

***Interactive comment on “Improving simulation of soil moisture in China using a multiple meteorological forcing ensemble approach” by J.-G. Liu and Z.-H. Xie***

**Anonymous Referee #1**

Received and published: 1 May 2013

Review of Liu and Xie

I recommend that this paper either be rejected or sent back to the authors for major revision. It has several flaws that would need to be addressed before it would be ready for publication:

1. What is the real bottom line? That if you test a land surface model, remove the bias in the simulations, and then weight simulations by the ones that already do the best, that you can get a good simulation? This is just curve-fitting and is obvious. What is the new science? That you give this procedure a fancy name, Bayesian Model Averaging?

C1361

2. The authors use four different input data sets, but do no analysis of those data sets. Why are they different? They each need to be validated with in situ observations before they can be trusted. They also need to be compared, with time series and maps. Some have identical data going into them, such as 2 and 3 both using surface station data. So what differentiates them? Obviously some are better than others. By forcing the model with some data sets that are not very good, and then given less weight to the bad results, you get a better answer – but how could you expect this not to be the case?

3. The paper needs to investigate why some simulations did better than others. This will follow from an analysis of the input data in its biases. If you did not force the model with poor data in the first place, you would not need to do corrections to the output.

4. The simulations were done incorrectly. The authors first did a long run with one of the forcings to 2011, but then used that balanced output to initialize simulations in 2005 with the four different forcings. This will obviously cause a drift due to spin-up for two reasons – the model will start with the incorrect soil moisture values, and the forcing (for the three not used in the first place) will produce a drift toward a different climate. Yet it seems that the authors used three years of this simulation to train the Bayesian model, without correcting for drift in at least several months at the beginning of the run. You can clearly see this drift in Fig. 5 for the 70–100 cm layer for several of the simulations.

5. Why was the CLM model used? How would the results depend on the land surface model used? Has this model done well in other intercomparisons? Why not use the well-tested VIC or NOAH models?

6. Why are there gaps in the results for three of the regions in Figs. 3–5? Even if the observations are missing, why are the simulations missing? Or were there gaps in the forcing data? If so, was the model restarted each time? From what initial conditions?

7. The English is pretty good, but there are a few mistakes. I attach an annotated manuscript with corrections, but also with comments on a number of other items, all of

C1362

which have to be addressed in any revision.

8. What time step was used for the simulations?

9. There needs to be a figure comparing the model levels and the observed levels, showing on a linear depth scale each of the layers of each. The text says observations were taken at 50 and 100 cm, but then they use data for 70-100 cm. How is that possible? I know that actually observations are taken for each 10 cm layer from 0 to 100 cm. The question is how they were saved and put into the archive – with what averaging.

10. This is a very valuable soil moisture data set. How can other scientists access the data to repeat these experiments? If they are not generally available, this is a problem for science. Are the data in the International Soil Moisture Network, <http://www.ipf.tuwien.ac.at/insitu/>? If not, why not?

Please also note the supplement to this comment:

<http://www.hydrol-earth-syst-sci-discuss.net/10/C1361/2013/hessd-10-C1361-2013-supplement.pdf>

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 3467, 2013.