Anonymous Referee #2, 03 Aug 2022

This manuscript is an implementation of the ParFlow-CLM at high resolution (3 km) focused upon the European domain. The validation of the model performance is a wide-ranging analysis based upon remotely sensed soil moisture, and ET, as well as ground-based data products of soil moisture, SWE, ET, groundwater depth, and streamflow. It is generally well written, although there is a lack of focus in the key findings. The authors attributed deviations from observed site level behavior (e.g. positive SM and ET bias) primarily to uncertainties with the incoming atmospheric forcing. However, it seems likely that uncalibrated parameters could have just as easily led to these biases.

We would like to thank the anonymous reviewer for his/her comments and constructive suggestions, which we believe resulted in an improved manuscript. We replied to your comments in the blue text below.

We agree with your comment that biases in our results could be due to errors in the model inputs or/and due to the fact that the model was not calibrated. To better characterize these uncertainties, we now compared our results with other global studies and provided a more detailed comparison with the CONUS implementation of ParFlow-CLM (O'Neill et al., 2021). In the revised manuscript, these discussion points are included in a separate "Discussion" section, which also puts more emphasis on the key findings.

The authors motivate the analysis by claiming high spatial resolution combined with a representation of lateral groundwater flow is necessary for improved region wide prediction of hydrological variables. However, this reviewer did not find compelling evidence to demonstrate these assertions from this analysis alone, partly because the model skill was not put in context of other simulations. For example, implementing a coarse version of ParFlow CLM, or a version without lateral ground-flow could have better demonstrated these points.

We agree with the reviewer's comments that a multi-model comparison for uncertainty assessment is important in order to better quantify whether biases stem from either model structural errors or from the model resolution, particularly for models with a lateral groundwater flow representation. However, the aim of this study is to implement and evaluate the ParFlow-CLM model performance in space and time relative to observations, which we believe is also helpful to identify biases in water balance components and problem areas that could be improved in future studies. The novelty of the model implementation lies in a fully 3D represented subsurface flow, integrated with 2D overland flow at a high km-scale resolution for a continental model domain. In order to use this implementation in a wide range of scientific applications where an accurate representation of groundwater and surface water interactions is critical (e.g. climate non-stationarity, coupled ESMs, water resources assessments) we think a comparison to observations is sufficient to evaluate the model's performance and that a sensitivity analysis with multiple model resolutions is beyond the scope of the manuscript. We value your suggestions to include a comparison with a coarser-resolution implementation of ParFlow-CLM, or a version without lateral flow; we have now discussed this as possible avenues for further tuning and future work. We believe that results from this study can be used as a baseline for future ParFlow-CLM implementations over Europe and will be used to guide future model development.

This manuscript is, in fact, complementary to a similar implementation of ParFlow-CLM for the CONUS domain (O'Neill et al). Yet, the author's do not fully address this point until late in the conclusions, and miss an opportunity to provide a more rigorous comparison between the CONUS and European domain performance with ParFlow-CLM.

We appreciate your suggestion and believe this will strengthen our manuscript. We now provide a detailed and extensive comparison of our results with the CONUS implementation of ParFlow-CLM (O'Neill et al., 2021). Please see our response below.

It is challenging to evaluate this manuscript because in one sense the methods behind the model implementation and evaluation are useful to the LSM or hydrology community. This validation approach (use of statistics based on comparison to RS and site-based observations) could be used as a template for benchmarking other models. Furthermore, this 'evaluation of a previously published model' does fulfill one of the criteria for publication in GMD. On the other hand, the comparison between the model simulation and remotely-sensed and ground based observations lacked a clear focus. Detailed comments are below.

Thanks for the positive response. We have revised the manuscript based on your constructive comments and suggestions. We agree that in some areas the focus could be strengthened, we have taken this on board and have revised the manuscript accordingly.

Line 21: It is a bit confusing what the authors mean by high resolution hydrological modeling, and large-scale hydrologic processes. Better quantification?

Thank you for your suggestion. We now clarify these terms by referring to high resolution hydrological modeling (< 5 km) and modify text "large-scale hydrological processes" to large-scale spatial patterns of hydrological processes (i.e. streamflow, evapotranspiration, soil moisture and total water storage).

Line 30: LSM's are also used commonly for carbon and nitrogen cycling research. Both LSM's and GHMs solve water balance equations.

Thanks for the comment. We revised the text as: *"Numerical models that attempt to simulate large-scale hydrology and associated processes are usually categorized as land surface models (LSMs) or global hydrological models (GHMs), which have been developed for simulating the land surface water, energy and momentum exchange (Sellers et al., 1988) to provide water balance estimates at global to continental-scale."*

Lines 40-50: The author seems to be conflating two things: issues of spatial resolution, or issues related to physical processes. It is true a coarse scale model will not capture fine scale hillslope topography which could be important for watershed scale studies, but is this necessary for global scale climate models?

Thanks for your comment. To address this comment, we revised the text (Line 30 – Line 50) as:

"Numerical models that attempt to simulate large-scale hydrology and associated processes are usually categorized as land surface models (LSMs) or global hydrological models (GHMs), which have been developed for simulating the land surface water, energy and momentum exchange (Sellers et al., 1988) to provide water balance estimates at global to continental-scale. Despite the extensive work in large-scale hydrology modeling (e.g. Clark et al., 2015), many of the existing large-scale hydrological models (both LSMs and GHMs), especially those intended for continental- to global-scale simulations are single-column models (e.g., Döll et al., 2003; Hunger and Döll, 2008; Gudmundsson et al., 2012; Haddeland et al., 2011), for which most hydrological processes are implemented empirically and at a coarse spatial resolution (typically 25 km to 100 km). As a result, many of the

important hydrological processes are simplified, including groundwater and surface water dynamics, soil moisture re-distribution and evapotranspiration (Clark et al., 2017). A physics-based integrated hydrological model, on the other hand, which can simultaneously solve surface and subsurface systems with lateral-groundwater flow may provide better prediction of both local and global water resources (Beven and Cloke, 2012). At finer resolution, processes-based integrated hydrologic models can better represent heterogeneity in the representation of water and energy states and fluxes when run at high spatial resolution (< 5 km) due to the higher resolved surface properties that help in providing a more accurate representation of the lateral transports of surface and subsurface water movements driven by topographic slopes (Ji et al., 2017; Shrestha et al., 2015). However, the effect of these important processes on water and energy states and fluxes is still not fully understood, especially over continental scales and comparison across different landscapes is needed."

Line 77: You need to spell out remote-sensing (RS) the first time you use it.

It has been modified in the revised manuscript.

Line 90: What is the difference between Parflow-CLM, PF-CLM and PF-CLM-EU3km?

We replaced the PF-CLM-EU3km to ParFlow-CLM throughout the manuscript.

Line 97: Renaming a model to PF-CLM-EU3km usually means you have changed the model equation/structures/parameterizations. I don' think the author's do that here – it is simply the PF-CLM or Parflow-CLM model run at a certain spatial domain (Europe) and at 3 km resolution. A 'new' model hasn't been designed or developed…..

We replaced the "PF-CLM-EU3km" to "ParFlow-CLM" throughout the manuscript.

Section 2.0.2 It is completely unclear what is novel about your implementation of ParFlow-CLM other than the domain and resolution. This seems like a model application and not novel development.

The focus of our study is the assessment of the model performance and for this reason we submit this manuscript as a "model evaluation" type to GMD. To make our objectives and research goals clearer, we expanded the Introduction and Discussion sections to emphasize and clarify the following points:

1. The aim of this study is to implement and evaluate the performance of ParFlow-CLM model which is a physically-based integrated hydrological model and simultaneously solve surface and subsurface systems with lateral-groundwater flow. The lateral groundwater flow is key - many modeling systems implemented at continental or global scales are one dimensional and contain a parameterized version of groundwater flow (Felfelani et al., 2020; Wada et al., 2016; Zeng et al., 2018; de Graaf et al., 2015). We have strengthened this point in the Introduction and Discussion sections. At finer resolution, a physically-based integrated hydrological model can better represent groundwater surface water interactions, and heterogeneities in the representation of states and fluxes of the water and energy cycles when run at high spatial resolution (< 5 km) due to the higher resolved surface properties. In addition, owing to ParFlow's 3D flow implementation and run in a continuum approach with 2D overland flow, this

model setup provides a more accurate representation of lateral transport of surface and subsurface water movements driven by topographic slopes (Shrestha et al., 2015).

2. The novelty of this study lies in the fact that it is the first study to implement ParFlow-CLM over the EU-CORDEX domain at high (km-scale) resolution, which allows fully three dimensional flow. In addition, a comprehensive model evaluation is given for multiple variables using both in-situ and remote sensing observations, in comparison to similar European studies such as Bouaziz et al. (2021); Rakovec et al. (2016); Zink et al. (2017). Several studies exist on ParFlow-CLM, but mostly concentrating over the CONUS region, therefore we believe that implementation of this model outside CONUS is a step forward towards “Hyperresolution global land surface modeling” which is considered a “grand challenge in hydrology” as described by Wood et al., (2011) and Bierkens et al., (2015), also in the context of coupled km-scale regional climate system models. We have strengthened this point in the Introduction.

3. Explicitly incorporating hydrological processes that are not included in the existing land surface models (LSMs) can also benefit the land surface or regional climate modeling community for a more improved representation of hydrological processes (Clark et al., 2015) such as the lateral transport of surface and subsurface processes across landscapes that are often ignored or crudely represented in LSMs. Many recent studies showed the importance of representing the lateral transport of subsurface water and/or interaction of groundwater with land-atmosphere water fluxes (e.g., Barlage et al., 2021; Keune et al., 2016; Maxwell and Kollet, 2008; Miguez-Macho and Fan, 2012; Miguez-Macho et al., 2007; Xie et al., 2012; Zeng et al., 2018). These studies suggested that explicitly simulating these processes can have a significant effect on the surface energy fluxes and flux partitioning (Maxwell and Condon, 2016). It can also affect the spatial redistribution of soil moisture through infiltration during lateral movement of water (Ji et al., 2017). Despite this important work, the effect of these important processes on water and energy states and fluxes is still not fully understood, especially over continental scales and comparison across different landscapes is needed. While representations of these important processes continue to improve in continental to global scale hydrological models, implementation and rigorous evaluation of these models over large areas is an important step and can be used to guide future modeling efforts at larger spatial scales and higher resolutions.

Line 134: Not clear what ‘inscribing’ into the Eur-11 grid means.
The model domain is inscribed into the official Coordinated Regional Downscaling Experiment (CORDEX) EUR-11 model grid (about 12 km). This has been clarified in the revised manuscript.

Line 144: CLM3.5 is from the Community Land Model, different than the Common Land Model (CLM) described here within ParFlow-CLM.
Correct. In the revised manuscript, we defined Community Land Model (v3.5) as CLM3.5 and Common Land Model as CLM.

Section 2.0.4 It seems unlikely that nine years of spinup would be enough to reach equilibrium between prescribing vegetation conditions and subsurface soil moisture state. Did the author’s check that the hydrological variables approached an equilibrium. It is also typically not normal to spinup with a single year (1997), you would want to spinup up overall several years (decade if possible) to capture variation in met forcing.

We followed a similar approach as used by other studies to spin up the ParFlow-CLM model (Maxwell and Condon, 2016, O'Neill et al., 2021, Shrestha et al., 2015; Shrestha et al., 2018). Most land surface models and water balance models need to spinup over several years owing to the absence of lateral flow and parameterization of physical processes in their model structure. Due to the physics-based model structure of ParFlow-CLM, spin up of the model over a period of one year, which is run multiple times in a closed loop, is deemed sufficient to reach equilibrium and has been shown to be sufficient in the previous studies mentioned. We ran the model continuously until the total water storage change was less than 2 % from the previous years, as per the methodology in the published studies. We have clarified this point in the revised manuscript.

Line 269: “Because of the explicit lateral groundwater and surface flow representation, we show that the PF-CLM270 EU3km model is able to resolve multi-scale spatial variability in hydrological states and fluxes such as simulated river flow, SM, ET and WTD distributions which are strongly correlated with the river network and topography as shown in Fig. 1.”

I am not sure I found any evidence of this causal relationship.
We revised Fig. 1 in the manuscript to include topography information as shown below.

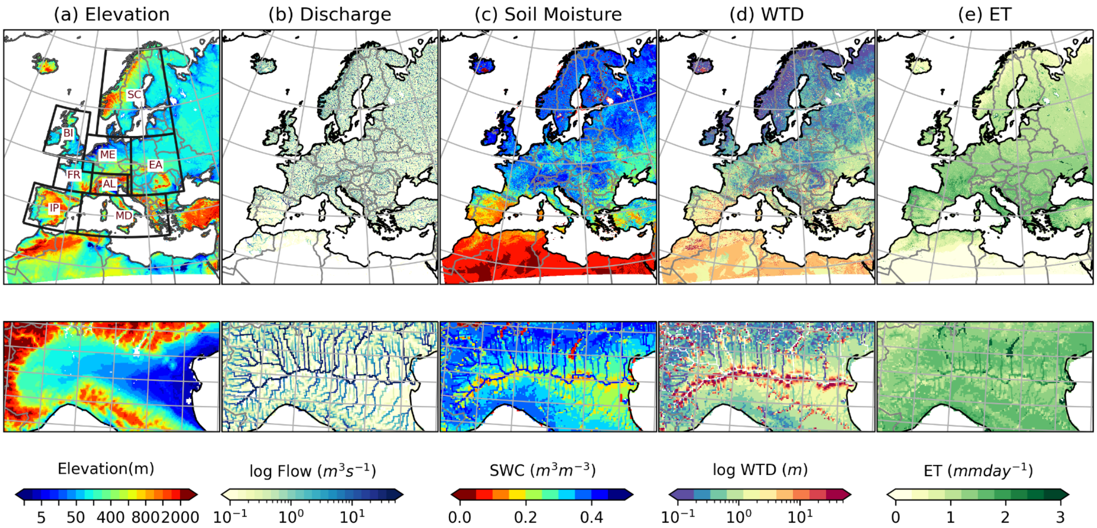

Figure 1: (a) Maps of EURO-CORDEX domain at 3 km resolution (1544 x 1592 grid cells) showing the spatially 10-year average distribution of (a) Elevation (b) discharge, (c) surface soil moisture, (d) water table depth, and (e) evapotranspiration (1997–2006) and close-up of Po river basin in Alpine (AL) region simulated by ParFlow-CLM model.

In addition, we compared as an example the spatial variability in surface soil moisture simulated by ParFlow-CLM for January and August months, 2000 for two regions (Alpine and Mid-Europe) with ESSMRA dataset (Naz et al., 2020) which is the assimilated soil moisture simulated by CLM3.5 to highlight the differences in spatial variability between the two models as shown below in Fig. S1 and Fig. S2. As shown in these figures, spatial structure simulated by the two models differs remarkably. CLM3.5 shows much larger spatial patterns of SM which are mostly related to the soil properties (e.g. soil texture information), while in ParFlow-CLM simulates more spatial variability, which can be attributed to the effects of 3D flows in river networks and across topography. Please note that both models use identical surface information (topography, soil and vegetation) and forcing datasets indicating that these differences are explained by the fine-scale processes (such as surface and subsurface lateral transport of water movements and the shallow groundwater system) simulated only by ParFlow-CLM.

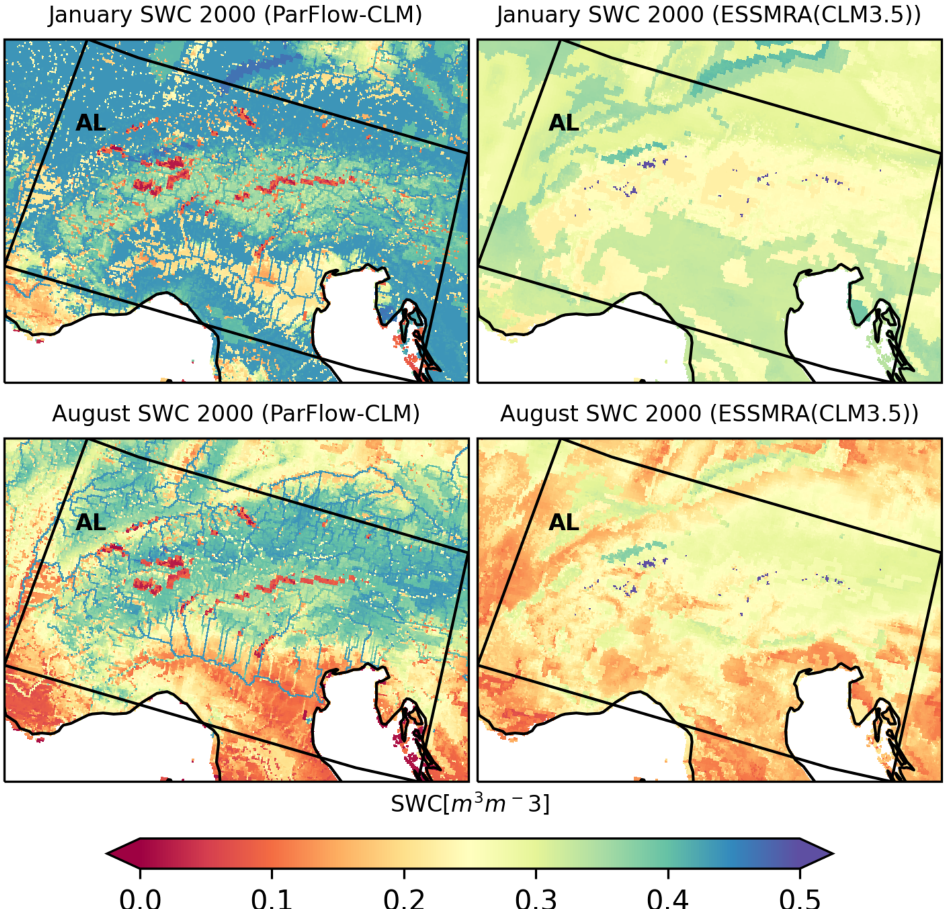

Figure S1. Spatial variability of surface soil moisture simulated by ParFlow-CLM and CLM3.5 at the surface soil layer for January and August months of year 2000 over the Alpine region. Please note that glacier areas were not simulated by ParFlow-CLM and soil moisture values are zero at those grid cells.

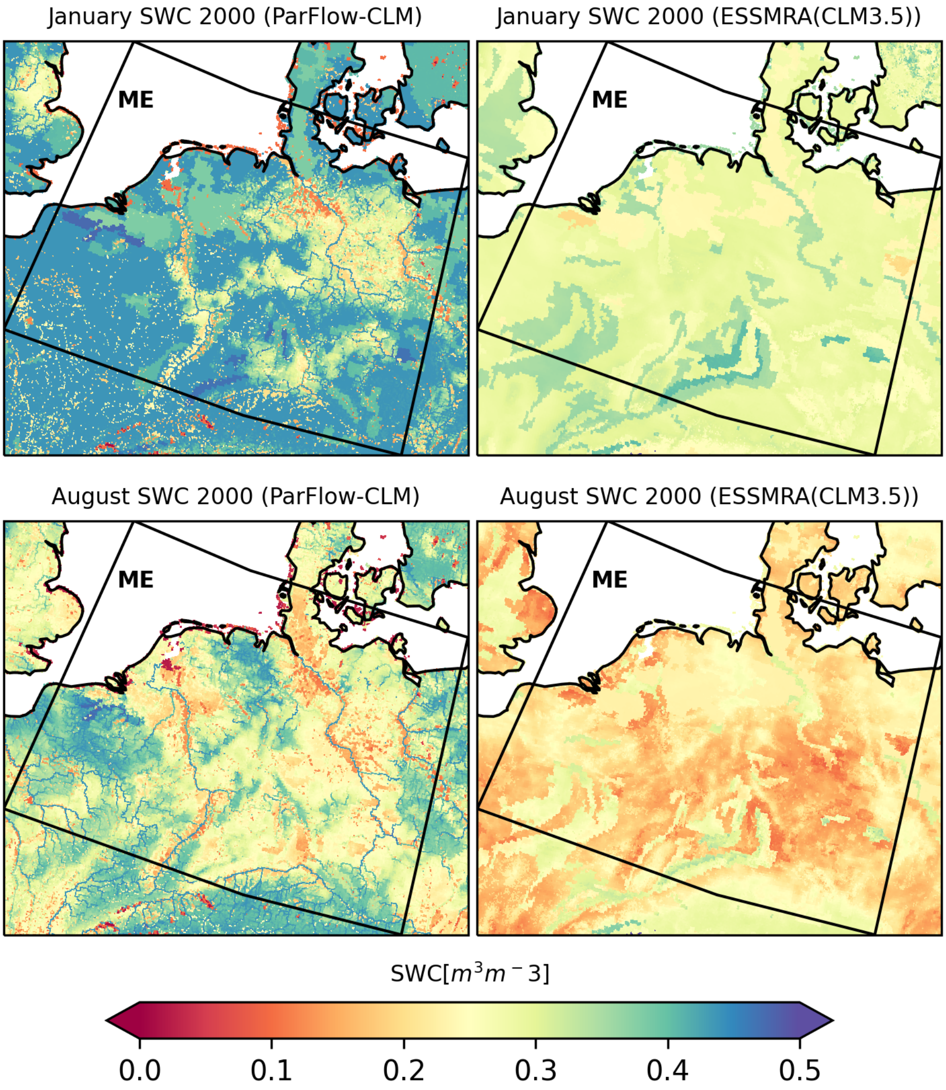

Figure S2. Spatial variability of surface soil moisture simulated by ParFlow-CLM and CLM3.5 at the surface soil layer for January and August months of year 2000 over the Mid-Europe region. Please note that glacier areas were not simulated by ParFlow-CLM and soil moisture was set to zero.

Line 339: "The difference is explained by the shallow groundwater system simulated only by PF-CLM-EU3km, which contributes to the saturation of the deeper soil layers leading to higher soil water content, whereas the standalone CLM3.5 model applies a simple approach to simulate groundwater recharge and discharge processes in a single column and neglects explicit lateral groundwater flow."

It appears here that the authors are attempting a comparison against CLM3.5 (the Community Land Model) which was used as the LSM to develop the ESSMRA product, and comparing against the PF-CLM-EU3km. Claiming the differences in SM can be accounted for by differences in the accounting of lateral groundwater flow. This is a complicated comparison for many reasons, one of them being that the ESSMRA product includes observations of the ESA-CCI 'observations'. The PF-CLM-Eu3km does not. It is not a controlled comparison to claim lateral groundwater flow is the cause for the differences…..

We agree that it is not a controlled comparison due to the fact that the ESSMRA product also includes CCI observation. However, ParFlow-CLM also shows higher soil moisture when comparing to the CLM3.5 simulated soil moisture with no assimilation of ESA CCI (not shown here) which indicates that the difference between the surface SM could be attributed to the shallow groundwater system simulated only be ParFlow-CLM. We will include this comparison in the supplementary material.

It's also extremely confusing that CLM3.5 (Community Land Model) is not the same as the "CLM" (Common Land Model) in PF-CLM-EU3km.
Sorry for the confusion. In the revised manuscript, we defined Community Land Model (v3.5) as CLM3.5 and Common Land Model as CLM.

Figure 4: Not clear what we can hope to learn by comparing 3 separate SM products against each other. Would it not be more helpful to compare the performance of the SM products against in-situ site ISMN observations? I see that this comparison is pushed to the supplement.

We agree that it would be helpful to compare the model performance of the SM against in-situ data (as we have shown in the Figure S5 - S8 in the Supplementary materials). However, as we have indicated previously, there is observational data sparsity across Europe and for the time period of 1997–2006, data for only 20 grid cells are available which are useful to evaluate model performance at those point locations but unfortunately useless to evaluate spatial variability in SM over large domains. Therefore, to evaluate the model performance at large spatial scale, we compared with other gridded products of surface SM which provide far greater coverage and helps to evaluate model performance for spatial signature over different regions influenced by different climatic characteristics.

Line 387: "Previous studies of PF-CLM-EU3km also indicate……"

Apparently this exact implementation of this configuration of the CLM ParFlow has been done before? Still failing to see the novelty of the study?
As we mentioned in the introduction section, that previously ParFlow model has been employed over the European CORDEX domain at 12 km resolution with 1D and 3D subsurface flow within the framework of fully integrated soil–vegetation–atmosphere model where the focus was to investigate the impact of extreme events on the water and energy fluxes through feedback mechanism (e.g. Keune et al., 2016; Keune et al., 2018; Furusho-Percot et al., 2019; Hartick et al., 2021), however, the model performance was not rigorously evaluated for all water balance components. This study is the first study that implemented and evaluated ParFlow-CLM over the EU-CORDEX domain at 3 km resolution, with fully 3D flow, for multiple variables using both in-situ and remote sensing observations. Implementation of process-based models such as ParFlow-CLM that allow to fully resolve the fine-scale surface and subsurface processes and spatial heterogeneities at continental scale is also a step forward towards "Hyperresolution global land surface modeling" which is considered as "grand challenge in hydrology" as described by Wood et al. (2011) and Bierkens et al. (2015). The results from this study can be used as a baseline for future ParFlow-CLM implementations over Europe and will be used as a guide for future model development.

Figure 5: It would be more compelling to show mean seasonal cycles for a sampling of sites (model vs. flux tower ET) across a variety of biomes. Seasonal correlations (as shown) should be strong, just based on phenology of vegetation, as well as increase/decreases in SW radiation. You show regional plots in Figure 6, but running at high resolution grid (3 km) should allow you to make direct comparison to flux tower ET data. It is less compelling to show seasonal variation with GLEAM and GLASS given these are data products.
Comparison with the eddy covariance sites from FLUXNET datasets has already been shown in the manuscript and in the supplementary materials as shown in Fig. 5 and Fig. S9.
As mentioned in the previous comment, the point source based sites provide hugely deficient coverage and therefore comparing with other satellite-based gridded ET products allows us to evaluate model performance over large spatial scales to better understand both seasonal and spatial variability for different regions influenced by different climatic conditions.

Line 417: “Figure ??” typos show up a few times in this manuscript.
Corrected.

Line 469: “Our comparison of simulated SWE with observed SWE reveals an overprediction of SWE in the Eastern regions which is more likely to be related to the uncertainties in precipitation.”

I don’t follow how the authors came to this conclusion. Could not biases in SWE be a result of uncertainties in temperature, or from issues with the snow/energy balance model which simulates accumulation and depletion of snowpack? If some sort of evaluation against in-situ site atmospheric observations was performed that could provide more credibility.

We agree with the reviewer that biases in SWE could be caused by many sources of uncertainties, as discussed in Section 3.1. In the discussion section we now revised this sentence as:

“Our comparison of simulated SWE with observed SWE reveals an overprediction of SWE in the Eastern regions which is more likely to be related to the uncertainties in forcing datasets or model structure errors in simulating the snow/energy balance.”

Line 481: “The rigorous evaluation of the PF-CLM-EU3km model over Europe together with the recent study by O’Neill et al. (2021) which evaluated model performance over CONUS paves the way towards a global application of fully distributed physically-based hydrologic models.”

This is the first time, at the end of the manuscript, where the authors mention this serves as a companion paper to the CONUS implementation of the same model. This manuscript would have been much more compelling if comparison in performance were discussed between the CONUS and EU implementations throughout. Or to quantify the benefit of high resolution implementation of this model, with subsurface, later flow against other LSM’s at coarse resolution, or lacking later, subsurface flow.
Thank you for your comment. To provide more discussion on how our model differs from other existing implementations of ParFlow-CLM, we compare our results with the CONUS implementation of ParFlow-CLM model (O'Neill et al., 2021) as shown in Table 1 below. As stated previously, the CONUS domain does not suffer the same data sparsity issues as the European domain and because of different domains, resolution and climatic conditions, a

direct quantitative comparison is not possible. We, however, concluded from this comparison the following points:

Streamflow: Both modeling setups show good agreement with observation from gauge stations in terms of temporal dynamics. However, the EU-CORDEX model shows negative biases for the majority of the stations, whereas, the CONUS model simulates higher positive biases for many gauge locations.

ET: A comparison to the FLUXNET sites shows that both model implementations show overall high correlations for all sites but overpredict ET for most sites. In regard to the remote sensing (RS) comparisons, CONUS implementation overpredicted ET in the dry regions (e.g., south west) but underpredicted ET in more wetter and snow dominated regions (i.e., in the northern and eastern part of the domain) relative to the MODIS ET data. We see a similar behavior of the EU-CORDEX model when compared with the GLEAM dataset, which showed a slight underprediction in the north eastern part of Europe (more snow dominated) and a small overpredication in the southern part (relatively dry regions). However, in comparison to the GLASS ET dataset, which is a MODIS based product, ParFlow-CLM underestimated ET. In addition, both model implementations show an underestimation of ET in mountainous regions, regardless of which product is used for validation.

Soil moisture: For surface soil moisture comparison, both EU-CORDEX and PfCONUSv1 models show similar performance with correlation (R) values between 0.17–0.77 and 0.25–0.77, respectively across different regions. Interestingly, overall both model implementations show an underestimation of surface SM in the dry regions and overestimation in the wetter regions. Similarly both implementation show lower correlation values for regions with dense vegetation, complex topography, snow cover and frozen soil (i.e. upper Colorado in the CONUS domain and Scandinavia in the EU-CORDEX domain), which might be due to the large uncertainties in the ESA CCI data for areas with such conditions.

WTD: We find a good agreement between the ParFlow-CLM and observed WTD with a mean difference of -3.60 m, RMSE of 4.25 m and 25th, 50th and 75th quantile for simulated minus observed WTD are -2.6 m, -1.37 m and -0.84 m, respectively. Negative values in WTD difference indicates shallower WTD simulated by ParFlow-CLM. Despite this wet bias, the model is able to capture the temporal dynamics well with $R > 0.5$ for more than 50% of locations. For the CONUS implementation, O'Neill et al. (2021) showed similar wet bias for most locations which they found to be aquifer-dependent with greatest wet biases occurring for aquifers experiencing the highest rate of depletion in the past.

TWS: Both models show good agreement for TWS anomalies relative to GRACE satellite data in terms of temporal dynamics. EU-CORDEX setup simulated much stronger dry anomalies in the dry regions (MD and IP regions) and overpredicted wet anomalies for snow dominated regions (e.g. Scandinavian region).

The summary Table 1 shows a similar performance among the EU-CORDEX domain setup and the CONUS domain setup, giving additional confidence that the EU-CORDEX model implementation is performing adequately.

Table 1: Summary of ParFlow-CLM model performance for different variables and its comparison with CONUS implementation described by O'Neill et al. (2021).

	This study (EU-CORDEX)			O'Neill el al 2021 (CONUS)			Comparison
Variable	Datasets used	R	pbias (%)	Datasets used	R	pbias (%)	
Streamflow	GRDC gauge stations (monthly)	0.77	-16 % (50th percentile)	USGS gauge stations (daily)	0.65 (50th percentile)	41.3 % (50th percentile)	PFCONUSv1: higher positive bias, EU-CORDEX: higher negative bias
ET	eddy covariance towers from FLUXNET dataset (daily)	0.94		eddy covariance towers from FLUXNET dataset (daily)	0.72 (50th percentile)	37.9% (50th percentile)	PFCONUSv1: positive bias, EU-CORDEX: positive bias
	RS-based GLEAM and GLASS datasets (monthly)	0.91, 0.91 (50th percentile)	-9.9% and -18.2% (50th percentile)	RS MODIS dataset (MOD16A2) and SSEBop (monthly)	0.85 and 0.91 (50th percentile)	14.2% and 13.2% (50th percentile)	PFCONUSv1: Underpredicts ET in the north/east (wet/snow regions) and overpredicts in the south (dry regions). Underpredicts ET in the mountainous regions. EU-CORDEX: underpredict ET in the wet/snow regions, small overpredications in the south (dry regions). Underpredicts ET in the mountainous regions.
Soil Moisture	ESA-CCI (monthly)	0.70 (50th percentile)		ESA-CCI	0.69 (50th percentile)		PFCONUSv1: shows overall lower amplitude in the west (dry) and higher amplitude in the east (wet) relative to the CCI product; EU-CORDEX: overall wet bias, dry bias in southern Europe
TWS	GRACE dataset (monthly)	ranging from 0.76 and 0.91 for major regions		GRACE dataset (monthly)	ranging from 0.43 to 0.94 for major basins		Both model setups show stronger dry anomalies and overpredict wet anomalies relative to the GRACE data.
WTD	groundwater monitoring wells	0.50 (50th percentile)		groundwater monitoring wells	0.46 (50th percentile)		PFCONUSv1: a shallow WTD bias, EU-CORDEX: a shallow WTD bias

Line 483: “The protocol of evaluation metrics and methods presented in this study and in O’Neill et al. (2021) can be used as a framework to benchmark future PF-CLM-EU3km model implementations to further improve model simulations in the areas that have been identified or to explore the impacts of groundwater on 485 simulated hydrological states and fluxes by comparing with other existing global land surface model applications.”

Again, it would be more compelling if this manuscript performed a direct comparison of performance against the CONUS implementation or existing global land surface model applications to demonstrate improved utility/skill.

Thank you for your comment. To provide more discussion on how our model differs from other existing implementations of ParFlow-CLM, we compared our results with the CONUS implementation of ParFlow-CLM model (O'Neill et al., 2021). Please see our response to the previous comment.

References:

Barlage, M., Chen, F., Rasmussen, R., Zhang, Z., and Miguez-Macho, G.: The Importance of Scale-Dependent Groundwater Processes in Land-Atmosphere Interactions Over the Central United States, Geophysical Research Letters, 48, e2020GL092171, https://doi.org/10.1029/2020GL092171, 2021.

Beven*, K.: How far can we go in distributed hydrological modelling?, Hydrol. Earth Syst. Sci., 5, 1–12, https://doi.org/10.5194/hess-5-1-2001, 2001.

Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., and Famiglietti, J. S.: Hyper-resolution global hydrological modelling: what is next? “Everywhere and locally relevant,” Hydrological processes, 29, 310–320, 2015.

Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., De Niel, J., Dewals, B. J., Drogue, G., Grelier, B., Melsen, L. A., Moustakas, S., Nossent, J., Pereira, F., Sprokkereef, E., Stam, J., Weerts, A. H., Willems, P., Savenije, H. H. G., and Hrachowitz, M.: Behind the scenes of streamflow model performance, Hydrol. Earth Syst. Sci., 25, 1069–1095, https://doi.org/10.5194/hess-25-1069-2021, 2021.

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., and Zeng, X.: Improving the representation of hydrologic processes in Earth System Models, Water Resources Research, 51, 5929–5956, https://doi.org/10.1002/2015WR017096, 2015.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, Hydrol. Earth Syst. Sci., 21, 3427–3440, https://doi.org/10.5194/hess-21-3427-2017, 2017.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, Journal of Hydrology, 270, 105–134, https://doi.org/10.1016/S0022-1694(02)00283-4, 2003.

Felfelani, F., Lawrence, D. M., and Pokhrel, Y.: Representing Intercell Lateral Groundwater Flow and Aquifer Pumping in the Community Land Model, Water Resources Research, 57, e2020WR027531, https://doi.org/10.1029/2020WR027531, 2021.

Furusho-Percot, C., Goergen, K., Hartick, C., Kulkarni, K., Keune, J., and Kollet, S.: Pan-European groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation, Sci Data, 6, 320, https://doi.org/10.1038/s41597-019-0328-7, 2019.

de Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model, Hydrology and Earth System Sciences, 19, 823–837, https://doi.org/10.5194/hess-19-823-2015, 2015.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, Journal of Hydrometeorology, 13, 604–620, https://doi.org/10.1175/JHM-D-11-083.1, 2012.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, Journal of Hydrometeorology, 12, 869–884, https://doi.org/10.1175/2011JHM1324.1, 2011.

Hartick, C., Furusho-Percot, C., Goergen, K., and Kollet, S.: An Interannual Probabilistic Assessment of Subsurface Water Storage Over Europe Using a Fully Coupled Terrestrial Model, Water Res, 57, https://doi.org/10.1029/2020WR027828, 2021.

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydrol. Earth Syst. Sci., 12, 841–861, https://doi.org/10.5194/hess-12-841-2008, 2008.

Ji, P., Yuan, X., and Liang, X.-Z.: Do Lateral Flows Matter for the Hyperresolution Land Surface Modeling?, Journal of Geophysical Research: Atmospheres, 122, 12,077-12,092, https://doi.org/10.1002/2017JD027366, 2017.

Keune, J., Gasper, F., Goergen, K., Hense, A., Shrestha, P., Sulis, M., and Kollet, S.: Studying the influence of groundwater representations on land surface-atmosphere feedbacks during the European heat wave in 2003, Journal of Geophysical Research: Atmospheres, 121, 2016.

Keune, J., Sulis, M., and Kollet, S. J.: Potential Added Value of Incorporating Human Water Use on the Simulation of Evapotranspiration and Precipitation in a Continental-Scale Bedrock-to-Atmosphere Modeling System: A Validation Study Considering Observational Uncertainty, J. Adv. Model. Earth Syst., 11, 1959–1980, https://doi.org/10.1029/2019MS001657, 2019.

Maxwell, R. M. and Condon, L. E.: Connections between groundwater flow and transpiration partitioning, Science, 353, 377–380, https://doi.org/10.1126/science.aaf7891, 2016.

Maxwell, R. M. and Kollet, S. J.: Interdependence of groundwater dynamics and land-energy feedbacks under climate change, Nature Geoscience, 1, 665, 2008.

Miguez-Macho, G., Fan, Y., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, J. Geophys. Res., 112, 2006JD008112, https://doi.org/10.1029/2006JD008112, 2007.

Naz, B. S., Kollet, S., Franssen, H.-J. H., Montzka, C., and Kurtz, W.: A 3 km spatially and temporally consistent European daily soil moisture reanalysis from 2000 to 2015, Sci Data, 7, 111, https://doi.org/10.1038/s41597-020-0450-6, 2020.

O'Neill, M. M. F., Tijerina, D. T., Condon, L. E., and Maxwell, R. M.: Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous United States, Geosci. Model Dev., 14, 7223–7254, https://doi.org/10.5194/gmd-14-7223-2021, 2021.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, Journal of Hydrometeorology, 17, 287–307, https://doi.org/10.1175/JHM-D-15-0054.1, 2016.

Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A.: A Brief Description of the Simple Biosphere Model (SiB), in: Physically-Based Modelling and Simulation of Climate and Climatic Change: Part 1, edited by: Schlesinger, M. E., Springer Netherlands, Dordrecht, 307–330, https://doi.org/10.1007/978-94-009-3041-4_7, 1988.

Shrestha, P., Sulis, M., Simmer, C., and Kollet, S.: Impacts of grid resolution on surface energy fluxes simulated with an integrated surface-groundwater flow model, Hydrol. Earth Syst. Sci., 19, 4317–4326, https://doi.org/10.5194/hess-19-4317-2015, 2015.

Shrestha, P., Sulis, M., Simmer, C., and Kollet, S.: Effects of horizontal grid resolution on evapotranspiration partitioning using TerrSysMP, Journal of Hydrology, 557, 910–915, https://doi.org/10.1016/j.jhydrol.2018.01.024, 2018.

Wada, Y., de Graaf, I. E. M., and van Beek, L. P. H.: High-resolution modeling of human and climate impacts on global water resources, Journal of Advances in Modeling Earth Systems, 8, 735–763, https://doi.org/10.1002/2015MS000618, 2016.

Wood, E. F., Roundy, J. K., Troy, T. J., Beek, L. P. H. van, Bierkens, M. F. P., Blyth, E., Roo, A. de, Döll, P., Ek, M., Famiglietti, J., Gochis, D., Giesen, N. van de, Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resources Research, 47, https://doi.org/10.1029/2010WR010090, n.d.

Xie, Z., Di, Z., Luo, Z., and Ma, Q.: A Quasi-Three-Dimensional Variably Saturated Groundwater Flow Model for Climate Modeling, Journal of Hydrometeorology, 13, 27–46, https://doi.org/10.1175/JHM-D-10-05019.1, 2012.

Zeng, Y., Xie, Z., Liu, S., Xie, J., Jia, B., Qin, P., and Gao, J.: Global Land Surface Modeling Including Lateral Groundwater Flow, Journal of Advances in Modeling Earth Systems, 10, 1882–1900, https://doi.org/10.1029/2018MS001304, 2018.

Zink, M., Kumar, R., Cuntz, M., and Samaniego, L.: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, Hydrology and Earth System Sciences, 21, 1769–1790, https://doi.org/10.5194/hess-21-1769-2017, 2017.