8-9, pp. 1353–1361.*

5. Page 6, line 21: as far as I know, the quantile-based mapping is a bias correction procedure. It is true that it has some downscaling component due to matching CDFs obtained at different spatial scales but, on a broader view, it is considered as a bias correction technique. In fact, you previously named it as a bias correction technique. Please fix this

   Following the reviewer suggestion, we will fix this point in the revised version by consistently characterizing the quantile mapping as a bias-correction technique.

6. Page 8, lines 15-16: the way in which the aggregation is performed it is not clear. I assume you aggregate the daily data of the same month, but it may also mean you aggregate the ensemble members. Please clarify it. If you aggregate the ensemble members to obtain a unique factor, I would rather suggest keeping the factor obtained by each ensemble member and generate synthetic daily time series with all of them. In this way, you will have a better representation of the extremes, which are flattened when taking the average.

   We agree with the reviewer that this step is not clear. We perform the following aggregations: first we aggregate the daily data of the same month, then we estimated a monthly perturbing factor for each ensemble members, and then we took the average factor across the ensemble's members. We are aware that in this way we lose some information on the extremes and we agree with the reviewer that performing the entire assessment on each single ensemble member would allow a better characterization of the extremes as well as exploring how this uncertainty is propagated when moving from the forecast quality to the operational value. Yet, this would be computationally challenging as it would require running 96 simulations per year, for a total of around 500 computational hours. This computational effort goes beyond the scope of this paper. Moreover, the use of large ensembles opens up a number of challenges (see the reply to the second point raised by R2) and the consequences of aggregating or not aggregating the ensemble members can be analyzed in detail, potentially focusing on a single forecast product, in a future work. In the revised version of the paper, we will clarify how we perform this aggregation and we will include this aspect in the list of assumption added in the conclusions section (see R1 suggestion), suggesting as a possible follow-up work the opportunity of refining our analysis keeping all the ensemble members separated.

7. Page 12, table 1: please include the ensemble members of each WCS used unless all the products provide just one ensemble. In this last case, you should

indicate in the text that all of them provide a unique ensemble member.

All the forecast products are in the form of ensembles: ECMWF products have 9 ensemble members (or 3 in case of decadal products), CanSIps have 10 ensemble members, CSFv2 has 4 ensemble members. We will include this information in the revised table 1.

8. Page 15, figure 6: from my point of view, the understandability of this figure would be increased by including the legend inside each individual plot as well as the name of the WCS product. Otherwise the reader needs to constantly go up and down the figure to find out what each bar refers to.

   We agree with the reviewer comment and in the revised version we will move the legends inside the subplots as suggested to improve the interpretability of the figure.

9. Page 16, figure 7: same comment as for figure 6.

   We agree with the reviewer comment and in the revised version we will move the legends inside the subplots as suggested to improve the interpretability of the figure.

10. Page 18, figure 8: I would include the name of the WCS product in each individual plot. Furthermore, I would also provide the value of an average score for the time series inside each plot (for example the MAE). In this way, the reader has a numerical way to easily compare the accuracy of each WCS product type for each plot.

    We thank the reviewer for the suggestion and, in the revised version of the paper, we will move the legends of the subplots as suggested and we will add a numerical score to facilitate the comparison across WCS products.

11. Page 19, lines 1-5: Did you generate 100 time series for each year between 2001 and 2005? Did you choose one year between 2001 and 2005 and then generate

100 series for it? Or did you spare the 100 time series between 2001 and 2005? Please add a clarification about it.

We actually generated 100 time series for each year over the evaluation horizon (2001-2005). We will clarify this point in the revised version of the paper.

12. Page 20, lines 1-14: In my opinion, the fact that the neutral or optimistic risk profiles did not obtain the best performance for the best forecast deserves more explanation. How can you justify this issue? In the absence of more information, I would doubt about the suitability of the score used (median and variance of MAE). Maybe the IFS/HOPE product does not predict extremes as ECHAM5/MPIOM does, and due to this reason the latter offers the best performance on both the neutral and the optimistic risk profiles. Please add some explanation or theory about this unexpected finding.

    This unexpected finding can be explained by the fact that forecast accuracy metrics quantify the error in predicting the agricultural production, while the operational value estimated through the decision model relies on the ranking of the available options (cropping patterns). Sub-optimal decisions are made when the forecasted productivity of the crops produces a different ranking with respect to the one resulting at the end of the agricultural season. However, such rank reversals are not linearly related to the forecast accuracy: large but consistent (e.g., systematic over/underestimation) errors for all the crops may produce the same ranking and result in optimal decisions, while smaller and variable errors can produce sub-optimal decisions. This is quite clear if we consider the forecast accuracy reported in Fig. 8 of ECMWF(annual) IFS/HOPE and ECHAM5/MPIOM: looking at the values in 2001, ECHAM5/MPIOM (which in Fig. 10 has the best performance) is systematically overestimating the productivity of all the crops; IFS/HOPE instead underestimates the productivity of tomato while overestimates the one of rice, potentially reverting the ranking of these crops and producing sub-optimal decisions. Following the reviewer suggestion, we will clarify this point in

the revised manuscript.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-304, 2017.