Scientific/Detailed Comments: Referee comments in bold and author answer non-bold.

<u>Review comments by Referee #1</u>

In this manuscript the authors described assimilation of soil moisture observations into the land surface model CLM5. In general, the paper is well structured, described and written. I have several questions which are still unclear for me.
1) as you mentioned in the end of section 2.1, you leave the data gaps. I'm wondering your observation frequency.

The observation frequency varies among data sources, e.g. FLUXNET provides data in 30-minute intervals. However, as mentioned earlier in section 2.1, we use daily averages of the soil water content, evapotranspiration, and sensible heat for the assimilation and analysis.

Changes in text:

Lines: 134 - 139

"We only assimilated (daily mean averaged) soil water content observations when measurements were available for a given day. The daily mean averages were calculated independent from the observation frequency for the different sites. Similarly, simulated evapotranspiration was only compared with observations when data were available, on the basis of daily mean averages."

2) for parameter updating in 2.3.2, you update the solid fractions instead of hydraulic conductivity. Please explain the reason. Does it work better, or what's the particular reason for this?

We performed data assimilation using a direct approach to update combinations of soil saturated hydraulic conductivity, hydraulic conductivity exponent, porosity, and soil matric potential but found that indirectly updating sand, clay, and organic matter fractions using pedotransfer functions provides better results. Specifically, we found that parameters often approached upper or lower defined limits, which did not happen with the indirect approach. Updating the solid fractions provided in general more stable parameter estimates and better simulation results.

Changes in text:

Lines: 220 - 230

"In previous studies parameters were updated indirectly (Naz et al., 2019; Han et al., 2014; Baatz et al., 2017). We tested directly updating saturated hydraulic conductivity, porosity, hydraulic conductivity exponent, and soil matric potential but this resulted in more unstable estimates than indirectly updating soil hydraulic parameters. The pedotransfer function which is used for the indirect updating results in reasonably correlated soil hydraulic parameters. In this study, the parameters are chosen to optimize the SWC estimation and not ET estimation to study the effects of SWC improvements on ET. To more directly improve the ET estimation, parameters affecting the ET process directly should be added, e.g. vegetation hydraulic parameters. "

3) line 175: how about the observation errors vs. perturbations?

(Line 175: "Only the PFT were manually assigned for each site. For the ensemble creation, the fractions ofsand, clay, and organic matter are modified for each ensemble member. The perturbations are normally distributed with mean zero and a standard deviation of 10%.")

For the data assimilation the observation error is assumed be constant and set to a root-mean square error of 0.02 cm^3/cm^3.

Changes in text:
Lines: 201 - 203

"In this study, the state vector depends on the simulation scenario (explained in more detail in section 2.3.2) and R is based on the measurement errors which are assumed to be constant and independent with a root-mean square error of 0.02 cm^3/cm^3."

4) line 190: parameter updating refers to fractions of sand, clay and organic? Are there any other parameters included, e.g. the vG parameters and porsity? If not, why do you only update these parameters?

Yes, parameter updating refers to the sand, clay, and organic matter fractions. We use the indirect approach in which the soil hydraulic parameters are calculated from these characteristics using the Clapp and Hornberger (1978) pedotransfer function as mentioned in 2.3.2. CLM5 uses the Brooks-Corey parameters and not the vG parameters as mentioned in section 2.2. As already mentioned in comment 2), we found that the indirect approach via pedotransfer functions provides better results than direct updating of hydraulic parameters.

Changes in text:
Lines: 220 - 230

"In previous studies parameters were updated indirectly (Naz et al., 2019; Han et al., 2014; Baatz et al., 2017). We tested directly updating saturated hydraulic conductivity, porosity, hydraulic conductivity exponent, and soil matric potential but this resulted in more unstable estimates than indirectly updating soil hydraulic parameters. The pedotransfer function which is used for the indirect updating results in reasonably correlated soil hydraulic parameters. In this study, the parameters are chosen to optimize the SWC estimation and not ET estimation to study the effects of SWC improvements on ET. To more directly improve the ET estimation, parameters affecting the ET process directly should be added, e.g. vegetation hydraulic parameters.

<u>Review comments by Referee #2</u>

This reviewer had two critical concerns about this manuscript related to the simulation of leaf area index, and the design and implementation of the parameter estimation:

First, I don't see how the author's came to the conclusion "[the] results suggest that state-of-the-art LSM such as CLM5 still suffer from uncertainties in the representation of soil hydrological processes in forests, e.g. deep root water uptake, uncertainties in the representation of biological processes of tree transpiration", without accoutning for biases of the simulated leaf area for the sites. Running CLM5-BGC can lead to erroneous leaf area values, and is why CLM users often use CLM-SP (prescribed leaf area) to diagnose to what extent the simulation of leaf area is impacting your soil moisture and ET relationship. Have a look at Li et al., (2022) or Fox et al., (2022) as an example of how prescribing or assimilating observations of leaf area can effect the representation of carbon and water cycling. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002747

If the simulated leaf area is incorrect in magnitude and timing it is unlikely that improving the soil water content through DA will improve ET. The authors do not account for this in the manuscript.

We agree that running CLM5-BGC can lead to erroneous LAI simulation results and that CLM-SP can be used to diagnose to what extent this uncertainty affects the simulation of soil moisture and ET. Therefore, we now also explore the difference between CLM-SP and CLM-BGC simulations and have included the corresponding results in the revised manuscript.

Changes in text:

Lines: 24 – 29, 303 - 313

"This finding indicates that only improving the SWC estimation of state of the art LSM such as CLM5 is not sufficient to improve evapotranspiration estimates for forest sites. To improve evapotranspiration estimates, it is also necessary to consider the representation of LAI in magnitude and timing, as well as uncertainties in water uptake by roots and vegetation parameters."

"This could be caused by the mismatch of simulated and actual LAI for these sites. To investigate this, we repeated the simulations using CLM5 with satellite-derived phenology (CLM5-SP). The results are shown in Fig. 9. For CLM5-SP we observe an average improvement in the RMSE of SWC between 57.6 % and 64.3 % and an average RMSE deterioration of 5.8 % for the ET estimation. These CLM5-SP simulations use the default datasets from CLM5 and without site specific calibration of the timing or magnitude of the seasonal phenology of LAI. Therefore, even for the CLM5-SP simulations, there is a mismatch between simulated and actual LAI. However, also for this case, there are sites with a large improvement in SWC estimation that show deterioration in the ET estimation.

Another possible explanation for the improvement in SWC estimation but no improvement of ET estimation is the underestimation of root water uptake from deeper soil layers for forest sites, as also suggested by Shrestha et al. (2018). "

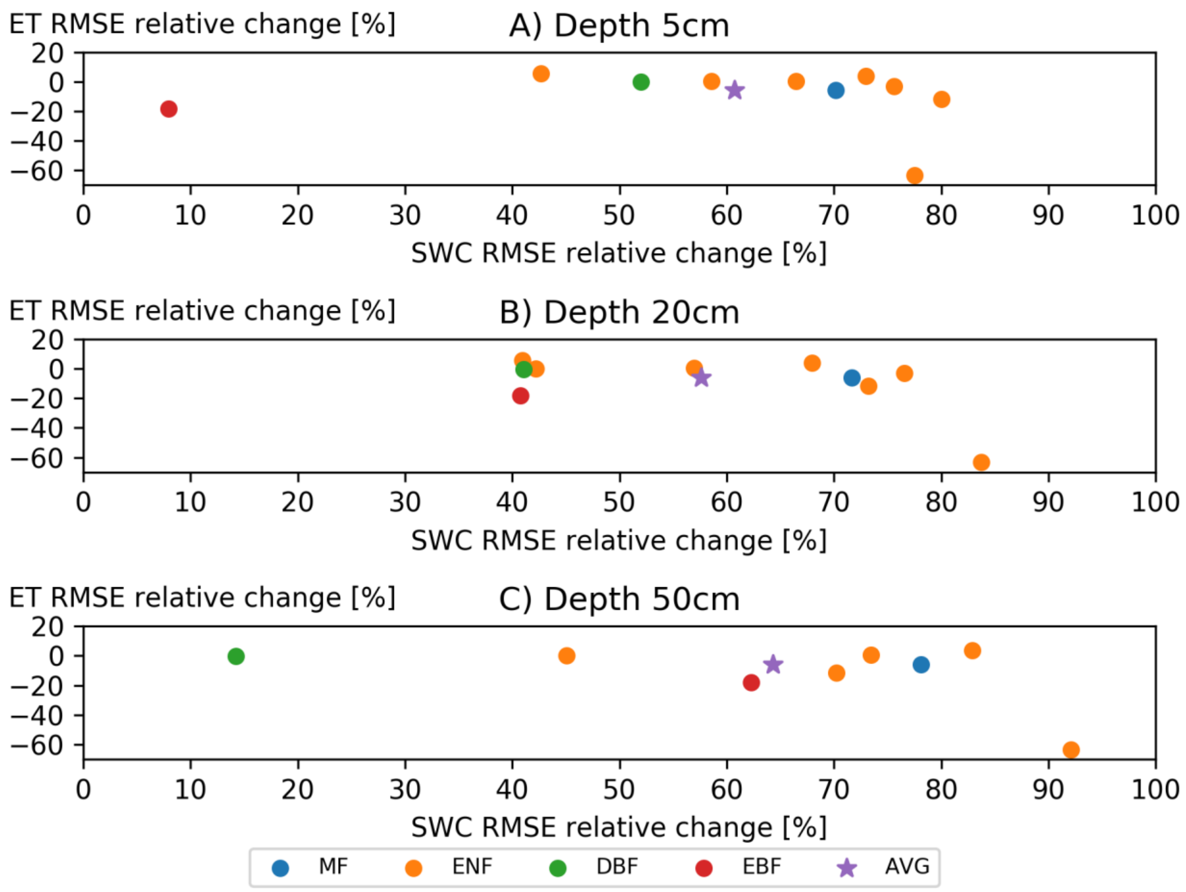

Figure 9: Comparing the SWC and ET characterization for the OL and DASP simulations using CLM5-SP. Each point represents the overall average RMSE change for one site. The color of the points indicate the forest type (MF: mixed forest, ENF: evergreen needle leaf forest, DBF: deciduous broad leaf forest, EBF: evergreen broad leaf forest, AVG: average over all forest types).

Second, in reference to the parameter estimation, the authors update the soil layer fractions sand/clay/organic which then impact to what extent the true soil hydraulic properties are weighted in the soil layer. It is a strange approach to access the 'true' global hydraulic parameters through a fixed land surface property. This would be equivalent to treating the PFT for each site as a 'parameter' rather than a fixed/prescribed property, and varying the PFT fraction across MF/ENF/DBF/ENF/WSA for each site, which would then weight the parameter set associated with each PFT. This is not done, and the authors don't do that here. Instead the author's prescribe the PFT for each site and keep that fixed. Similarly, it is unclear why the author's would then not prescribe the sand/clay/organic layers either from the default CLM files, or prescribe the sand/clay/organic fractions from the FLUXNET site data.

Soil texture data are always subject to uncertainties and therefore it is reasonable to not fix them. Therefore, the indirect approach of updating soil characteristics, i.e. sand, clay, and organic matter fraction, to update the soil hydraulic parameters using a pedotransfer function was used in various studies (Naz et al., 2019; Han et al., 2014; Baatz et al. 2017).

Secondly, we performed data assimilation using a direct approach to update combinations of soil saturated hydraulic conductivity, hydraulic conductivity exponent, porosity, and soil matric potential but found that indirectly updating sand, clay, and organic matter fractions using pedotransfer functions provides better results. Specifically, we found that parameters often approached upper or lower defined limits, which did not happen with the indirect approach. Updating the solid fractions provided in general more stable parameter estimates and better simulation results.

Changes in text:

Lines: 220 – 230

“In previous studies parameters were updated indirectly (Naz et al., 2019; Han et al., 2014; Baatz et al., 2017). We tested directly updating saturated hydraulic conductivity, porosity, hydraulic conductivity exponent, and soil matric potential but this resulted in more unstable estimates than indirectly updating soil hydraulic parameters. The pedotransfer function which is used for the indirect updating results in reasonably correlated soil hydraulic parameters. In this study, the parameters are chosen to optimize the SWC estimation and not ET estimation to study the effects of SWC improvements on ET. To more directly improve the ET estimation, parameters affecting the ET process directly should be added, e.g. vegetation hydraulic parameters.

Furthermore, the choice of parameters to be estimated should more closely relate to processes controlling the SWC-ET relationship, such as stomatal conductance, and the *vegetation* hydraulic parameters specifically involved in the Plant Hydraulic Stress formulation in CLM5. The soil hydraulic parameters estimated in the manuscript, seem to control the dynamics of moisture within the soil layer, but not the transfer of water from root to leaf, which is controlled by the stomatal conductance and PHS (Kennedy et al., 2019). Since the author’s are already adjusting for SWC, adjusting parameters inherent to the soil hydraulics seems somewhat redundant.

We agree that we could improve the simulation of ET by updating vegetation hydraulic parameters and will do so in further studies. However, for this particular study we chose to use parameter updating to improve the SWC estimation as best as possible and analyze the effect of this improvement to the related evapotranspiration estimation. This decision is related to a central objective of this study: To investigate whether precise in situ soil moisture measurements can improve ET characterization, as several studies have found that this cannot be achieved with more uncertain remote sensing soil moisture data. Please notice that assimilating soil moisture data with the aim to improve the simulation of land surface processes is a common modelling strategy.

Changes in text:

Lines: 227 - 230

“In this study, the parameters are chosen to optimize the SWC estimation and not ET estimation to study the effects of SWC improvements on ET. To more directly improve the ET estimation, parameters affecting the ET process directly should be added, e.g. vegetation hydraulic parameters.”

Finally, the authors never provide figures/tables of the influence of the DA on the parameter values. Do the parameter values converge to a new value, or are they temporally changing and random? Without this information it is impossible to assess why the parameter estimation had little influence on ET.

We agree that we did not provide enough information about the parameter updates, since we did not see it as a focus of this study. The parameters updated with DA do change when observations are assimilated and converge to new values. We included a new figure to show this.

Changes in text:

Lines: 286 - 291

“Figure 5 shows time series of the estimated saturated soil hydraulic conductivity for each of the sites and the three observation layer depths. The DASP scenario results in parameter changes when the first observations are available but converge over the time of the simulation to a new value. The corresponding time series for the other soil hydraulic parameters can be found in the appendix (Figures A8, A9, and A10). The sand, clay, and organic matter fraction and thus the soil hydraulic parameters can vary with depth but as shown in Fig. 5 the DA updates to the parameters affect the different layers similarly.”

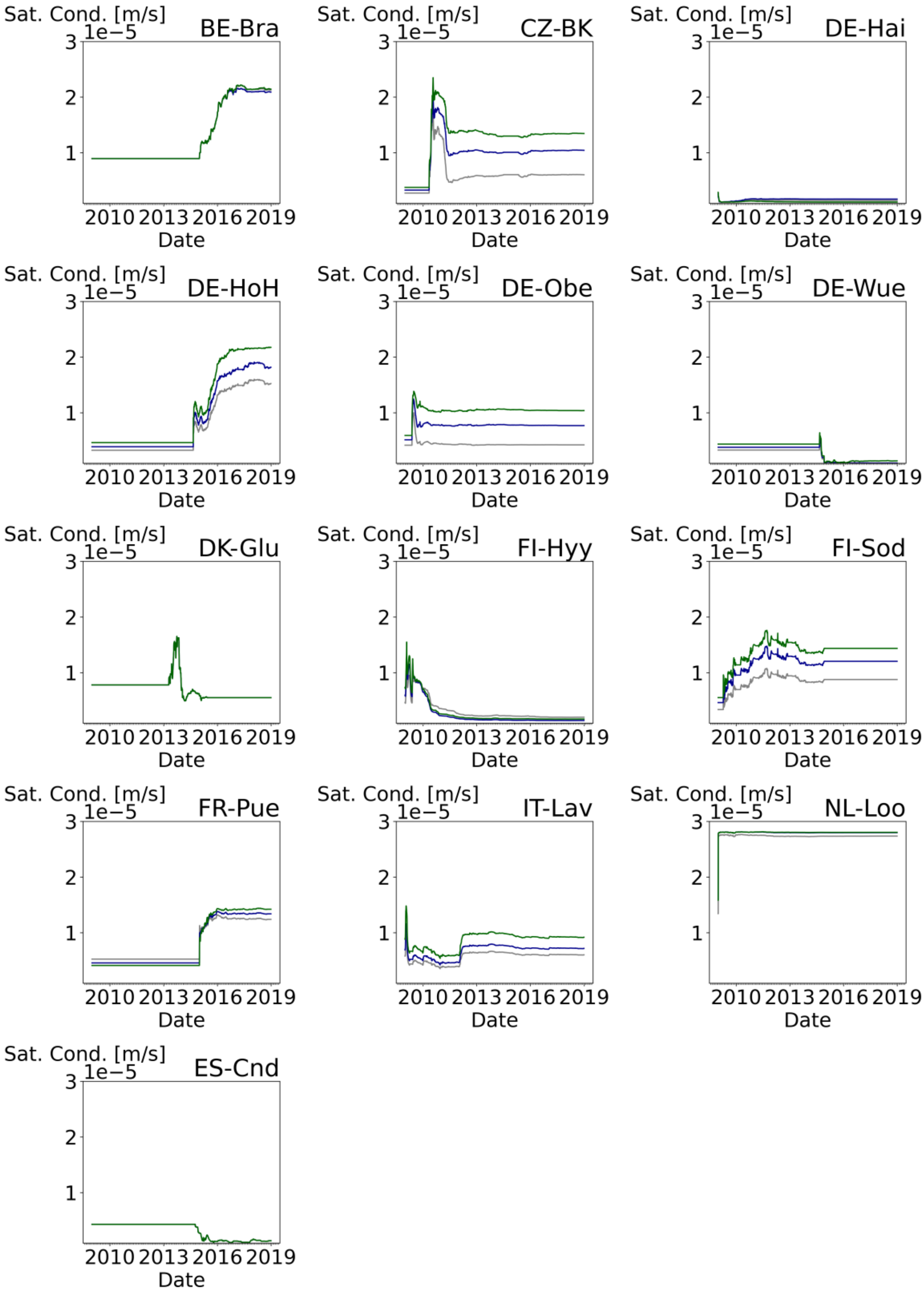

Figure 5: Time series of the saturated soil hydraulic conductivity for each site in the DASP simulation. The grey line is the value at 5 cm depth, the blue line at 20 cm depth, and the green line at 50 cm depth.

Detailed comments:

Line 55-60: The literature review seems a bit incomplete. Site level studies where parameters were hand tuned for the site of interest for forested sites also seems relevant: Duarte at al, 2017; https://doi.org/10.5194/bg-14-4315-2017 and Raczka et al., 2016; https://doi.org/10.5194/bg-13-5183-2016.

We focused in our literature review more on studies using data assimilation and agree that we can extend the literature review to include references to studies using manual tuning of parameters.

Changes in text:

Lines: 80 - 85

"Other studies have used manual tuning of parameters to improve CLM simulations for forests. For instance, Duarte et al. (2017) calibrated CLM4.5 for an old-growth coniferous forest and found good agreement between simulated and observed response of canopy conductance to atmospheric vapor pressure deficit and soil water content. Raczka et al. (2016) used CLM4.5 and implemented a seasonally varying calibration of vegetation parameters and accurately simulated net carbon exchange, latent heat exchange, and biomass."

Line 61: crop cover, not cover crop

The referenced study is about the inclusion of the management practice of cover crops, i.e. crops specifically planted to cover the soil during winter to reduce soil erosion, soil compaction, and nitrogen leaching and to increase agricultural productivity by nitrogen fixation. Therefore, we think the term "cover crop" is correct here.

Line 69: "The point scale measurements use invasive equipment and the specific measurement volume, exact depth of the sensors, number of sensors, and number of stations varies from site to site. For a few sites we use soil water content measurements from Cosmic Ray Neutron Sensing (CRNS) from the COSMOS-Europe data set (Bogena et al., 2022)."

Seems worthwhile to diagnose the in-situ observations with the CRNS data for the same sites. This could help quantify representation of point level measurements for a flux tower.

We agree that a study comparing the in-situ observation with the CRNS observations for sites where both data exist would be useful. However, in this study we focus on the data assimilation and the effect on modeled evapotranspiration and not the direct comparison between observational methods. This is already investigated and documented in other papers and goes beyond the scope of this study.

Changes in text:

Lines: 102 - 106

“CRNS use neutrons as proxy for SWC and the vertical measurement depth varies with the soil moisture conditions. Additionally, the uncertainty of CRNS-derived soil moisture varies not only with the different neutron detectors but also with the number of counts in a time period and therefore results under lower soil moisture conditions are more accurate (Bogena et al., 2022).”

Line 73: “The CRNS provides continuous and non-invasive soil water content measurements over a spatial footprint of hundreds-of-meters and integrates from the surface to a depth of 10-70 cm vertically in the soil (Zreda et al., 2008; Köhli et al., 2015).”

If the COSMIC measurements are non-invasive, how are they integrating the soil water content from 10-70 cm vertical depth? Clearly these are not just measurements, but modeled output that is likely constrained from the surface measurements (1-5 cm). What are the uncertainty values for the CRNS values? More details are needed.

The CRNS measures neutrons as proxy for SWC and use conversion functions and weighting to provide soil water content for the upper 10 cm of soil (under very wet conditions) to upper 70cm of soil (very dry conditions). The uncertainty of CRNS-derived soil moisture varies not only with the different neutron detectors but also with the number of counts in a time period and thus results under lower soil moisture conditions are more accurate.

Changes in text:

Lines: 102 - 106

“CRNS use neutrons as proxy for SWC and the vertical measurement depth varies with the soil moisture conditions. Additionally, the uncertainty of CRNS-derived soil moisture varies not only with the different neutron detectors but also with the number of counts in a time period and therefore results under lower soil moisture conditions are more accurate (Bogena et al., 2022).”

Lines 65-83: Seems odd the authors would not mention their previous manuscript Strebel at al., (2022) in this review section (Wustebach catchment), which essentially uses the same exact DA setup, but expands the analysis to more sites in this analysis. Seems the authors should state their previous findings here, and use that to motivate and distinguish this analysis from the previous one.

The previous study mentioned here focused more on the implementation details of the software coupling and therefore we reference it in the method section.

Changes in text:

Lines: 89 - 93

“In a previous study (Strebel et al. 2022), we investigated the potential for data assimilation of in-situ SWC measurements to improve model estimation for a single forest site. This study expands this method to more forest sites and investigates the effect of improved SWC estimation on ET.”

Line 97: 'daily average soil water content data are assimilated'. Need more detail of what depth these observations are taken from. This needs to be explained either here or more clearly in the methods.

We describe the vertical layout in section 2.4.1.

Changes in text:

Lines: 187 - 191

"For FLUXNET sites, measured soil water content is provided for up to three depths described as superficial, medium, and deep. Since data assimilation in CLM5-PDAF requires a specific vertical layer, we assigned 5, 20 and 50 cm to the respective FLUXNET SWC layers. For the CRNS sites, the measurement depth for each individual measurement is calculated following Schrön et al. (2017) and is included in the dataset from Bogena et al. (2022). "

Table 1: I think elevation would also be relevant to report here.

Added elevation to table 1.

Changes in text:

Table 1: Overview of the study sites. Classification uses the International Geosphere-Biosphere Program Code (IGBP) as is used for FLUXNET: MF for mixed forests, ENF for evergreen needle leaf forests, DBF for deciduous broad leaf forests, EBF for evergreen broad leaf forests, WSA for woody savannah. LON is longitude and LAT latitude.

Site name	Country	Abbreviation	Code	LON	LAT	Elevation [m.a.s.l.]	Data source	Mean annual temperature [°C]	Mean annual precipitation [mm]	Typical tree species
Brasschaat	Belgium	BE-Bra	MF	4.51	51.3	16	FLUXNET	9.8	750	Scots pine
Bílý Kříž forest	Czech Republic	CZ-BK	ENF	18.53	49.5	875	FLUXNET	7	1316	Norway spruce
Hainich	Germany	DE-Hai	DBF	10.45	51.07	430	FLUXNET	8.3	720	Mixed Beech
Hohes Holz	Germany	DE-HoH	DBF	11.21	52.08	217	COSMOS Europe	10	820	Mixed beech
Oberbärenburg	Germany	DE-Obe	ENF	13.72	50.78	734	FLUXNET	5.5	996	Norway spruce
Wüstebach	Germany	DE-Wue	ENF	6.33	50.5	605	COSMOS Europe	7	1180	Spruce
Gludsted	Denmark	DK-Glu	ENF	9.33	56.07	86	COSMOS Europe	8.2	1080	Spruce
Conde	Spain	ES-Cnd	WSA	-3.22	37.91	370	FLUXNET	15.8	474	Olive grove
Hyytiälä	Finland	FI-Hyy	ENF	24.29	61.84	181	LTER Europe	3.8	709	Boreal Scots pine
Sodankylä	Finland	FI-Sod	ENF	26.63	67.36	180	FLUXNET	-1	500	Boreal Scots pine
Puéchabon	France	FR-Pue	EBF	3.59	43.69	270	FLUXNET	13.5	883	Evergreen oak
Lavarone	Italy	IT-Lav	ENF	11.28	45.95	1353	FLUXNET	7.8	1291	Coniferous forest
Loobos	Netherlands	NL-Loo	ENF	5.74	52.16	25	FLUXNET	9.8	786	Scots pine

Figure 1: Would also be helpful to include information about data source in this figure too, similar to Table 1.

We agree and we have included the information the legend of Figure 1.

Line 116: 'partitions' is strange in this context. Suggest 'CLM5 simulates sensible and latent heat flux for both vegetated and ground fluxes'

We agree.

Changes in text:

Lines: 154 - 155

"CLM5 simulates sensible and latent heat flux for both vegetated and ground fluxes."

Line 120: same comment as above, suggest: 'canopy evaporation is represented as the sum of steam and leaf evaporation as a function of temperature.'

We agree.

Changes in text:

Lines: 158 - 160

"Interception, throughfall and canopy drip are explicitly modeled in CLM5 and canopy evaporation is represented as from the sum of stem and leaf surface evaporation as a function of temperature."

Line 127: How many ensemble members were used to sample the model uncertainty?

We used an ensemble of size 96 for all our simulations.

Changes in text:

Lines: 174 - 175

"In this study we use an ensemble of 96 member to sample the model uncertainty."

Line 144: "For simulations assimilating CRNS, H assigns the mean observed SWC to all the layers down to the measurement depth."

It is unclear what this means. Guessing from previous statements that CRNS is the integrated water content from 10-70 cm, so are you comparing against the CLM modeled soil layers that

coincide with 10-70 cm? You need to be more explicit of how you calculate your observation operator in general. What CLM soil layer variables are you using to calculate the expected observation? Also for the point measurement of soil water content, where are the depth measurements? I don't see this mentioned in this section where it would be appropriate.

I see you describe the FLUXNET soil water content in lines 176 through 179, but it seems out of place. This description should come earlier and discussion of the observation operator should be in one place. The statement 'For the CRNS sites, the measurement depth for each individual measurement is calculated following Schrön et al. (2017) and is included in the dataset from Bogena et al. (2022).' Seems to conflict with an earlier statement that the CRNS integrates between 10-70 cm. depth.

For CRNS the measurement depth varies as function of soil water content. The cited "10-70" cm is the usual range of the measurement depth. Since the reference data set provides the measurement depth for each observation the depth which is used by H also varies.

Changes in text:

Lines: 187 - 191

"For FLUXNET sites, measured soil water content is provided for up to three depths described as superficial, medium, and deep. Since data assimilation in CLM5-PDAF requires a specific vertical layer, we assigned 5, 20 and 50 cm to the respective FLUXNET SWC layers. For the CRNS sites, the measurement depth for each individual measurement is calculated following Schrön et al. (2017) and is included in the dataset from Bogena et al. (2022). "

Line 159: "By default, CLM5-PDAF updates soil hydraulic parameters through changes to fractions of sand, clay, and organic matter and the pedotransfer function of Clapp and Hornberger (1978)."

This statement and methodology is problematic. See my comment in summary section above.

See our comment above.

Line 162: It seems you are allowing the % sand, clay and organic to vary by layer? Does this make physical sense? This would allow for each layer to vary independently with no fixed bounds other than 0 or 100%? I cannot find any results that show the adjusted parameter values.

The sand, clay, and organic matter fraction already vary by layer even in the default CLM5 surface files. This makes physical sense, especially for organic matter that is usually very high in the first few layers and rapidly decreases for any deeper layers. During parameter updates the layers are updated using the Kalman gain, which ensures that spatial correlation in soil texture between soil layers is taking into account.

Changes in text:

Lines: 221 - 227

“We tested directly updating saturated hydraulic conductivity, porosity, hydraulic conductivity exponent, and soil matric potential but this resulted in more unstable estimates than indirectly updating soil hydraulic parameters. The pedotransfer function which is used for the indirect updating results in reasonably correlated soil hydraulic parameters. “In testing a direct approach to updating saturated hydraulic conductivity, porosity, hydraulic conductivity exponent B, and soil matric potential we found that updating the parameters indirectly to provide more stable simulations. The pedotransfer function keeps the soil hydraulic parameters reasonably correlated to each other.”

Line 174: You are perturbing the % sand, clay and organic to generate ensemble spread, but you are also estimating these parameters at the same time. How do you maintain ensemble spread, or prevent the system from collapsing on a solution?

It seems you address this in Line 192, but seems out of place to mention it here. Should come earlier.

Yes, the sand, clay, and organic matter fractions are all perturbed to create the ensemble spread and updated. However, the atmospheric forcing variables are also perturbed and create ensemble spread and prevent a collapsing on a solution.

Changes in text:

Lines: 218 - 219

“A damping factor of 0.1 is used on the parameter updates to avoid filter inbreeding and keep the ensemble spread larger so that the model error covariance matrix is a good approximation for model uncertainty.”

Line 226: “This indicates that the SWC-ET relation is incorrect for these sites. A possible explanation is the water uptake by roots in the deeper layers is underestimated for forest sites, as also suggested by Shrestha et al. (2018).”

I don’t see how you can come to this conclusion without diagnosing your leaf area (LAI) mismatch. See my general comment/concern above.

We rephrased the statement.

Change in text:

Lines: 311 - 313

“Another possible explanation for the improvement in SWC estimation but no improvement of ET estimation is the underestimation of root water uptake from deeper soil layers for forest sites, as also suggested by Shrestha et al. (2018).”

Line 315-325: Exactly. Prescribing LAI through CLM5-SP or assimilating LAI observations is necessary to demonstrate that soil water content does not improve ET, and that it is a parameter problem.

We agree that assimilating LAI will improve ET. However, we still think that it is noteworthy that data assimilation of high-quality in-situ soil water content provides no improvement to ET estimation for forested sites with CLM5-BGC.

Line 337: "These results suggest that state-of-the-art LSM such as CLM5 still suffer from uncertainties in the representation of soil hydrological processes in forests, e.g. deep root water uptake, uncertainties in the representation of biological processes of tree transpiration, partly related to uncertain vegetation parameters."

I disagree, at least I don't think your results demonstrate this. The inability to prescribe leaf area, or account for the assimilation of leaf area (Fox et al., 2022) does not allow you to make this statement with certainty.

We rephrased the statement to account for the lack of LAI adjustment in this study.

Changes in text:

Lines: 451 - 456

"These results suggest that improving the SWC estimation of state-of-the-art LSM such as CLM5 is not sufficient to improve ET estimation for forest sites. To improve ET estimation it is also necessary to consider the representation of LAI in magnitude and timing, as well as uncertainties in water uptake by roots and vegetation parameters."

Furthermore the parameter estimation performed here does not seem to directly update 'true' vegetation hydraulic parameters (e.g. vegetation hydraulic parameters; Kennedy et al., 2019) to further test this theory. You adjust soil hydraulic parameters which could influence soil moisture dynamics, but seems redundant in that soil moisture is already adjusted directly.

It also would have been useful to include the prior and posterior values for your parameter estimation. Its unclear if the lack of impact the parameter estimation had on ET, was because of lack of sensitivity to controlling the ET, or because the parameters were not updated significantly from their prior values.

Yes, we did not take vegetation hydraulic parameters into account in this study as we only update soil hydraulic parameters which is not necessarily redundant but usually provides increased improvement in SWC estimation. However, we agree that it is not enough for ET estimation and we will take vegetation parameters and LAI assimilation into account in our future studies.

Changes in text:

Lines: 227 - 230

“In this study, the parameters are chosen to optimize the SWC estimation and not ET estimation to study the effects of SWC improvements on ET. To more directly improve the ET estimation, parameters affecting the ET process directly should be added, e.g. vegetation hydraulic parameters.”

Figures 2-4: You could combine these plots to present a better comparison between OL-DAS and OL-DASP.

We combined the Figures 2-4.

Changes in text:

Lines: 272 - 281

“Figures 2 and 3 show the results of the soil water content simulations at 20 cm depth of the OL, DAS and DASP simulations compared to the soil water content observed at the nine sites. Figure 2 compares the OL and DAS results and Fig. 3 compares the OL and DASP results. The corresponding scatter diagrams for the depths 5, 20, and 50 cm can be found in the appendix (Figures A1 - A7). Overall, the results show expected improvements by data assimilation of observed soil water content. For the OL simulations, Fig. 2 shows particularly large RMSE values for CZ-BK, DE-Obe, FI-Sod and NL-Loo. Fig. 2 also illustrates the improved performance achieved by DASP, with a RMSE reduction from 29.3 cm^3/cm^3 to 6.25 cm^3/cm^3 and a MBE reduction from 28.06 cm^3/cm^3 to -2.94 cm^3/cm^3 for FI-Sod. Parameter updating, as shown in Fig. 3, further improves the simulation results, but the improvement from DAS to DASP is significantly less than from OL to DAS.”

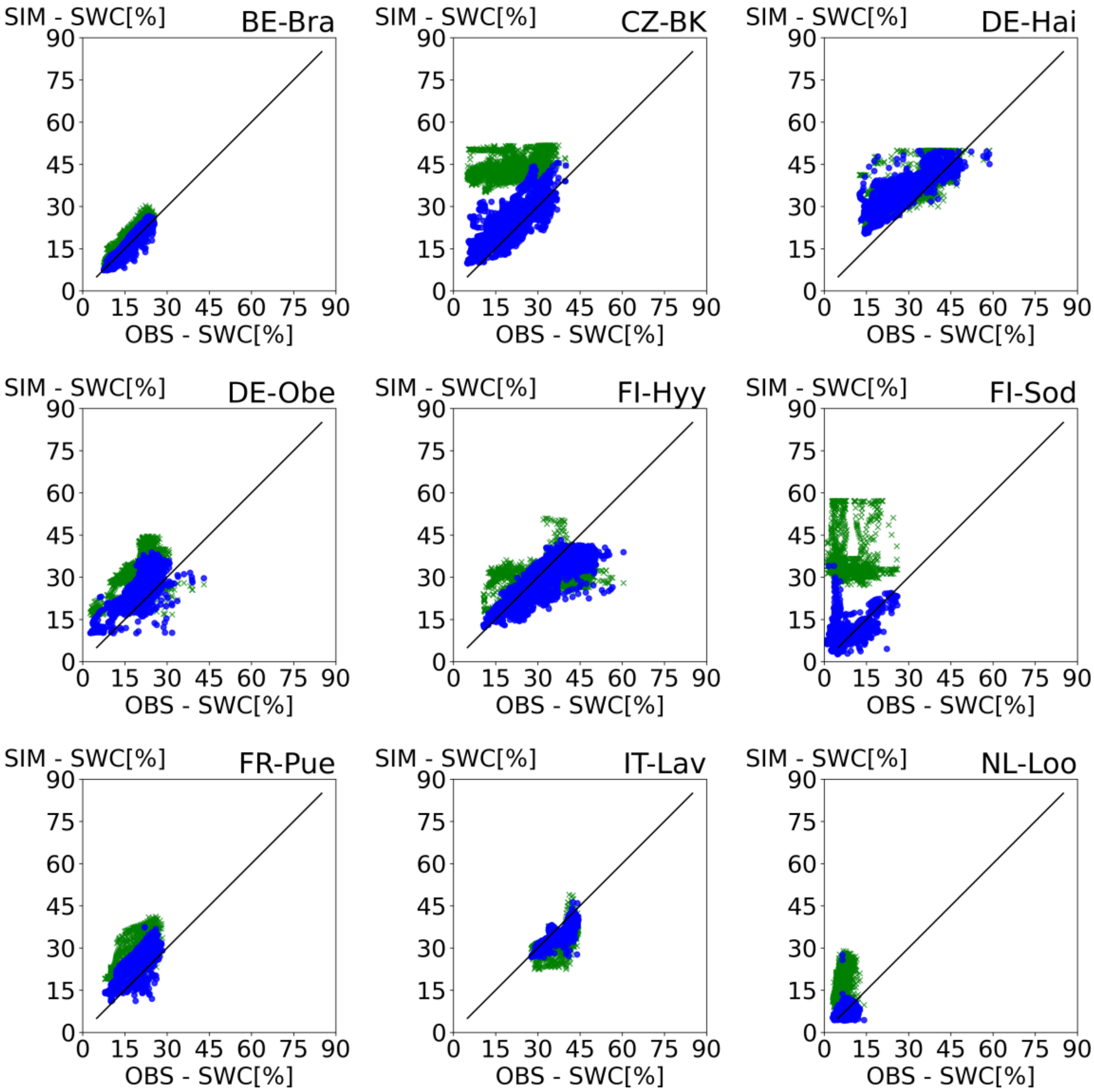

Figure 2: Scatter plots of observed soil water content at 20 cm depth at nine study sites versus OL and DAS simulated soil water content. The points represent daily averages for the days observation data are available. Green points are OL and blue points are DAS results.

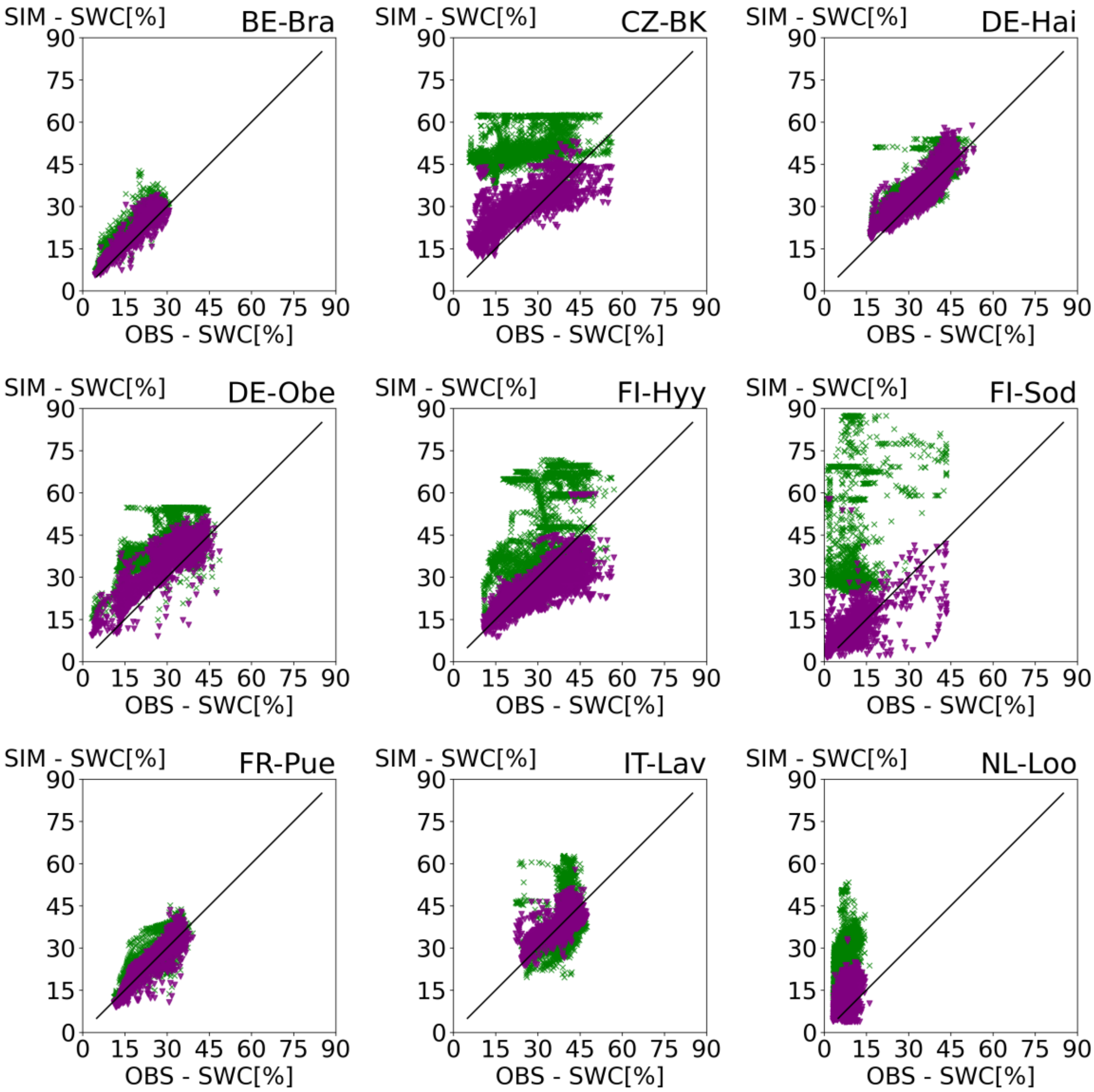

Figure 3: Scatter plots of observed soil water content at 20 cm depth at nine study sites versus OL and DASP simulated soil water content results at 20 cm depth. The points represent daily averages for the days observation data are available. Green points are OL and purple points are DASP results.

Figure 11: This figure seems very critical, given the potential mismatch between observed and simulated LAI, and the potential impact that has on the soil water content vs. ET relationship. I feel like you also need to present the average season cycle of LAI for all these sites, because the the annual average is not enough. CLM has trouble simulating the timing of seasonal phenology of LAI especially for evergreen forested sites (not just magnitude). Would help to simulate a CLM5-BGC and CLM5-SP simulation here, where LAI is prescribed, or assimilate LAI observations directly for your experiment.

We added a figure with the seasonal LAI cycle for each site and include a comparison with simulations where LAI is prescribed as mentioned earlier. In addition, we will work with LAI assimilation in future studies.

Changes in text:

Lines: 303 – 306, 354 - 355

“This could be caused by the mismatch of simulated and actual LAI for these sites. To investigate this, we repeated the simulations using CLM5 with satellite-derived phenology (CLM5-SP) The results are shown in Fig. 9. For CLM5-SP we observe an average improvement in the RMSE of SWC between 57.6 % and 64.3 % and an average RMSE deterioration of 5.8 % for the ET estimation. These CLM5-SP simulations use the default datasets from CLM5 and without site specific calibration of the timing or magnitude of the seasonal phenology of LAI. Therefore, even for the CLM5-SP simulations, there is a mismatch between simulated and actual LAI. However, also for this case, there are sites with a large improvement in SWC estimation that show deterioration in the ET estimation.”

“Figure 14 shows the LAI for each site averaged over all the simulated years. Sites with the same PFT show clear differences in the yearly LAI cycle. ”

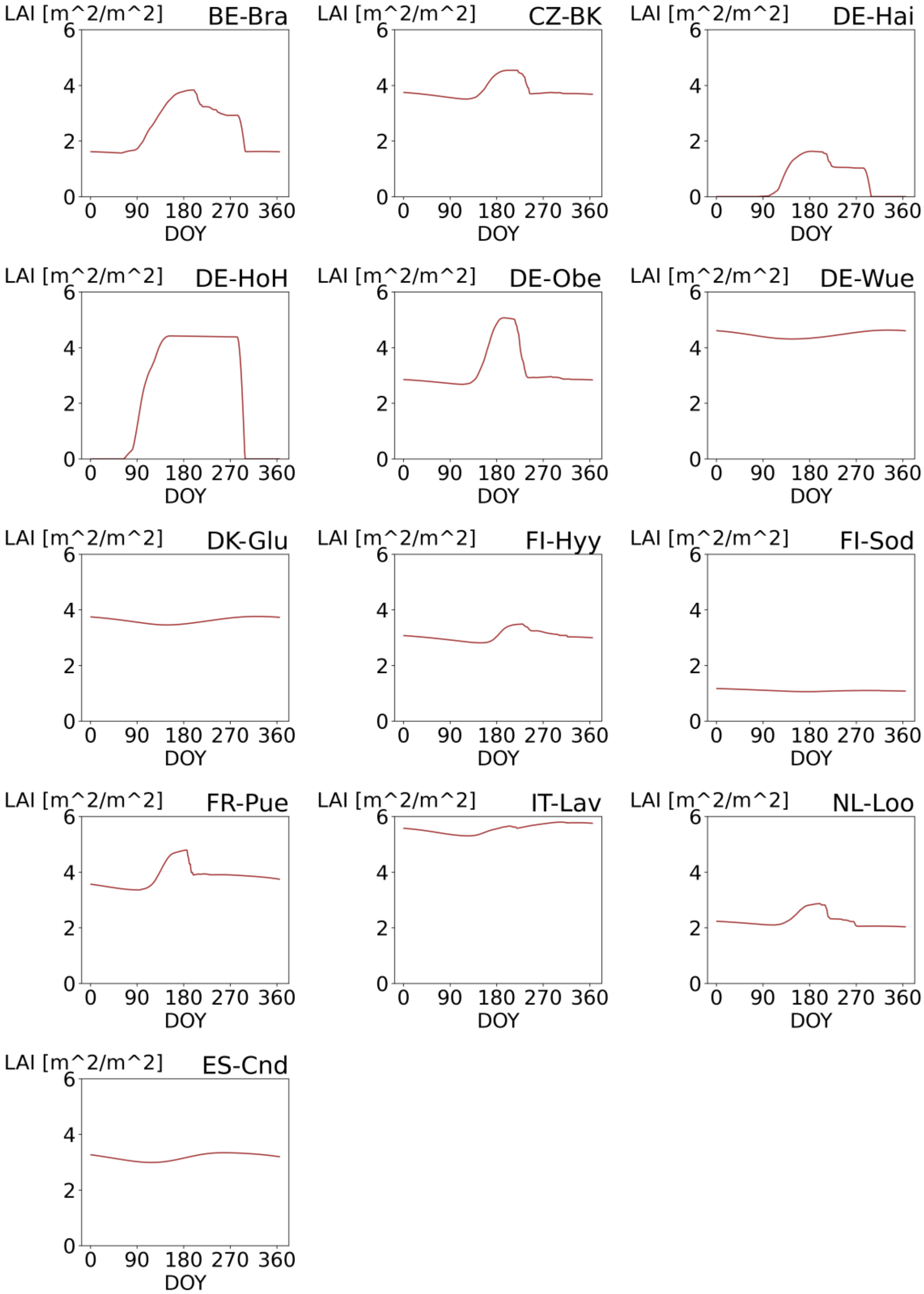

Figure 14: Seasonality of simulated LAI (DASP) for each of the sites.

Figure 12: These plots could be useful to diagnose the behavior, but I am afraid information is 'washed out' when averaging over all years and all sites. This brings up some really important questions, such as, did the assimilation adjust the SWC all the way through the

root zone, or was the adjustment primarily superficial, and limited to the upper 25-50 cm. only? Is the majority of the root zone below this, and contribute to the lack of impact on ET?

We show results averaged over all years and sites to give an overview on general trends in the results. However, we agree that additional figures can shed a further light on the results. The data assimilation adjusts the SWC for all layers, but as shown in this figure the change is usually larger in the upper layers than in deeper layers.

Changes in text:

Lines: 341 - 343

“Although DASP adjusts SWC at 5cm towards the observations, the correction for SWC at 50cm depth is smaller because not all sites provide data at this depth. However, for all sites the data assimilation provides some improvement for SWC estimation even in layers below the observation depth.”