Reviewer 3 (R#3)

The paper explores the use of SMAP soil moisture products with the JULES land surface model with a data assimilation framework. The framework is applied in a region of the UK where soil properties from pedotransfer functions are constrained with data assimilation. The topic has potential and the paper started very well with its Introduction and Methods sections. However, I found the Results and Discussion very weakly presented, without in-depth analyses and implications. It is not clear what is the lessons learned and how it can benefit the wider community. In addition, these two sections read much more like a technical report. There are many additional tests that can be made to improve this study (I've made some suggestions). For that reason, I believe this paper manuscript requires considerable changes, hence I recommend major revisions before making my decision on its acceptance.
We thank the reviewer for their comments and hope to be able to strengthen the paper in line with their specifications. Below we present our responses to comments and proposed changes.

List of comments:
L63-64: Notice there are several approaches that constrain model parameters that do account for uncertainties, please refer to works by Keith Beven, Jim Freer, Jasper Vrugt, Grey Nearing, Hamid Moradkhani, Martyn Clark; to name a few.
We agree that it is beneficial to mention such studies with relation to the catchment scale.

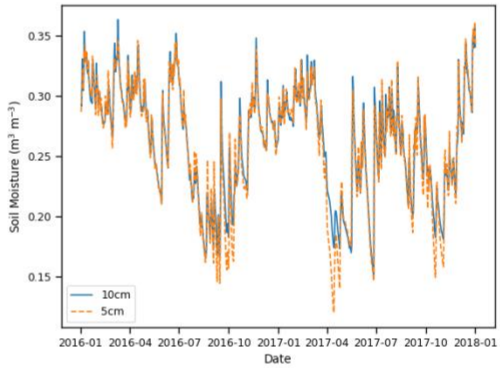
L64-65: First, can the authors please point out the references for the 'Previous studies' mentioned in the sentence?
Apologies, these were included in the previous sentence but missed here.

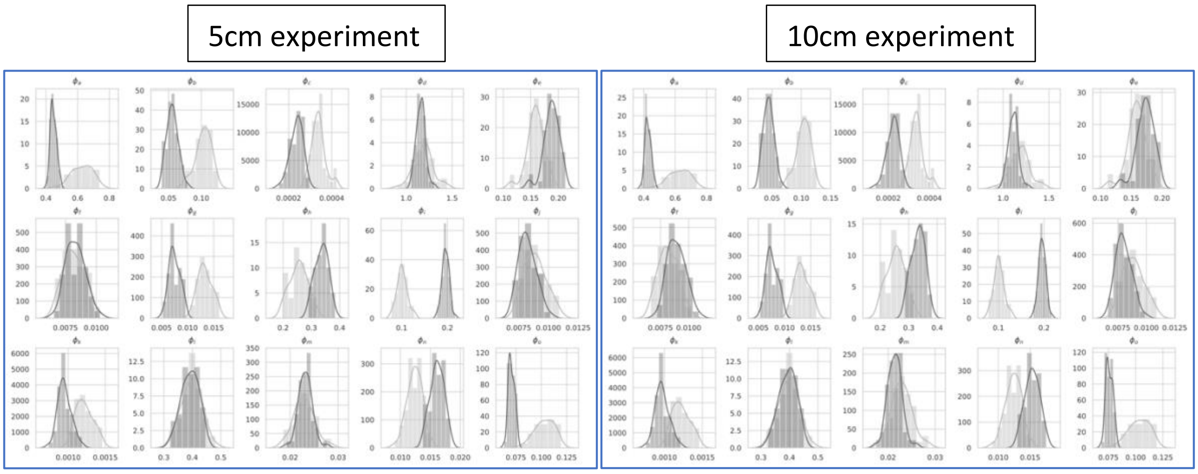
L65-66: Note that usually, the term data assimilation has been used in different ways by the atmospheric sciences and land surface modeling community in relation to the hydrological modeling community. 'Data assimilation' in general refers to using/fusing observed quantities to better constrain model components (i.e., parameter, states, etc...). Typically, the use of 'parameter estimation', 'state estimation', or 'dual parameter-state estimation' would be more clear. The reason I am mentioning this is because, although not technically a classic data assimilation application, the group by Luis Samaniego in UFZ Germany has explored similar approaches to this one using their mHM with their MPR framework. Additional work 'assimilating' both state and parameters include groups from Harrie-Jan Hendricks-Franssen, for example.
We agree. We are coming at this problem from a different background and so accept it may be helpful to update the wording here to make things more clear for the reader. We will also discuss the work and approaches you mention more in the manuscript discussion.

L79-81: This seems to be related to Results, not sure why it is included at the end of the Introducion section.
We will move this to relevant section.

L94-95: The direct information obtained from SMAP is typically for the first few centimeters of soil; yet your JULES model is configured with a relatively thick initial soil layer and only 4 layers in general. Have the authors considered revising their soil layers in JULES? Have they done any simple sensitivity study to check how influential the choice of soil layer discretization is when assimilating SMAP data. If I recall correctly, CLM (which is similar to JULES) is run with a much finer soil layer discretization.
Reviewer #2 had a similar concern (comment number 1). We chose to keep JULES in its default soil layer configuration so that the optimized soil ancillaries would be useful to the wider JULES community. We have run JULES with a 5cm top layer to confirm that the vertical variability at this depth in the model is not too great (see below). We have also re-run the entire experiment using this

soil layer to confirm that we retrieve very similar parameter distributions from the DA procedure for both the 5cm and 10cm soil depth JULES models.





L113-115 and Table 1: It is unclear to me how the prior is used. Don't you need an emsemble (i.e., range) for each prior factor shown in this Table? How is a single prior applied in this case?
For our prior we have 50 realisations of the parameters in Table 1. Each realisation is drawn from a normal distribution defined by the mean (shown in Table 1) and standard deviation (taken as 10% of the mean). This gives us the light grey distributions shown in the plots above. We will expand the text here to improve this description.

L138-140: Can the authors be more specific about this? There are many studies that have used the COSMIC operator which is available (refer to works by Jim Shuttleworth, Rafael Rosolem, Harrie-Jan Hendricks-Franssen, as examples). Have the authors consider implementing this operator?
Section 2.6: Needs to be expanded as it is very vague and general.
Apologies, we should have added more here and also referenced the works noted above. 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 (Our paper also forms part of this project) by colleagues at UKCEH (Cooper et al. https://doi.org/10.5194/hess-2020-359). We will highlight this in the text.

Figure 3: Typically, DA are justified as an operational tool for models (in the case of state estimation). This figure here shows the Bayesian optimization approach (prior –> likelihood –> posterior) which is fine. However, I'd be interested to see the time-series of the final soil parameters (produced with the updated pedotransfer function) to check for any inconsistencies in the way a particular parameter change from time to time. I'd expect soil properties to be fairly constant (relatively to the fluxes and states in the JULES model). Also, the authors should consider checking which of the PDFs shown in the

figure are expected to be significantly different. One way to do this is for example by checking whether two samples come (or not) from the same probability distribution. This can be easily done with a two-sample Kolmogorov-Smirnov test.

There is no optimization through time here. All the data is used at once over the whole time window and spatial domain to find optimized values for the 15 pedotransfer function (PTF) parameters valid in time and space. In this way we are avoiding the physically unrealistic artefact of time-varying parameters for the PTF's but also for the final soil hydraulic parameters. We do not believe we have made the DA approach clear enough in the paper as R#1 (comments number 4-7) also had issues in understanding what had been done, also assuming we were using a sequential assimilation method. This is a variational method and we will try to bring this out in the updated manuscript by also including a diagram (see below). Our technique is analogous to carrying out a single step of a sequential method such as the EnKF but using all information from observations and model dynamics for the whole time period and spatial domain to find a single set of PTF parameters that are valid in time and space.

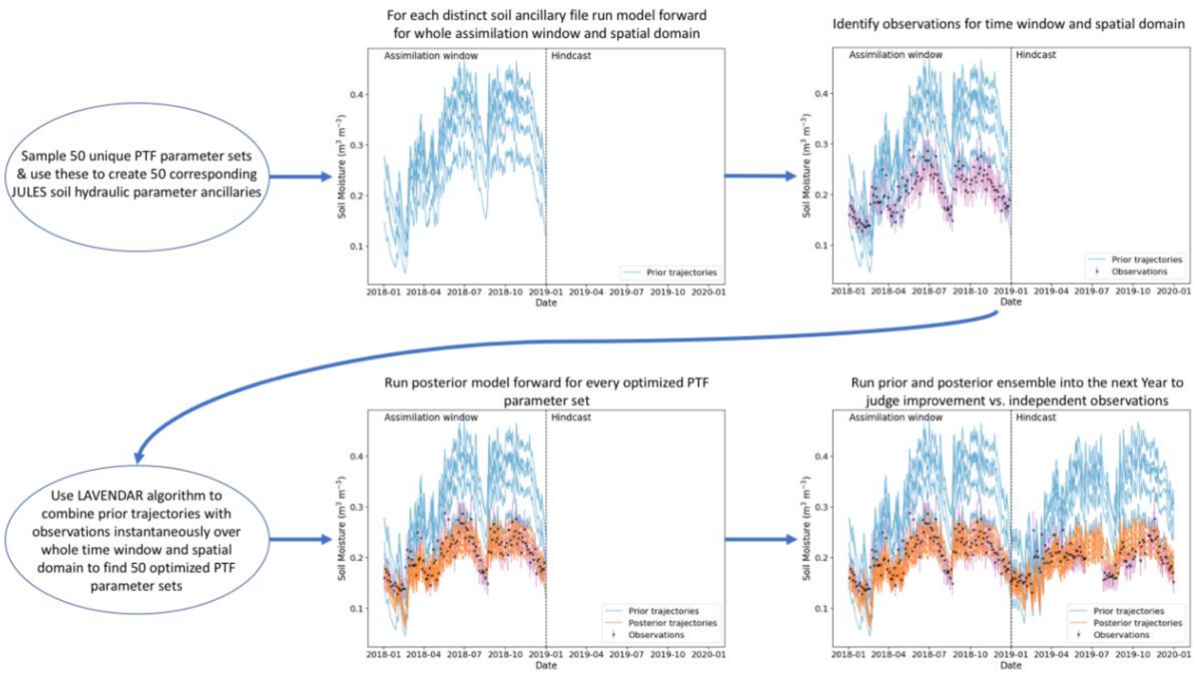
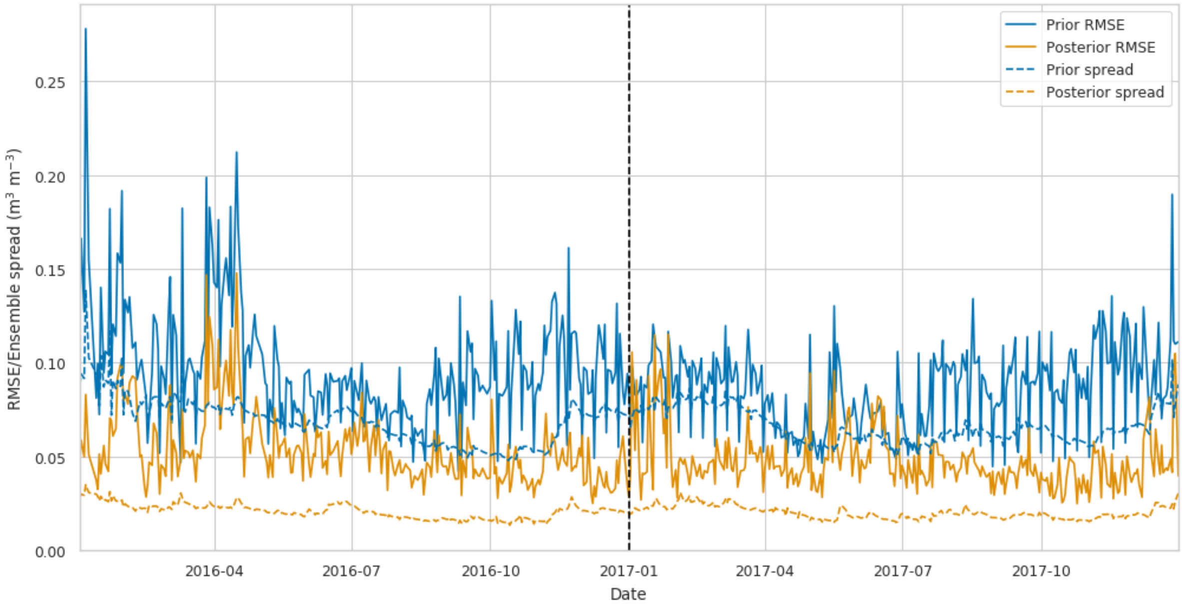


Figure 6: It is important to show how the prior and posterior spread compare with the actual RMSE calculated against the actual observation to check for consistencies with the DA setup. Without this analysis shown (for some points and maybe regionally), it is hard to diagnose the DA results. The goal is for the spread to have the same magnitude of the RMSE (not too large, nor too small)

Figure 4: It is not clear to me how RMSE is calculated in percentage. Maybe I missed something. Can the authors made this clear in the captions.

We agree this is a useful check to make. We will include plots of the prior and posterior spread in the model predictions of soil moisture averaged in space to compare to the model RMSE averaged in space (see below). We can see from this plot that for the prior we have the desired relationship with the ensemble spread being around the same magnitude as the prior RMSE. However, for the posterior we do find an ensemble spread with a slightly lower magnitude than the posterior RMSE. This is perhaps unsurprising as we are conducting just a single assimilation step but using all <30000 observations at once in space and time, so we may find that some of the posterior parameter distributions become too narrow, as with increasing observations we increase the confidence in our posterior, thus tightening the retrieved distributions. If we were to use our posterior optimized parameters in onward experiments we would require some form of ensemble inflation.

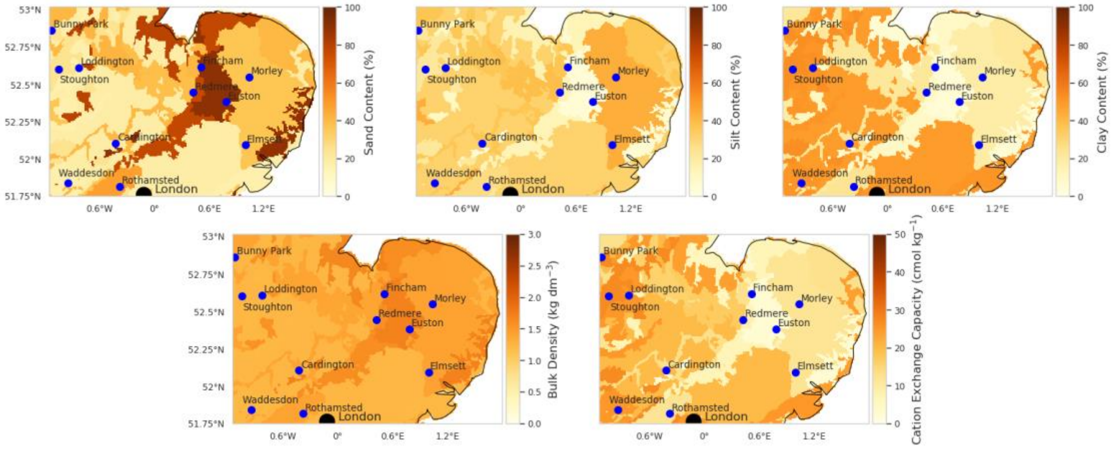
For the spatial plots of error reduction we have just calculated the percentage change between the prior JULES prediction RMSE (compared to SMAP) and the posterior JULES prediction RMSE, both averaged in time. We will define this in the manuscript.



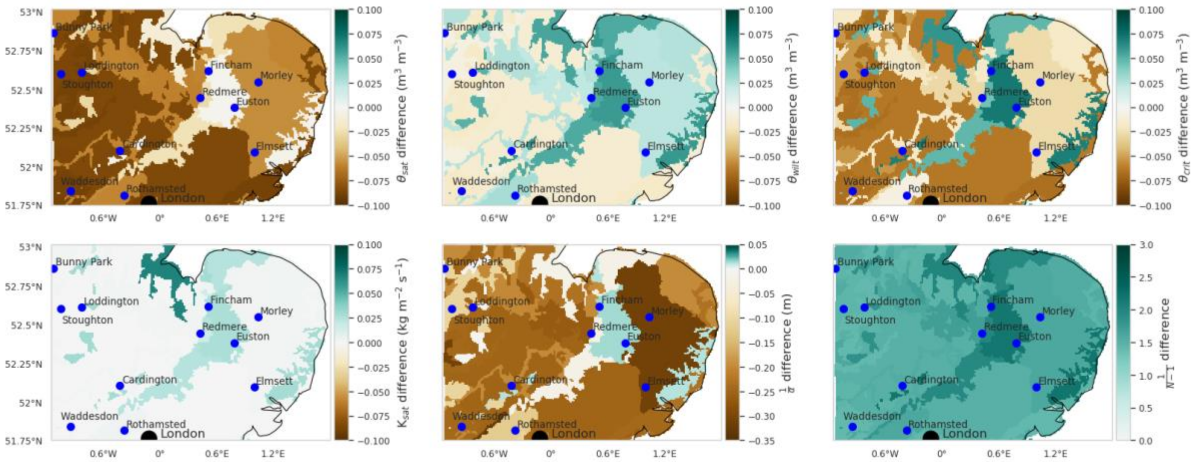
Results section: I found the results section to be presented in a very weak way. It seems to be rushed with the same regional map shown only for different metrics. The section is written almost like a technical report just going from figure to figure with very little in-depth analysis. How does the soil moisture in the region change from time to time (the metrics are only aggregated for the period)? Are the soil properties and consequently soil moisture profiles realistic? What are the impacts on other components of the model? Does 'improving' soil moisture improves other fluxes in JULES? My understanding is that COSMOS-UK also has flux data that can be used (H, LE, G???). The simple exercise of assimilating soil moisture to constrain parameters and/or states and evaluate the impact on soil moisture only does not seem to be particularly novel in my opinion (the DA frameowrk and the use of COSMOS-UK do, but should be explored further). This item is a major issue I have with the current manuscript.

We agree that the results section could be expanded to add more detail on the impact of the DA to other model components. We have had similar comments from other reviewers. We suggest adding plots of soil texture and how the resultant soil parameters change after DA. Also we agree looking at the performance of the models through time would be useful (see plot above). Unfortunately flux observations are not available from COSMOS-UK, we do have soil temperature which would allow us to judge the performance of another model component and we will include this. We will also include other water budget variables (ET, Run off) at each of the current 4 flux sites shown so that we can judge the impact of the DA on these variables (please see below for proposed plots). We believe that the ability to calibrate pedotransfer functions at a large scale using a considerable amount of satellite data (<30000 observations) in an instantaneous innovative data assimilation system does present novelty. Especially when it is shown that from this very large scale we are able to improve independent in-situ estimates from the model. However, we do still agree that including extra variables will strengthen the paper and we will include those stated above.

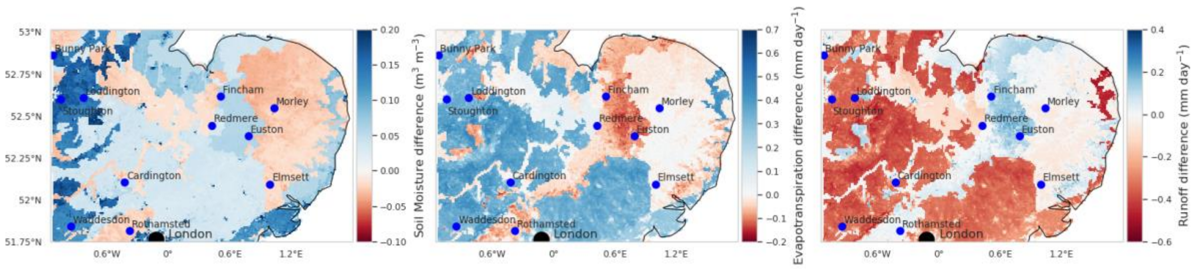
Soil properties:



Soil hydraulic parameter updates:



Impact on water budget variables:



COSMOS-UK site example (Cardington):

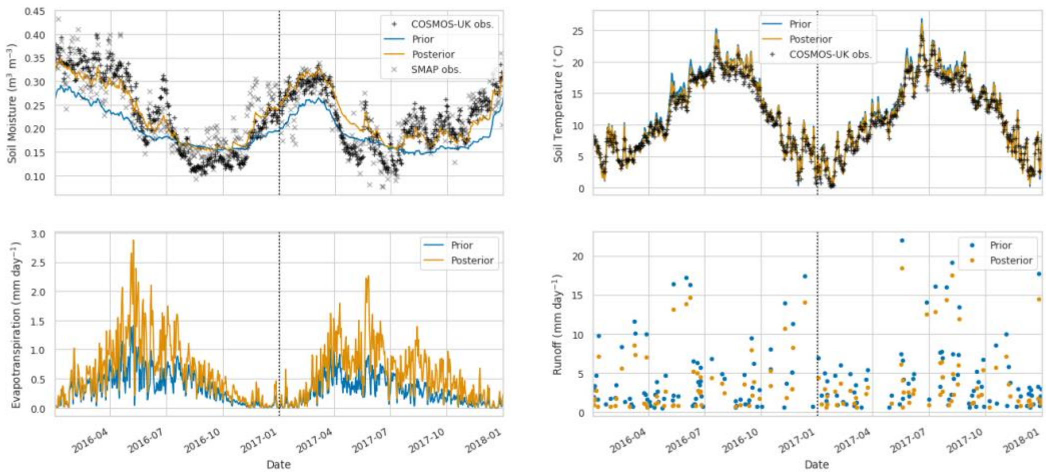


Figure 10: There seems to be some systematic biases in the model that suggests non-optimal DA setup (DA requires errors to be around a zero mean). How much that impacts the results? Are there other sites with similar issues (can you expand the discussion)? Have you tried some initial pre-calibration prior to runninng the DA to reduce/remove the biases?

The observations shown here are independent of the data assimilation. We agree it would be optimal to have errors centred around a zero mean for the assimilated SMAP observations. However, for independent in-situ validation data there will many competing errors that may make this impossible. There will be errors in the forcing meteorology (here we are using CHESS 1km forcing data and not observed in-situ meteorology), errors in the model grid and its representativity to the in-situ location, structural model errors (we currently have no ground water model in JULES and some in-situ sites may be more ground water dominated), errors in the vegetation fractions, and many more. Although at the larger scale for SMAP some pre-calibration could help we do not believe this is necessarily the case for the in-situ data. We will expand the discussion here to outline these issues.

Discussion section: I also found the discussion a bit weak. Very little is further discussed and explored. Sometimes the discussion is mainly focused on aspects that can be done in the future. I'd suggest the authors to define 2-5 clear objectives –> questions –> hypotheses that can be presented in more detail in the Results section, and discussed more in-depth in this current section.

We agree the discussion section could be improved and had similar comments from R#1. We suggested to include more literature in the discussion and will also make more comment on the include points above and new figures proposed for the results section.

Our suggested objectives are:

1) To examine the ability of 9km SMAP data to update PTF parameters in a 1km model.

2) To assess the resulting predictions of modelled soil moisture against (a) SMAP data from a different time period and (b) COSMOS probe data.