Authors' response letter (hess-2023-304)

The following is a point-to-point response to comments by reviewer #1

This work compared the simulation performance of Noah-MP land surface model over Ireland using two global soil property datasets, including a high-resolution SOILGRID data and a coarse-resolution STASGO data. The results showed that, the coarse STATSGO performs as good as the fine-scale SOILGRIDS soil database, although they both have dry biases. Overall, I think comparing the added value of high-resolution soil dataset to land surface modeling is an important and meaningful topic for both data and model developers. However, there are lots of uncertainties caused by other processes, such as the observation and the model physical parameterization, instead of the soil database. Following major issues should be carefully considered before the further consideration of the publication.

Dear Reviewer #1,

We thank you very much for your constructive comments and suggestions. The comments provided are very useful to improve the quality of our paper. All concerns and suggestions raised are addressed accordingly. Please find below your comments and our responses. Note that the authors' responses are highlighted in red.

Comment #1:

The innovation. The current innovation is somewhat weak for the HESS journal. The work is very similar to Zhang et al. (2023), although the authors said that they used the SOILGRID dataset. The work compared the difference of soil moisture/temperature simulation and drought processes (2 drought events) over Ireland, and made some conclusions. This makes the work like a technical comparison without in depth analysis of the uncertainties and related reasons. For example, why the high-resolution soil database performs similarly with the coarse-resolution? Is it because the uncertainties from soil database itself or the model structure and physical parameteritzations (for example, the uncertainties of PFT function to accurately derive the soil hydrological properties based on soil texture data)?

Response #1:

Zhang et al. (2023) did an excellent piece of work, incorporating global soil datasets into the WRF-Hydro model over southern Africa. However, our application differs in terms of climatic region, nature of the managed and highly heterogeneous soil landscapes, focus on specific weather events and soil physics. Ireland lies in the maritime temperature region with cool temperatures year round and no marked seasonality to precipitation. As a consequence, growing conditions are near optimal and grassland land cover accounts for almost 60% of the total land area. Grasslands here are also highly managed, for grazing and the provision of overwintering grass fodder for animals. Ireland also has far younger soils that are heterogeneous over small spatial scales; compared to South Africa, where soils are older and more consistent over large spatial areas.

In spite of its maritime climate, Ireland can and does experience periodic/seasonal soil moisture deficits, particularly in the sandy soils located in the south-east of the island. To the north and west, soils tend to have high clay content, which can act as a buffer to prolonged dry periods. Improved understanding of the potential impacts of climate change, specifically changes in the frequency and magnitude of drought/heat waves is of particular importance, not just for agricultural productivity, but on grasslands across Europe and more generally – grassland land cover represent an estimated 20-40% of global land cover (estimates vary depending on how grasslands are categorized).

Critically, the use of direct ground observations from sites with different soil characteristics allows us to more robustly evaluate the efficacy of the selected global soil databases in representing complex soil regimes. Compared to the work of Zhang et al., we evaluated different soil physics, including those based on PTFs, to provide insights into advancing soil hydrothermal extremes by evaluating the added benefit of vertical soil properties derived from 250 m SOILGRID maps. We also focus on the ability of the land surface model – NOAH-MP, which provides the only physical boundary to WRF climate model, to estimate soil hydro thermal properties under both mean and extreme conditions.

On basis of the background dynamic climate, grassland land cover/land cover use and soil conditions found here, and the evaluation of different soil physics schemes against a network of in-situ sensors (rather than remote), we believe the study, while strongly complementary to Zhang et al. (2023), is novel in application and relevant for a global audience.

The question about why both soil databases perform similarly was addressed in the paper (P15L21-34 and P14L18-30). We demonstrate that on one hand is a potential issue with misrepresentation of soil textural classes, but this does not fully explain why as sites with different soil texture types between STATSGO and SOILGRIDS (Table 1) (e.g. Valentia and Ballyhaise) produce similar results (Figures 5b, e in the paper). The reason for the similar performance between the soil databases is linked to the uncertainties of the empirical PTFs in the NOAH-MP model as discussed in P15L21-34. We have now carried out an uncertainty analysis as shown below (Figure 1) to further support our points.

The grid-scale uncertainty is quantified using the standard deviation difference between the experiments at the model topmost soil layer. Though model internal variability shows higher tendency with STATSGO than the SOILGRIDS, especially from Spring to Autumn, the standard deviation is generally below 0.08 m^3m^{-3} and the difference between the experiments is relatively small. The spatial patterns of standard deviation difference and the mean difference (see Figure A2 in the paper) are consistent with the topsoil textural classes (see Figure 2c in the paper) and field capacity (see Figure 3f in the paper). Our calculation of Pearson correlation coefficients between mean difference (standard deviation difference) of soil moisture and difference in the field capacity yields approximately 0.65 (0.45), and -0.40 (-0.1) with the soil hydraulic conductivity difference. This suggests that the mean and variability of soil moisture difference between STATSGO and SOILGRIDS are significantly ($p < 0.05$) related to the soil parameter values. This is further confirmed at a point location (see Figure 12 in the paper). However, for a more robust understanding on whether the remaining errors in the simulations are still linked to soil parameter uncertainty, it would involve performing ensemble simulations based on ensemble of parameter values which is beyond the scope of this work.

We will revise the paper accordingly to incorporate some of the information provided here.

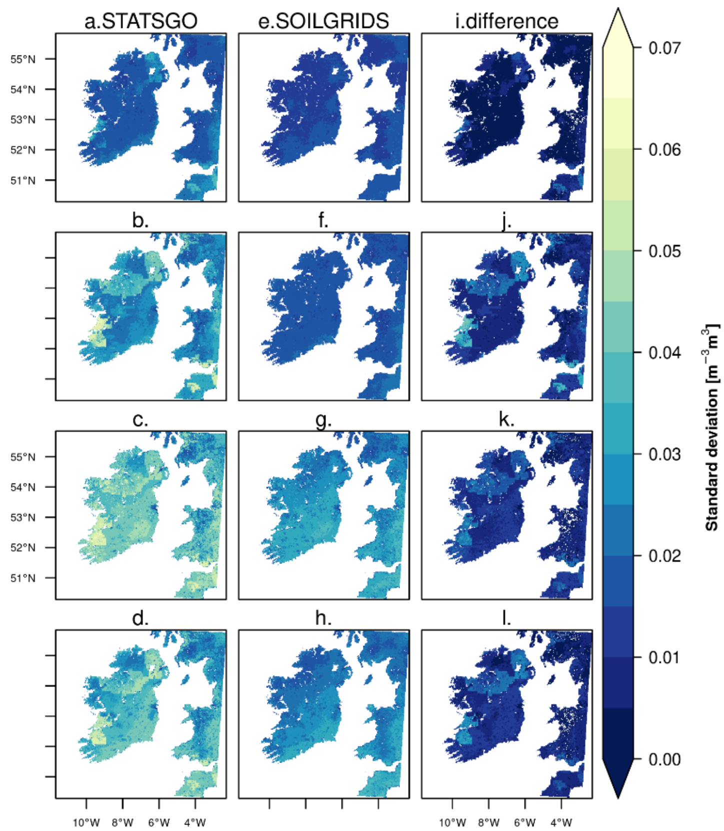

Figure 1. Spatial and seasonal comparisons of top-soil soil moisture internal variability between STATSGO [a-d] and SOILGRIDS [e-h] for the period 2009-2022. The rows represent Winter to Autumn.

Comment #2:

The station observation. There are 6 observation stations used in this work, and 4 of them are from a new network (Terrain-AI) using Time Domain Reflectometry (TDR) sensors. I found noteworthy difference between the Terrain-AI based observations and the 2 long-time eddy covariance grass flux sites. For example, most of the Terrain-AI based observations show high values during wet seasons (larger than 0.5 m3/m3), while observations from 2 eddy covariance grass flux sites are generally lower than 0.5. I wonder whether these high-values in Terrain-AI stations are true or not? This is very important to the conclusion, as the dry bias is mainly due to these stations. If the observation is true, then why there are such large differences considering they are all loam or loam-sand stations?

Response #2:

We agree that the Terrain-AI based soil moisture data should not be far from their counterpart flux sites, given that the soil textures, and instrumentation are similar.

At Terrain-AI, we have ensured standard, globally acceptable and well calibrated TDR sensors (Campbell Scientific CS615/CS616) across the stations. Therefore, to investigate these concerns, we analysed additional measured soil moisture records from 2023 to present across the Terrain-AI stations (Figure 1 below). In fact, the 2022 high values during the wet period are almost the same as 2023 and 2024. Hence, we should be safe to respond that the values are true and there is no evidence of sensor decay in these measurements.

However, we note that the soil moisture values at flux sites were measured in top 20 cm (Kiely et al., 2018; Murphy et al., 2022), whereas the Terrain-AI soil moisture used were measured at 5 cm depth (Figure 2 below, blue lines). Kiely et al. (2018) also mentioned that the porosity for top 5 cm at Dripsey site is 65 % which is broadly within the range of 5 cm values measured across Terrain-AI stations (Figure 2 below, blue lines). Unfortunately, we do not have near-surface soil moisture measurements from the flux sites, but we analysed the measured 20 cm soil moisture values from Terrain-AI stations (Figure 2 below, green lines), the values (0.4 - 0.45 m^3m^{-3}) at this depth during the wet period are broadly close to that of the flux sites. Therefore, we can state with some degree of confidence that the large differences in the wet soil moisture values between the Terrain-AI and the flux stations are due to different soil depths, as the near surface soil will be wetter than the deep soil layer between rainfall events. Again, this points to the more complicated soil landscapes experienced here, compared to Zhang et al. (2023).

For clarity, we have now separated the model evaluation analysis between near-surface (5 cm soil depth) and subsurface (top 20 cm) as shown in Figures 3 and 4 below.

We will revise the paper to provide more information on the station observations and associated soil depths. We will also incorporate the analysis in Figures 3 and 4 below to ensure clarity in relation to model evaluation and performance at different depths.

References

Kiely, G., Leahy, P., Lewis, C., Sottocornola, M., Laine, A., Koehler, A.-K.: GHG Fluxes from Terrestrial Ecosystems in Ireland. Research report No. 227.EPA Research Programme, Wexford. Available online at https://www.epa.ie/publications/research/climate-change/research-227.php, 2018

Murphy, R. M., Saunders, M., Richards, K. G., Krol, D. J., Gebremichael, A. W., Rambaud, J., Cowan, N., Lanigan, G. J.: Nitrous oxide emission factors from an intensively grazed temperate grassland: A comparison of cumulative emissions determined by eddy covariance and static chamber methods, Agric. Ecosys. Environ., 324 107725, https://doi.org/10.1016/j.agee.2021.107725 , 2022

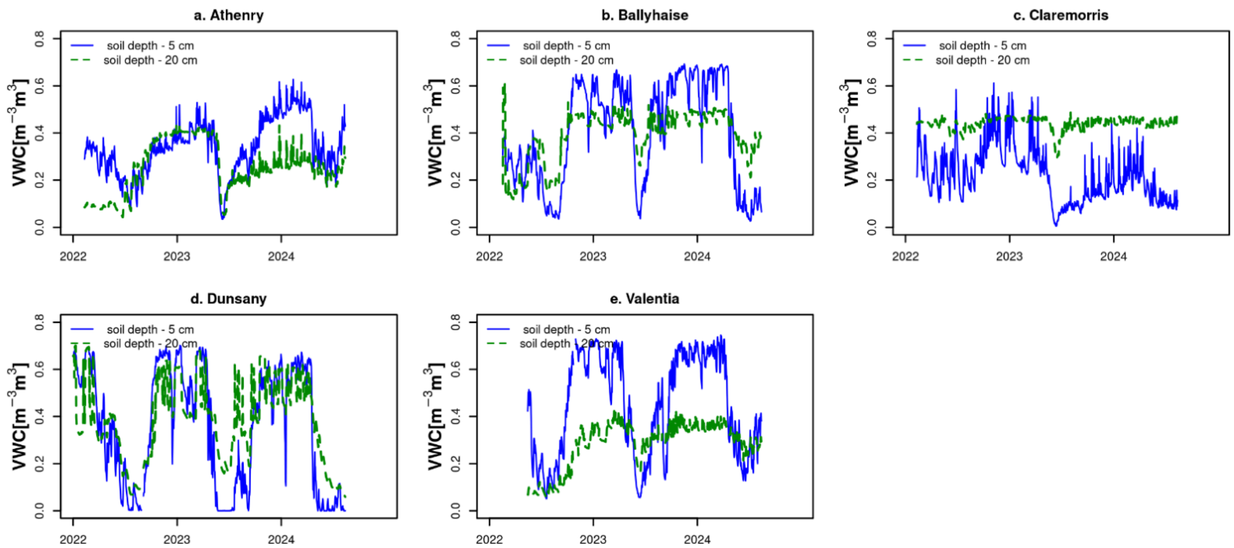

Figure 2 . Observed 5 cm and 20 cm depths TDR soil moisture from 2022 to present across the Terrain-AI stations

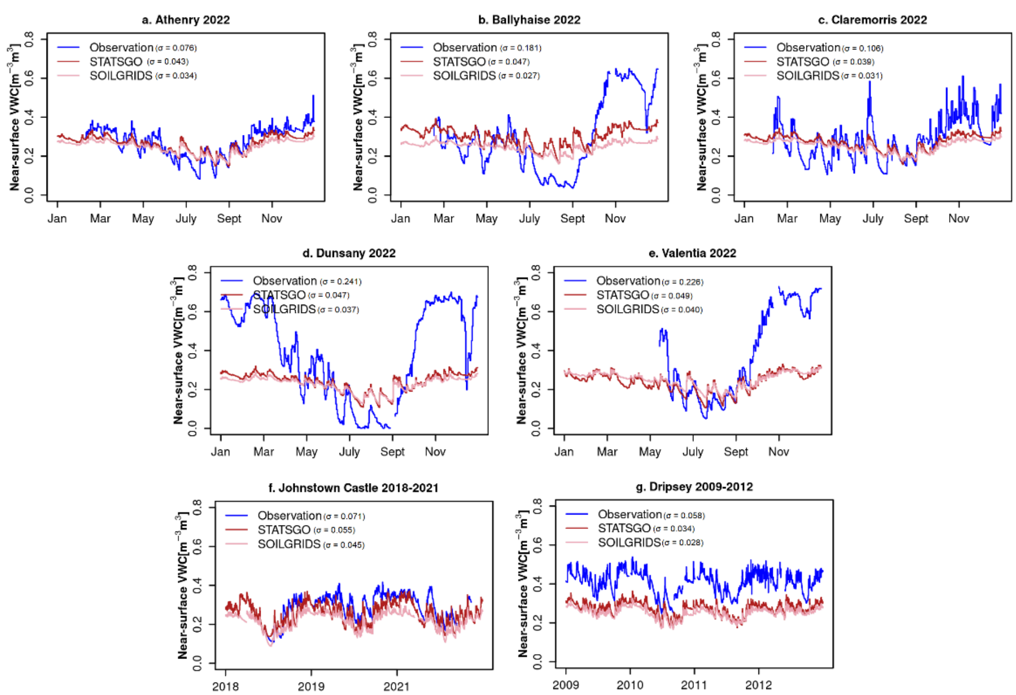

Figure 3. Temporal comparisons of near-surface volumetric water contents between observations at 5 cm depth and model simulations centered at 3.5 cm depth. The model simulations are contrasted with 20 cm depth due to unavailable near-surface observations for Johnstown Castle and Dripsey

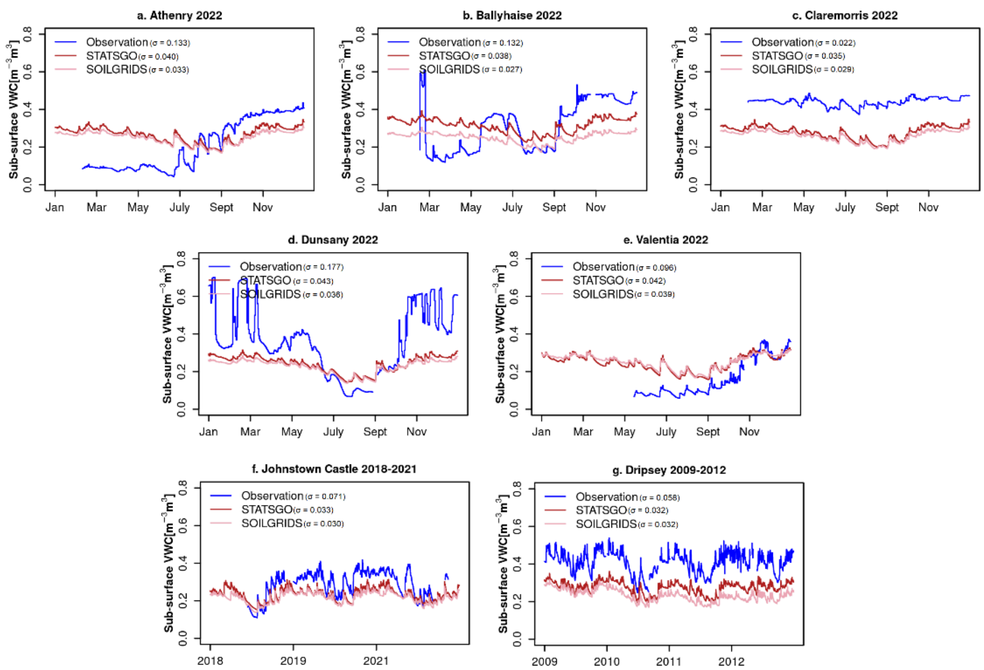

Figure 4. Temporal comparisons of sub-surface volumetric water contents between observations top 20 cm depth and model simulations centered at 17.5 cm depth.

Comment #3:

The satellite observation. Why you chose the ASCAT database as a reference? In my practice, the ESA-CCI or SMAP datasets are usually perform better than ASCAT. Is it because the ASCAT has best performance (use the station observation as a reference) or just ASCAT dataset can reproduce a dry bias pattern? In addition, how do you consider the influence of uncertainties of ASCAT dataset on the evaluation results?

Response #3:

While it may be true that ESA-CCI are a better performing product than ASCAT, this also depends on the types of products employed, for example, ASCAT has achieved the best performance among non-blended products, including ESA-CCI combined products (e.g. Mazzariello et al., 2023).

We have independently evaluated ESACCI SSM (25 km resolution), GSSM (1 km machine learning based surface soil moisture products) (Han et al., 2023) and ASCAT 1 km SWI (linearly transformed using variance matching with station observations), and carefully chosen the latter as the reference. Evidence from our evaluation of these products against the station observations (Figures 5 and 6 below) suggests that ASCAT yields better performance than ESA CCI SSM and GSSM 1 km products, though the latter products show higher temporal dynamics as shown by the higher temporal correlations with the ground observations. The rising and falling trends are also better captured by ASCAT. While the uncertainty in GSSM products is likely linked to lack of training data from Ireland, the biases in ESA CCI SSM may be attributed to its native grid resolution which is too coarse to effectively represent the soil heterogeneity, and/or differences in soil depths.

In addition, the model standard errors of 0-0.07 m^3 m^{-3} (Figure 1 above), accounting for model uncertainty, are below the ASCAT uncertainty threshold of 0.1 m^3 m^{-3}. Therefore, model errors may not be overestimated because of large ASCAT uncertainty. However, because ASCAT products do not account for soil textural properties, the use of a characteristic time length (e.g. T2) without soil texture differentiation may influence our results, as the optimal characteristic time lengths differ for different soil texture categories (de Lange et al., 2008).

Reviewer #2 also agrees that other products can be evaluated to support the analysis. We will revise the paper accordingly to reflect this information.

References

Han, Q., Zeng, Y., Zhang, L. et al. Global long term daily 1 km surface soil moisture dataset with physics informed machine learning. *Sci Data* 10, 101 (2023). https://doi.org/10.1038/s41597-023-02011-7

Mazzariello, A, Albano, R., Lacava, T., et al. Intercomparison of recent microwave satellite soil moisture products on European ecoregions, *J. Hydrology*, 626, 130311 (2023). https://doi.org/10.1016/j.jhydrol.2023.130311

R. de Lange, R. Beck, N. van de Giesen, J. Friesen, A. de Wit and W. Wagner, "Scatterometer-Derived Soil Moisture Calibrated for Soil Texture with a One-Dimensional Water-Flow Model," in *IEEE Transactions on Geoscience and Remote Sensing*, 46, 12, 4041-4049, 2008. doi: 10.1109/TGRS.2008.2000796

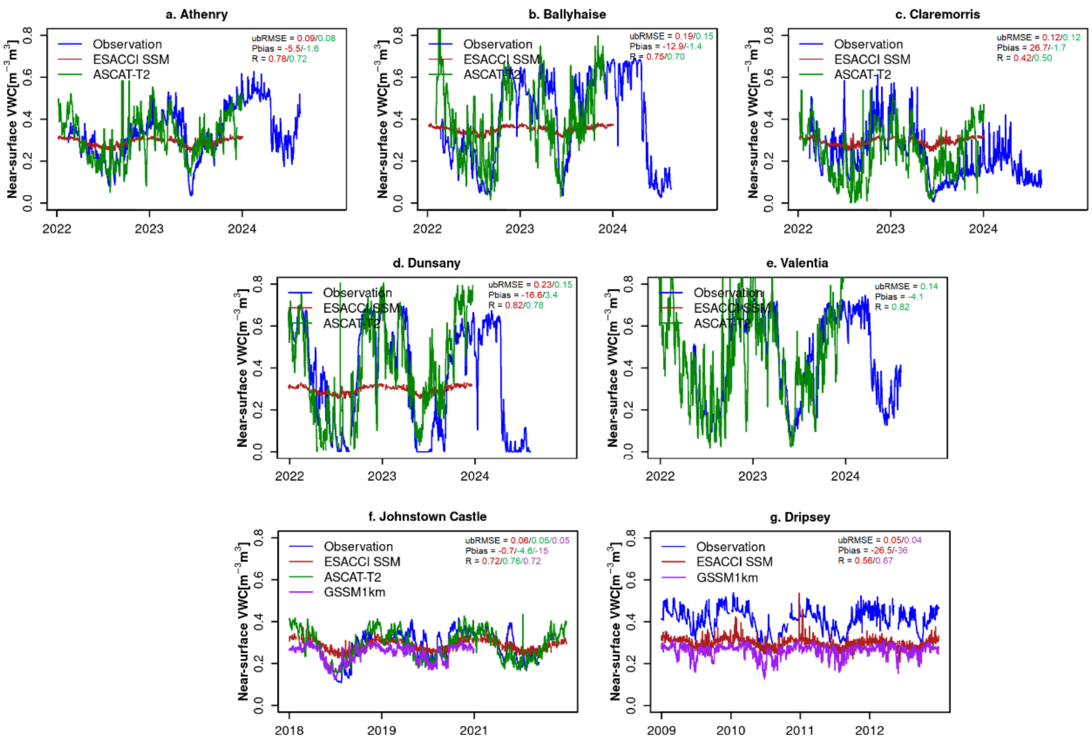

Figure 5. Evaluation of satellite-derived 1 km ASCAT-T2 (0-10 cm), 1 km GSSM (0-5 cm) and 25 km ESACCI near-surface soil moisture against the station observations. No available ESACCI SSM grid values for Valentia, and due to ASCAT later year of operation in 2015, no available ASCAT values also for Dripsey.

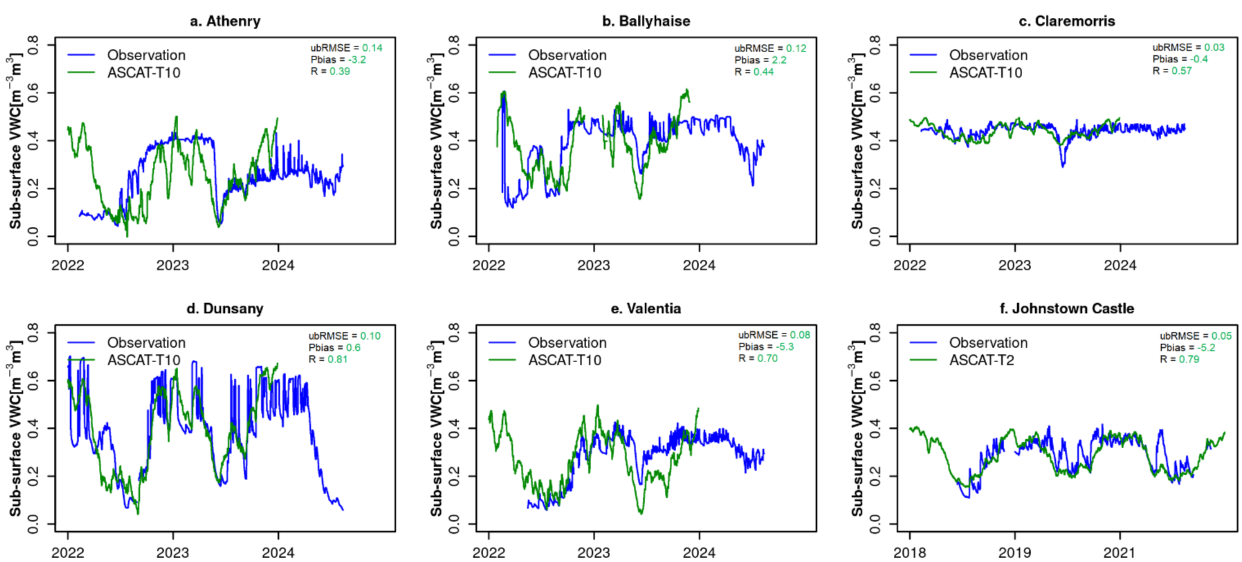

Figure 6. Evaluation of satellite-derived 1 km ASCAT-T10 (10-30 cm) sub-surface soil moisture against the station observations (20 cm). No sub-surface values for ESACCI and GSSM products.

Comment #4:

The soil moisture or soil moisture anomaly. Although SOILGRID seems to show larger negative bias in modeling soil moisture, it improves the correlation coefficient. An important issue is whether the soil moisture absolute value is more important than the soil moisture anomaly (dynamics)? Actually, the observed soil moisture and model simulated soil moisture are physically different. Model simulation is a mean state of a grid box with specific thickness, while observations only represent a point at a fixed depth.

Response #4:

The soil moisture anomaly may be more important if the interest is in the relative state of soil moisture, rather than the absolute soil moisture values and how they may vary seasonally.

There is clear seasonality to soil moisture in Ireland, thus we have looked at evaluating both absolute soil moisture values at point scale and the relative state at grid scale. We demonstrated that the improved correlation coefficient in SOILGRID is consistent regardless, though the improvement is very small relative to STATSGO. The reason for SOILGRID showing higher negative biases and correlation coefficient is due to uncertainty in PTFs-derived soil parameter values and stronger seasonal effect or increased spatial scale, respectively. We have addressed this issue in our response #5 below.

We acknowledge that due to soil heterogeneity across the landscape, the scale discrepancies between point and model grid data or soil depths may have introduced uncertainties in our results. These issues are out of the scope of the current work. However, we have evaluated the model against the comparable and better performing 1 km ASCAT soil moisture products, to reduce the issues of scale mismatch. But again, these satellite products also have inherent uncertainties associated with them as shown above in our response #3. Hence, we have provided a note of caution in P8L18-20 and P9L34-36. Additionally, we have carried out further analysis and demonstrated that 1 km ASCAT SWI is better performing with ground observations (see Figures 5 and 6 above)

We will revise the paper accordingly to properly acknowledge the potential uncertainties due to different limitations. These will include spatial scale mismatches between point-based observations and model 1 km grids; differences in soil depths between observations and model; uncertainties in soil moisture measurements/satellite products.

Comment #5:

Why the SOILGRID improves the simulation of soil moisture dynamics but increases the dry biases? Some in-depth analysis should be provided. In addition, the differences in soil moisture drought may not simply related to the simulation of soil moisture absolute values because the soil moisture percentiles are used here. I wonder whether the soil hydraulic conductivity or diffusivity is responsible for the difference here. For example, a higher conductivity can cause a faster response of soil moisture to the water deficit.

Response #5:

Thanks for this comment. We have further investigated this, as a common practice we calculated the soil moisture anomalies for ground observations and model outputs to remove the potential seasonal effect on model evaluation. The calculation was based on a z-score anomaly over a 35-day moving window where the number of soil moisture samples is greater than 6. While Pearson's correlations for anomalies reduce significantly (ranging between 0.45 and 0.78) compared to the absolute values for both model experiments across the stations, the soil moisture dynamics are still sometimes higher in SOILGRID than the STATSGO. Hence, the improved SOILGRID soil moisture dynamics may be linked to the increased resolution (250 m) soil input data which allowed us to better constrain the model soil heterogeneity in terms of the texture and vertical profiles.

In addition, we agree with the reviewer's observations. The SOILGRID dry biases are evident particularly in areas with higher soil hydraulic conductivity, as demonstrated in Figure 7 below. These areas are mostly associated with increasing grain size (from Clay to Loam or Loam to Sandy Loam), thereby increasing the pore spaces, faster decline in soil moisture memory and rapid drying, relative to the STATSGO. This is in addition to the significant influence of lower soil field capacity as earlier explained in our response #1.

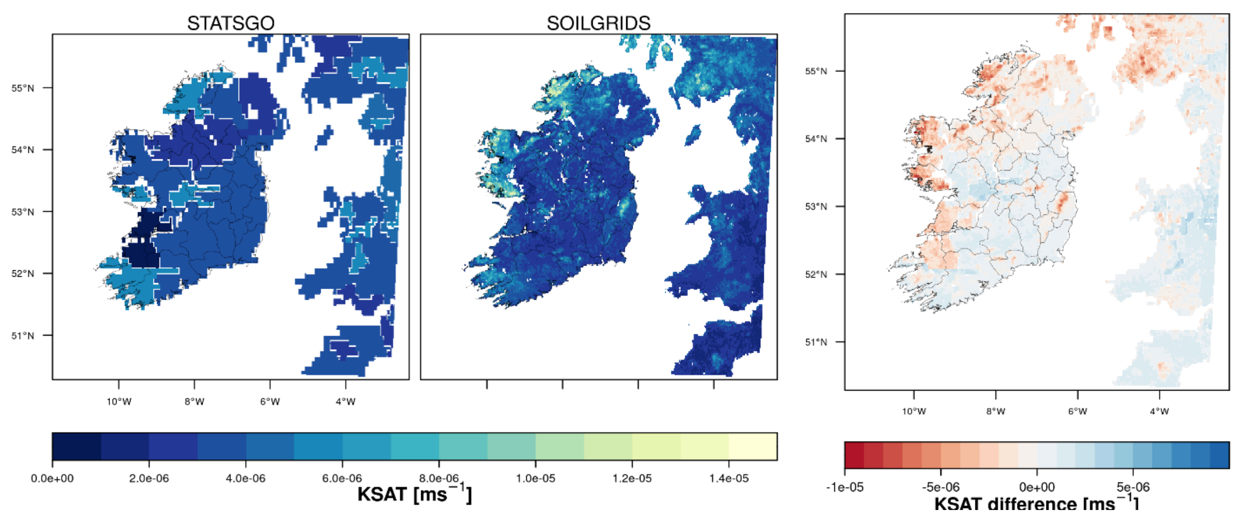

Figure 7. Spatial characteristics of absolute and difference between STATSGO and SOILGRIDS for soil hydraulic conductivity (KSAT)