Response to comments from Referee 1

October 26, 2015

Firstly, we would like to thank the reviewer for many constructive comments. The corrections for reviewer 1 are presented in red in the revised paper.

Response to major comments:

1

1.1

Referee comment

The ensemble-bias correction method may have done more harm than good. The WG2 bias of 4 mm was fairly small in the first place, and while the ensemble bias correction method did reduce this soil moisture bias, it has introduced very large biases in the fluxes in Figure 2. For many users the fluxes are of more importance than the states.

Response:

We agree with the comment that the ensemble bias correction did more harm than good. This was one of the conclusions of the paper. We acknowledge that we have not made clear that for many users the water fluxes can be as important as the soil moisture states - we propose adding this to the conclusion.

1.2

Since the problem arises largely when the ensemble perturbations generate model states outside of the usual bounds of the model (<wilt, >field capacity), it may be better to limit the ensemble members to not go outside these bounds, particularly for the field capacity (if there is no mechanism by which the model would dry the soil moisture below this point).

Response:

We propose answering this question in the discussion (Section 4.5): “In this study, the nonlinearity issues were most prevalent when the model descended below the wilting point or ascended above the field capacity. For this reason it may seem intuitive to introduce lower and upper bounds at these thresholds. However, water can be slowly lost through the leaves by cuticular conductance below the wilting point or though incomplete closure of the stomata. Soil moisture may also temporarily increase above the field capacity before it drains out (Mahfouf and Noilhan, 1996). These features are part of the NIT version of the ISBA-A-gs model. Therefore, imposing bounds would not be realistic. Boundary problems can also affect the analysis. For example, Ryu *et al.* (2009) used a bounded land surface model and found the model bounds were partly responsible for the positive perturbation bias in their study."

2

Referee comment:
The synthetic experiments are incorrectly designed. The observations were generated by running the model (single member?) with perturbed precipitation, then adding a random error. The same model is then used in the synthetic assimilation experiment, and if I understand the manuscript correctly (this is not totally clear), the same precipitation perturbations were used to perturb the ensemble in the assimilation experiments. Hence the same precipitation perturbations are used to represent errors in the observations (for the generation of the synthetic obs), and then to represent errors in the model (for the assimilation). Please review the literature on synthetic experiments to redesign these experiments. The use of an observation error just 10% the size of that in the real experiments also limits the relevance of these experiments.
Response:
We acknowledge that there was a mistake in the explanation of the design of these experiments. Please see the revised Section 2.6 for a full description.

For the synthetic experiments the in situ observations were not used, although the model was used for the same 12 sites. The truth was generated from a single model simulation. The WG1 observations were extracted from the truth with the addition of a random normally distributed observation error with zero mean and standard deviation equal to to 10% of the higher value used in the real experiments ($\sigma^{\mathrm{o}} = 0.05(w_{\mathrm{fc}} - w_{\mathrm{wilt}})$). The size of the observation error was small enough for the observations to have a noticeable impact on the analysis.

A perfect model was used for the DA. However, errors were introduced in the precipitation forcing by adding random hourly perturbations ϵ_{Pr} to the hourly precipitation accumulations (Pr):

$$Pr* = Pr + \epsilon_{\mathrm{Pr}}, \tag{1}$$

where $Pr*$ is the perturbed hourly precipitation. The hourly perturbations were randomly sampled from a normal distribution with standard deviation equal to 50% of the hourly precipitation and zero mean. The probability distribution function (pdf) was truncated in order to prevent negative values of $Pr*$ and to maintain a mean of zero ($-Pr \leq \epsilon_{\mathrm{Pr}} \leq Pr$). A single precipitation time series from Eq. (1) was generated over 2007-2010. This was used to force the model.

The SEKF has no means of capturing the uncertainty in the precipitation forcing from Eq. (1) directly. Therefore it was necessary to calibrate the **B** matrix to capture the background errors that resulted from the precipitation errors. The background-error variances (σ^{b}) were calibrated with values a 10th of the values used for the real experiments, since the open loop errors in the synthetic experiments were about 10% of the errors in the real experiments.

In the synthetic experiments, the EnSRF ensemble members used the single precipitation time series from Eq. (1) to force the model. However, each ensemble member was then perturbed using Eq. (1) with different random seeds for each member. This enabled the EnSRF to capture the uncertainty in the precipitation forcing directly.

This experiment was designed to assess the advantage the EnSRF could gain over the SEKF by using a perfect stochastic representation of random precipitation forcing errors. The results clearly demonstrated an advantage in the EnSRF performance for this idealized regime (perfect model and small observation errors). We did try a much larger observation error originally, but the impact of observations on the analysis was noticeably smaller.

3

Referee comment:
The manuscript assigns all of the difference in results between the sites to differences in soil class, however other differences between the sites that may affect the assimilation are not accounted for (differences in climate!). Also, it looks like the conclusions regarding soil class are made from comparing just two sites. The statements about the role of soil class need to be greatly de-emphasized (including removed from the title and abstract).

Response:
This is a very good point. We recognise now that we did not investigate the impact of climate on the results, nor did we properly investigate the impact of soil texture over the 12 sites. Therefore, we performed extra experiments to determine whether the differences in the ensemble perturbation bias between the sites can be partly attributed

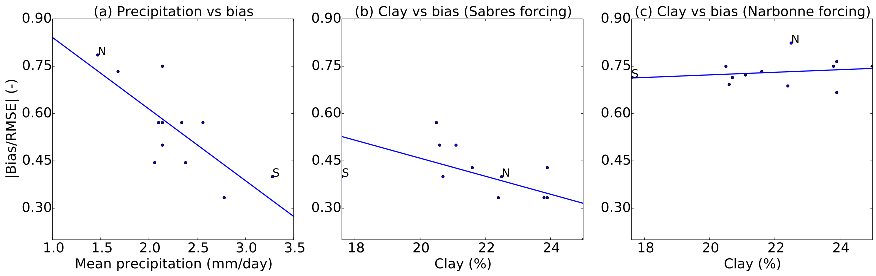

Figure 1: (a) Average perturbation bias normalized by the RMSE against average daily precipitation for each site; (b) Average perturbation bias normalized by the RMSE against clay percentage for each site, with the climatological forcing of the Sabres site applied to all the sites; and (c) same as (b) but the climatological forcing of the Narbonne site has been applied to all the sites. The Sabres and Narbonne sites are labelled as ‘S’ and ‘N’ respectively. The line of best fit (linear regression) is shown for each plot.

to soil clay content or to precipitation. We propose adding these results to Section 3.1.2. We clearly demonstrated that precipitation suppresses the bias for the 12 sites, while clay content has little influence. A scatter plot of the average daily precipitation against the normalized bias is shown in Figure 1(a). The linear regression line shows a strong negative correlation between the precipitation amount and the magnitude of the perturbation bias.

We then performed an experiment to determine the impact of clay content on the bias. In this experiment we used the same atmospheric forcing of the wettest site (Sabres) for all the sites. This eliminates the impact of different climate on the results and leaves only differences in soil class. The clay percentage is plotted against the perturbation bias in Figure 1(b). We then repeated the experiment in Figure 1(b) but instead using the same atmospheric forcing of the driest site (Narbonne) for all the sites. The results are shown in Figure 1(c). Neither Figure 1(b) nor 1(c) show a strong correlation between the clay percentage and the bias. On the other hand, the perturbation bias for the drier climate in Figure 1(c) is much greater for all the sites than for the wetter climate 1(b). These results demonstrate that precipitation acts to suppress the perturbation bias, while clay content has little influence on the bias for these 12 sites.

These results contradict the original conclusion we made that clay content was the main factor influencing the size of the bias. We propose correct the conclusions of the paper accordingly. We will discuss the reasons for these findings in Section 4.3. We propose replacing “contrasting soil conditions” with “contrasting conditions” in the title.

4

4.1

Referee comment:
The evaluation statistics need refining, and better description: -P7367: Edit these equations to make a clear distinction between the assimilated observations, and the observations used for evaluation.
Response:
Agreed.

4.2

Referee comment:
The fact that the evaluation observations are from the same network as the assimilated observations (but at a different depth) needs to be very prominently acknowledged in the text. This must include a discussion of any potential dependence between the assimilated and evaluation observations.
Response:
Agreed. We will mention this in the conclusions and propose adding that the "triple colocation" re-scaling approach might also be more appropriate than CDF matching for independent data sources.

4.3

Referee comment:
The ACC presented here is the correlation, not the anomaly correlation. Change all reference to anomaly correlation to correlation, and include a true anomaly correlation in the evaluation.
Response:
We mistakenly wrote the equation for the ACC as the CC. In fact the equation used for the ACC in the experiments is the correct one.

4.4

Referee comment:
P7367 implies that the observations used for the evaluation were CDF-matched. This should have been clearly stated with the introduction of the evaluation metrics.
Response:

Agreed.

4.5

Referee comment: There is also some discussion of the introduction biases in the RMSE due to inconsis- tencies in time period why not use consistent time periods?
Response:
We will add the following to Section 2.4.1: "The CDF matching was performed over 2007-2010. The results were calculated over 2008-2010 because a one year spin-up was used in the experiments. The small bias that remains, as a result of the different time periods, is not significant."

4.6

Referee comment: Also, presenting biases to the in situ data (or RMSE, which is not bias robust) for the different experiment is not very informative since the in situ data were arbitrarily rescaled. It is more usual practice to rescale each experiment separately to the in situ observations before comparing them for evaluation (thus removing the mean difference between each experiment and the in situ data).
Response:
Since the WG2 observations (used for evaluation) were re-scaled in the same way as WG1 (used for assimilation), we assume that the bias and the RMSE are meaningful. We wanted to show the ensemble perturbation bias and re-scaling the analysis by the in situ data would remove it. The ACC is bias robust and was therefore not affected by the bias nor the re-scaling.

4.7

Referee comment: Before proceeding to investigate the role of model physics in generating the bias, confirm that the perturbation time series is mean-zero (see comments below: re the precip perturbations). Include a note in the text that this has been confirmed
Response:
The perturbation time series is mean-zero by design. We have even confirmed this by performing a separate experiment. We will include a note in the text.

5 Response to minor comments

Response to minor comments:

1. **Referee comment:** It was several pages in before I realized that in situ observations were assimilated in this study. This needs to be stated clearly in the abstract and introduction.
   Response: Agreed

2. **Referee comment:** P7355, L6. Soil moisture assimilation is not the main objective of DA. Many other variables are assimilated. Please rephrase.
   Response: "An important application of data assimilation (DA) for land surface models is to assimilate observed surface soil moisture to produce an analysis of root-zone soil moisture"

3. **Referee comment:** P7355, L35. Uncorrelated with what?
   Response: "uncorrelated between layers and gridpoints"

4. P7355, L27: change with an NWP model to with an NWP model at Meteo-France
   Response: OK

5. P7356, L2: specify that the SEKF at ECMWF assimilated screen-level variables, and not soil moisture.
   Response: OK

6. P7357, L23: Remove Calvet and Noilhan reference if they did not do CDF-matching (i.e., match the full CDF, not just the mean and variance)
   Response: The term "CDF-matching" was originally used to include higher-order moments (Reichle *et al.*, 2004; Drusch *et al.*, 2005). However, the term "CDF-matching" has also been used in other studies when applying re-scaling to just the first two moments (see e.g. Scipal *et al.* (2008); Barbu *et al.* (2014)). Calvet and Noilhan (2000) also used this approach but did not use the term "CDF-matching". We suggest using the following phrase in Section 2.4.1: "A linear CDF matching approach was employed in this study, which scales the observations such that the mean and the variance match that of the model (Calvet and Noilhan, 2000; Scipal *et al.*, 2008). This is a linear approximation of the CDF matching approach of Reichle *et al.* (2004); Drusch *et al.* (2005), which uses

higher order moments."

7. P7359, last paragraph: it is not clear here why only grassland was used. Is this the land cover at all of the SMOSMANIA sites? Include a map or table with the SMOSMANIA site locations.
   Response: Yes it is the only land cover type used as vegetation within the weather station boundaries mainly consist of grasses. We will make this clear and will add a map of the SMOSMANIA stations and the average precipitation for each site (Figure 1).

8. P7361, L19: CDF-matching also matches the higher order moments. Please rephrase.
   Response: Please see comment above.

9. P7363, L5: H does not equal [1 0] for the SEFK, since there is also a time integration. Please edit.
   Response: We will mention that H is applied at the end of the window, which is at the analysis time.

10. Equation 4 needs time indexes for x and Delta x and M. The l superscript on Delta x is also not defined. State in the text that equation 4 requires an extra model run for each element in the state update vector. State at what time the analysis update is made (at the start or end of the perturbed runs for equation 4).
    Response: OK

11. P7363, L19: kl superscript is not defined.
    Response: Will define

12. A Jacobian > 1 does not itself indicate a non-linear model. This is implied in several places, please rephrase.
    Response: OK

13. P7366: the presentation of the SEnKF needs equations for the analysis update (not just n del xa. Also the j subscript is not defined.
    Response: Agreed.

14. P7368: Since both assimilation methods are tuned using the same data as presented for the evaluation (rather than independent data) it must be prominently acknowledged that the presented results will not necessarily generalize to sites

where data are not available for calibration.
Response: Agreed (see response to major comment 4.2).

15. L7368, P20. I cannot make sense of the discussion here. The ensemble spread is used to represent the background error. This will include estimation of forcing and model errors if the ensemble is appropriately perturbed. Also, here(and elsewhere): The references to specifying or tuning the error covariance matrix for the SEnKF is misleading, as this the matrix itself is not specified (or even estimated) for the En methods. Please rephrase throughout to avoid referring to the matrix for the En methods (and replace with something like ensemble spread).
Response: Agreed. We will re-phrase this part.

16. L7369: The precip was perturbed using Gaussian noise with standard deviation of 50%, I assume these were additive perturbations. It is not stated over what time period of precip the 50% is taken. It might be worth rethinking this I am concerned that the long term mean of these perturbations may not be zero, since the standard deviation of the added noise is not stationary. A lognormal multiplicative perturbation would be more appropriate. Also 50% of precip is a very large perturbation.
Response: Please see major comment 2 regarding the synthetic experiment design. In fact the 50% magnitude comes from Reichle *et al.* (2002) and is based on order of magnitude approximations (by comparing the errors with the difference between two different data sets). The perturbations are multiplicative in the sense that they are a percentage of the precipitation amount itself. They are also additive in the sense that the perturbations are added to each ensemble member. This is standard practice for precipitation perturbations. We agree that a a lognormal perturbation would be more appropriate than a normal perturbation - we will add this information in the conclusion.

17. Equation 14: introduce notation to distinguish between the bias corrected and original xb
Response: OK

18. P7371, L6. I would not say that these biases are unexpected (only unexpected / or inconsistent under the linear assumption). The model is non-linear and soil moisture is bounded, so these biases are to some extent expected.
Response: Agreed - we will remove this.

19. Split Table 2 into separate tables for synthetic and real-data experiments.
Response: We would prefer not to change this because having the results in

the same table allows the reader to compare the performance of the DA methods between different experiments.

References

A.L. Barbu, J.C. Calvet, J.F. Mahfouf, and S. Lafont. Integrating ascat surface soil moisture and geov1 leaf area index into the surfex modelling platform: a land data assimilation application over france. *Hydrol. Earth Syst. Sci.*, 18:173192, 2014.

J.-C. Calvet and J. Noilhan. From Near-Surface to Root-Zone Soil Moisture Using Year-Round Data. *J. Hydrometeor*, 1:393–411, 2000.

M.K. Drusch, M.E. Wood, and H. Gao. Observation operators for the direct assimilation of TRMM Microwave Imager retrieved soil moisture. *Geophys. Res. Lett*, 32:L15403, 2005.

J.-F. Mahfouf and J. Noilhan. Inclusion of gravitational drainage in a land surface scheme based on the force-restore method. *J. Appl. Meteorol.*, 35:987992, 1996.

R.H. Reichle, R.D. Koster, J. Dong, and A.A. Berg. Global Soil Moisture from Satellite Observations, Land Surface Models, and Ground Data: Implications for Data Assimilation. *J. Hydrometeor*, 5:430442, 2004.

R.H. Reichle, J.B. Walker, R.D. Koster, and P.R. Houser. Extended vs Ensemble Kalman Filtering for Land Data Assimilation. *J. Hydrometeor*, 3:728–740, 2002.

D. Ryu, W.D. Crow, X. Zhan, and T.J. Jackson. Correcting Unintended Perturbation Biases in Hydrologic Data Assimilation. *J. Hydrometeor*, 10:734–750, 2009.

K. Scipal, M. Drusch, and W. Wagner. Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Adv. Water. Resour*, 31:11011112, 2008.