

## ***Interactive comment on “The impact of land model structural, parameter, and forcing errors on the characterization of soil moisture uncertainty” by V. Maggioni et al.***

**Anonymous Referee #1**

Received and published: 16 March 2012

In this paper, the authors try to assess the contribution of different kinds of uncertainty on total forecast uncertainty. The idea is certainly interesting. However, I do have my doubts about a number of issues.

- My major problem is that essentially with this kind of study, you will get back what you enter into the models (the authors actually state this also on page 2296). For example, in the GLUE analysis, the authors define a threshold for the goodness of fit. The ensemble members that pass this threshold are then retained for further analysis. To me, this implies that the ensemble does not really represent forecast uncertainty. An ensemble that represents forecast uncertainty is generated starting from known

C369

or calibrated parameters, initial conditions, and forcings, to which noise or errors are added. That would mean that a number of ensemble members would not pass the threshold for the goodness of fit.

- I also have a problem with the adding of noise to the model variables. The paper does not provide a good explanation of where the noise parameters come from. Essentially, one will again get back what one puts in (more noise, more uncertainty). One could apply an ensemble verification (enough papers have been written about that), and calibrate this noise.

- As a summary of my first two remarks, I think that the paper does not deliver what it promises. The impact of the different error sources on forecast uncertainty is not investigated, the authors demonstrate that adding two different kinds of noise leads to a larger ensemble spread than only using the individual kinds of noise only.

Some further remarks:

- How can it be justified to generate an ensemble by disturbing only two parameters, while this kind of models for sure uses many parameters? As the authors state, you need to identify the most sensitive parameters, but I do not believe that two parameters is enough for this kind of analysis.

- Also, is fourteen ensemble members really enough for a statistical analysis?

- I am also not convinced that looking at anomaly time series is a good way to assess model performance. This tends to make model results look much better than they really are.

- Are there references for the ensemble metrics that the authors use? How are these related to more frequently used statistics?

- The references need to be checked, a number of times the journal names are not listed completely.

C370

- I am also confused about the way the parameters have been disturbed to generate the ensembles. Looking at table 1 it seems like they have been changed with predefined steps. This is not the way GLUE (and any ensemble generation in general) works: one has to draw parameter values randomly from a specified distribution, and specify the amount of parameter combinations. This would allow sampling from a more-dimensional parameter space. If one works with intervals, the number of model runs increases to the same power as the number of parameters. This quickly becomes infeasible. I suspect that this is why only two parameters were used in the analysis.

- A detail: in the caption of figure 1, put a dot between "period" and "Black".

In the light of these comments, I do not recommend the paper for publication.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 2283, 2012.