Reviewer 1 (R#1)

The study describes results of a data assimilation experiment, assimilating soil moisture data of the Soil Moisture Active Passive mission into the UK land surface model JULES. The assimilation updates states and parameters. Resulting soil moisture is compared to SMAP data and data of an independent network of cosmic ray neutron probes.
The title and general content of the manuscript are promising, while the manuscript itself exhibits lack of detail which would be required for following the study and reproducing the results. Below, my concerns, starting with the general ones, and followed by detailed comments.
We thank the reviewer for their comments which will undoubtedly help to strengthen this manuscript. We outline below our responses and proposed changes.

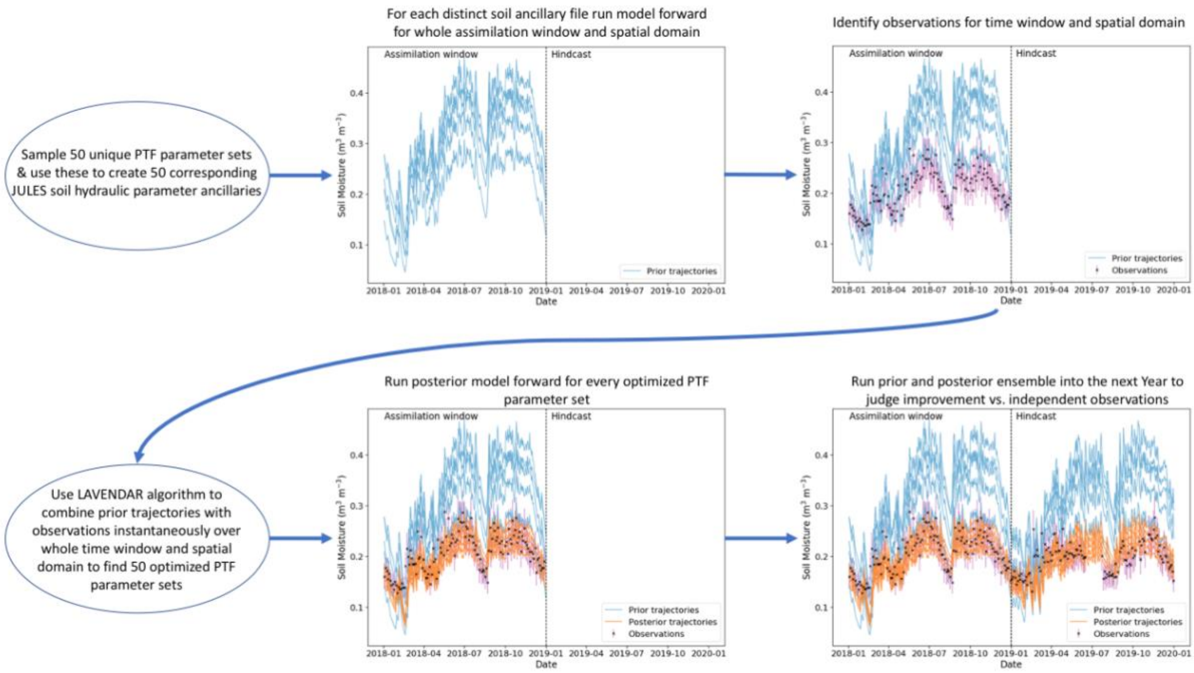
1. Well known bias in the SMAP satellite product and impact on pedotransfer functions is not discussed (e.g. Reichle et al. https://doi.org/10.1029/2019MS001729 or Colliander et al 2017 https://doi.org/10.1016/j.rse.2017.01.021). This would be a key asset of the paper.
This is a very good point. As per the papers mentioned, if the SMAP product is biased high there could possibly be an impact on the retrieved pedotransfer function (PTF) parameters. This would likely exhibit itself in PTF parameters that would artificially increase the values of the saturated soil moisture and possibly decrease saturated conductivity given the underlying soil textural information. The comparison to COSMOS probe data should also allow us to comment further on this and whether the bias in SMAP has had a significant impact on the posterior model skill. We will add this discussion and further analysis into the paper alongside the stated references.

2. Which SMAP level data was used. It will help the reader in understanding the results. Please point this out in the introduction and methods sections. What are the implications?
We used the L3 SMAP v3 9-km radiometer-radar combined product. We will include this and possible implications as requested.

3. Discussion is not based on literature but merely on own postulations. A good guide is located here: https://www.biosciencewriters.com/How-to-Write-a-Strong-Discussionin-Scientific-Manuscripts.aspx
We agree the discussion could be strengthened and will endeavour to do so (see later related points).

4. Please add conceptual details on how the 4DEnVar (an optimization method) is combined with EnKF (optimization) (see page 7 lines 159-164). I imagine this can be done by text or together with a figure. Also address why are both optimization methods combined at all?
5. Please add how is the state vector in Appendix A is composed in the present case (variables, parameters, IenKS posterior?) and which units do the variables in Appendix A have.
6. Please clarify, what are prior and posterior with respect to two data assimilation methods? How can posteriors be worse then priors considering that the results are optimized using the evaluation data? Please plot as well the data assimilation performance over time with regard to RMSE and parameter convergence as for example in Poterjoy et al. 2017 https://doi.org/10.1175/MWR-D-16-0298.1, Botto et al. 2018 https://doi.org/10.5194/hess-22-4251-2018 and Baatz et al. https://doi.org/10.5194/hess-21-2509-2017 .
7. Please add results after the 4DEnVar assimilation in order to demonstrate what an additional assimilation yields in terms of skill.
We have grouped together points 4-7 here as we believe these all stem from us not adequately describing the data assimilation technique used in the current manuscript. We have referenced a previous paper centred around the development of the technique and have not supplied enough information here for readers to properly understand what we have done.
4. 4DEnVar is not combined with the EnKF, 4DEnVar is a hybrid technique combining elements of both ensemble and variational data assimilation methods. This is done in practice because we want to use

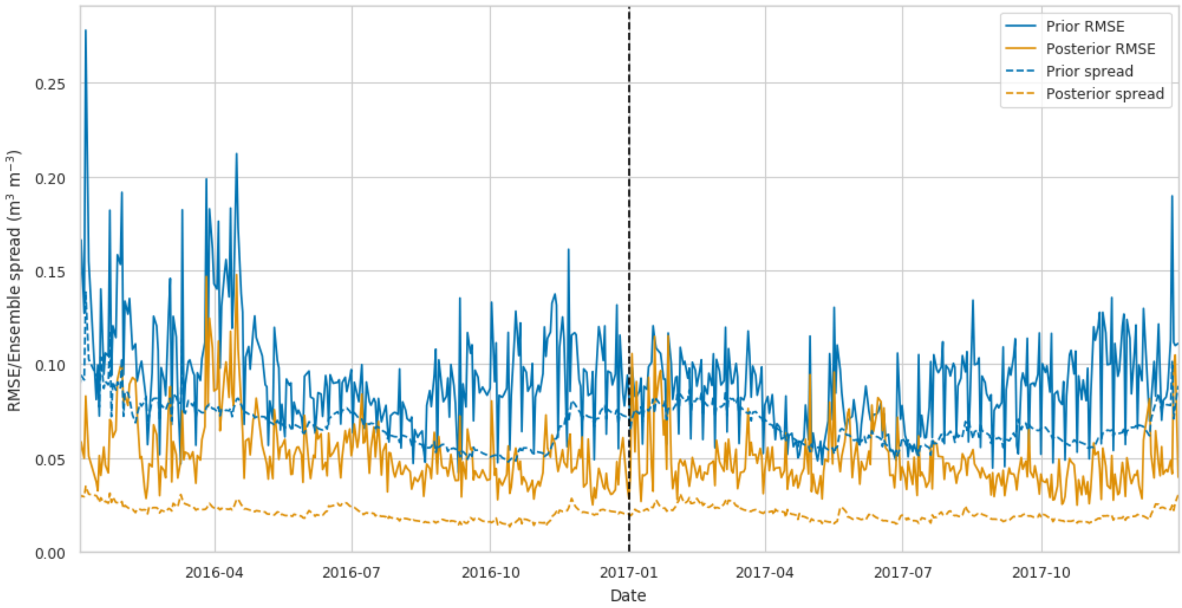
a variational technique here, combining all observations over a time window with a prior prediction, to retrieve a set of optimised parameters that do not vary in time (which you would retrieve from a technique such as the EnKF or other sequential methods). However, the majority of variational techniques require the adjoint and derivative of the model code. We do not have this for JULES and it is very costly to compute. 4DEnVar, the IEnKS and other related hybrid methods allow us to approximate this adjoint from an ensemble of model runs. On reflection the way we have described this in the manuscript is not clear and we agree that the use of a diagram could be beneficial to illustrate the technique (see example diagram below). The description shall be improved and a diagram added.



5. In Appendix A the state vector is just the vector of 15 PTF parameters as defined in section 2.2 Table1 We will include this in the Appendix and make it clearer as to what the different variables relate too.

6. There is only a single assimilation step being used, we will aim to make this clearer. In Figures 4 to 7 the prior is just the mean and standard deviation of the 50 prior JULES ensemble members before DA and the posterior is the mean and standard deviation of the 50 posterior JULES ensemble members after DA. We are optimizing 15 PTF parameters for the whole time window (<28000) and the whole spatial domain (<30000 gridcells) in a single assimilation step by minimising a cost function. This is unlike sequential methods such as the EnKF or ETKF which step through time updating estimates at each step with available observations. We retrieve a single set of 15 PTF parameters valid over the whole domain and for the whole time period. This means that the optimisation may have to degrade the fit at certain locations to allow the 15 PTF parameters to improve the picture as a whole. This could be due to errors at these locations in driving data, the underlaying soil property map or indeed in the model structure (as is the case over urban areas in our results). As the DA method here is fundamentally different from the techniques in the papers mentioned, we are not able to reproduce the stated plots for parameter convergence as we retrieve just one set of parameters valid for the whole time window. However, we can plot the RMSE over time for both the prior and posterior ensemble members (see plot below). We will also aim to increase the distinction between previous

sequential DA methods and the variational method we have used for this paper.



7. We believe this is already shown and hopefully once we have strengthened the description of the DA algorithm will become more clear.

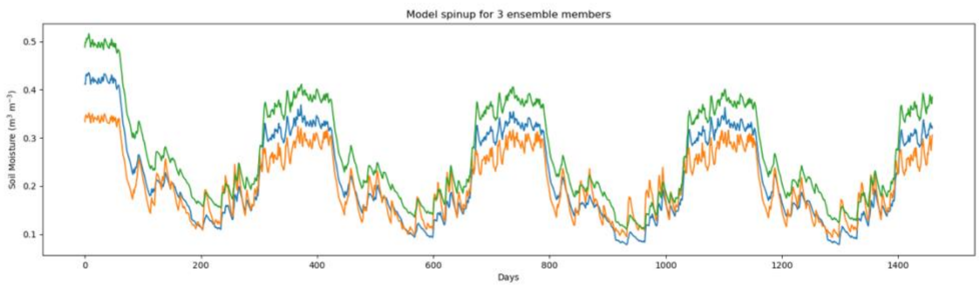
8. Please expand on why to add another 1% SWC error to SMAP (from 0.04 to 0.05 cm3/cm3, page 6 line 123) and multiply by four (20cm3/cm3 error?) for observation inflation, a rather seldomly used method. Inflation is rather used for covariance inflation during the run time of the data assimilation experiment (e.g. Jamal and Linker 2020, https://doi.org/10.1002/vzj2.20000 or Whitaker et al. 2011 DOI: 10.1175/MWR-D-11-00276.1). Please cite more studies were observation inflation is directly used and discuss why a bias aware data assimilation method was not used (e.g. Ridler et al. 2018 https://doi.org/10.2166/nh.2017.117)
Although the baseline aim for SMAP is 0.04 cm3/cm3 other studies have found higher values; 0.043 (Colliander et al 2017 https://doi.org/10.1016/j.rse.2017.01.021), 0.054 (Zhang et al. https://doi.org/10.1016/j.rse.2019.01.015), 0.054 (Li et al. https://doi.org/10.3390/rs10040535). We therefore chose a value between these studies of 0.05 as a form of expert elicitation. Although observation error inflation is seldom used in sequential data assimilation it is quite common place in variational methods (such as the one in this paper) and especially in numerical weather prediction (Wang et al. http://dx.doi.org/10.1029/2019JD031029, Bormann et al. http://dx.doi.org/10.21957/gq8j2gjp7, Fowler et al. https://doi.org/10.1002/qj.3183, Hilton et al. https://www.ecmwf.int/node/15331). The observation error inflation is required due to the fact that all observations are used at once in the assimilation whereby we minimise a cost function containing a prior term and an observational term. The greater the number of observations in the observational cost function term, the higher the weight they have in the optimization. This can lead to the prior term being completely negated and hence the retrieval of unphysical parameters. Observation error inflation would not be required if the correct specification for the observation error correlations (in space and time) and model error was included. These, however, are hard to diagnose and it has been shown that in the absence of such information inflation is required for an optimal DA system (Stewart et al. https://doi.org/10.1002/qj.2211). It has also been shown that for variational DA model errors can be included in the observational cost function term by inflating the diagonal variances, Howes et al. https://doi.org/10.1002/qj.2996. We will include further references to this in the text and strengthen the discussion around the inflation. Hopefully the improved description of the DA technique will also help here and the distinction between sequential and variational DA. Although we

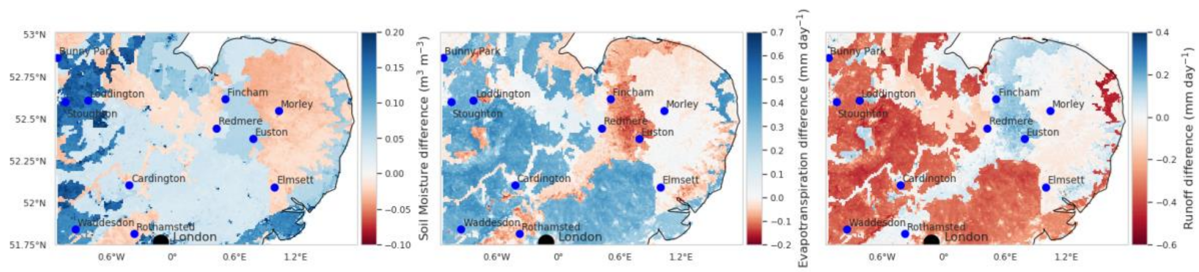
agree a bias aware data assimilation could be more optimal, the one proposed is in relation to a sequential technique (the ETKF) and we are using a variational method.

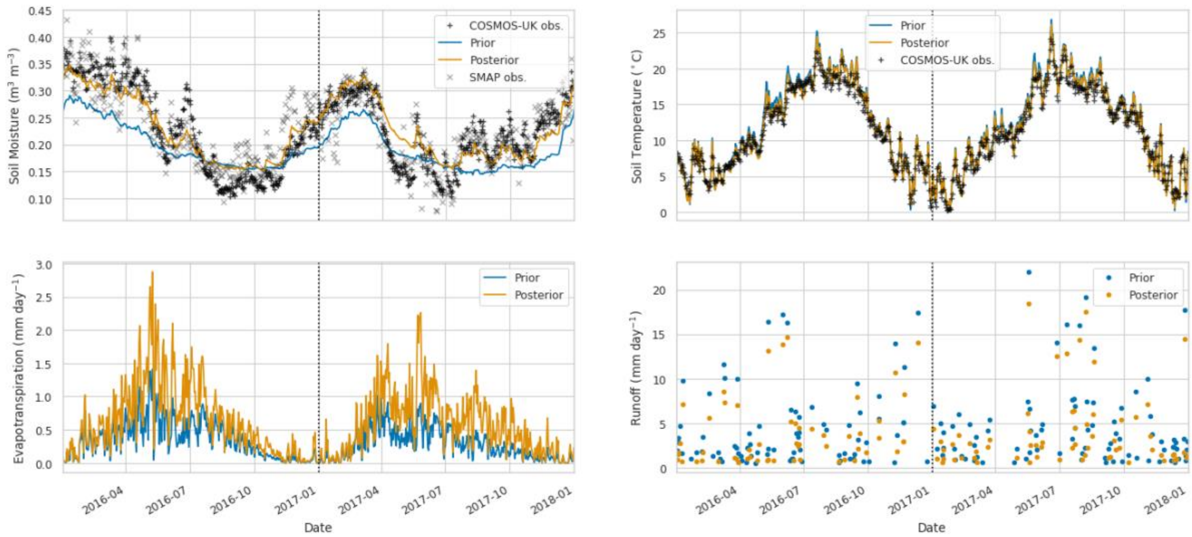
9. Please add legend to the graphs (Figure 6, 7 etc.).
We will add legends as requested.

10. Please discuss cross-correlation among the parameters of pedotransfer functions. From Equation 1 in the author's paper, it is clear that many parameters cross-correlate. Take for example Phi a and Phi c crosscorrelate strongly. What is the impact on saturated soil hydraulic conductivity?
We agree added discussion on this would be beneficial. It is also possible that the inclusion of such correlations could improve data assimilation results and we have shown this in previous examples Pinnington et al. https://doi.org/10.1016/j.agrformet.2016.07.006 . As this is a first attempt at DA with pedotransfer functions this has not been investigated yet but comment on this will be added.

11. Please expand on the JULES hydrologic water components (ET, ground water, surface water flow, overland flow, infiltration, snow). How exactly was the 4 year spin up done? Was it done in ensemble mode? How were parameters perturbed? Please provide groundwater and soil moisture development over time at four cosmic ray neutron probe locations during the spinup period to elucidate the reader about the spinup performance.
We will add plots of other water budget components and details on the spin up to the text. The spin up is done for each prior and posterior ensemble member, with the parameters either being sampled from the defined prior distribution or as outputs from the DA system in the case of the posterior. The model is run from an initial value (defined by the saturated soil moisture model parameter) over the same year of forcing data to reach an equilibrium soil moisture state for any given set of parameters. The plot below shows this for three distinct ensemble members which are all defined by unique sets of PTF parameters. We can see how these unique realisations of PTF parameters define unique soil moisture trajectories. The JULES model does not contain a groundwater component in the current configuration but we will add spinup plots for soil moisture and other relevant variables to the text. We will also add plots of the other model water components (see below).







12. In this realm, a discussion of main characteristics, limitations and specifics of the study area with regard to SMAP data is essential to understand the manuscript. This would include addressing topography, land cover, other factors.
We will included a broader description as requested.

13. Equation 1 – please list the units of the parameters in these physical equations.
To be included.

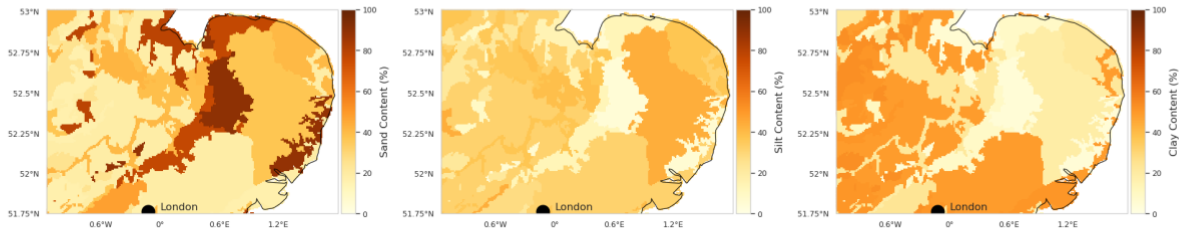
14. Page 7 line 145 – why did the authors chose 10% standard deviation when it is well known that many van Genuchten parameters and soil hydraulic conductivity is logarithmic scale. What does 10% standard devation mean? Does it mean 0.63+/-0.063 for phi a and 0.0003 +/-0.00003 for phi c for example?
The reviewer is correct in their example of a 10% standard deviation, this is used to define a Gaussian distribution that 50 unique parameter sets are sampled from. It is true that van Genuchten and soil hydraulic conductivity parameters can be described by logarithmic distributions, but it is less clear what the best distributions are for the PTF parameters that are used to calculate the van Genuchten and soil hydraulic conductivity parameters. We therefore made a naïve assumption of a 10% standard deviation for our prior distribution and did not look further at this as we achieve good results when compared to in-situ COSMOS probe data. It is an important point that this is an area that could be investigated further in future studies and we will make sure this is communicated within the manuscript.

15. Why did the authors not use a known weighting function for JULES soil moisture to compare with cosmic ray neutron sensors. Köhli et al. 2014 https://doi.org/10.1002/2015WR017169 Baatz et al. https://doi.org/10.5194/hess-21-2509-2017 or Shuttleworth et al. 2014 doi:10.5194/hess-17-3205-2013 provide already well tested methods. How does the author's method compare with these results?
Apologies we did not make this clear; the COSMOS-UK network does use the method of Köhli et al. 2015 https://doi.org/10.1002/2015WR017169 to diagnose SM and relative depth of the COSMOS probe measurements. We then use a simple operator on this information and the JULES model output to compare to the COSMOS SM estimates. This operator was developed as part of the Hydro-JULES project by colleagues at UKCEH by Cooper et al. https://doi.org/10.5194/hess-2020-359. We will highlight this in the text.

16. Aside, Desilets and Zreda, 2013 doi:10.1002/wrcr.20187, 2013 consider the diameter being 600 meter, not the radius.

Noted will update.

17. Figure 2: please add a map of soil textures. Please discuss the sharp light blue – dark red gradient at 0.9E. Is this an artifact from data assimilation?
The adding of soil texture maps is a great idea and will help with interpretation of the results (we include these below also). We can see that the dark red gradient at 0.9E in Figure 2 is a result of a distinct area of soil texture in the HWSD and how this is responding to the pedotransfer functions.



18. Page 9 line 196 – adding London in all maps for the non UK citizens would be a great asset.
Noted, see above.
19. Page 10 line 206 – please define observation operator and outline the details on how this operator was developed, calibrated and validated. There are existing operators already (see point 15).
We have addressed this in point 15 and will ensure the operator is further defined.
20. Page 13: Please separate discussion and outlook clearly. The authors use repeatedly phrases on future work e.g. 'work is being undertaken' (line 238), 'we will' (line 241), 'is possible' (line 244) 'could be' (line 245) 'it may' (line 247) and so on. . . Also references to e.g. GRACE are missing.
Noted will split into subsections.

21. Also, a discussion on literature with previous published assimilation experiments on soil hydraulic parameters will be useful. Here, the paper can give a valuable contribution to exhisting literature. Especially considering the authors going the extra step to assimilate often cross correlated parameters of pedotransfer functions.
We agree including additional literature is important here. Also, given responses to the previous comments on how this DA method differs from previous examples, we will make sure to include the cited literature and add comment.

22. Figure 11: Symbols with a center point are more precise and clearer than circles. Please use smaller dots, or even better symbols with a center point such as +,*, x and use different symbols for Cosmic Ray Calibration data and SMAP data points. Also please add SMAP soil moisture to the plots with cosmic ray neutron probe data, although these are not the equivalent depths as cosmic ray neutron probe soil moisture.
We will update plots accordingly (see subplot included for the Cardington site above).