Anonymous Referee #3, 05 Aug 2022

In their manuscript Naz et al. evaluate a pan-European, high-resolution (0.0275°) simulation with the coupled land surface groundwater model ParFlow-CLM, using observations and re-analysis data for streamflow, near-surface soil moisture, evapotranspiration, water table depth and snow water equivalent. In general, the manuscript is well written, the metrics for evaluation seem to be appropriate and the authors go into great detail discussing the potential sources for some of the biases – with respect to possible shortcomings of the model but also of the observational data.

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.

Having said that, there was one aspect of the evaluation that did not fully convince me, namely the evaluation of the simulated water table depths, where only the anomalies were being investigated. I understand that it may not be easy to define the reference elevation, but with the sophisticated ground water fluxes being the key component of the model that sets ParFlow-CLM apart from most LSMs, the authors should really think about discussing a comparison of the absolute values – maybe indicating the uncertainty due to the reference surface elevation. Also I did not understand, why the authors limited their comparison to those points with simulated WTD < 10m ?

This concern was also raised by one of the other reviewers which has prompted us to further clarifications in our revised manuscript. Reported water table depth data across Europe is only poorly quality controlled, and inconsistent methods and standards are used for the calculation of the depth (Fan et al., 2013). Because of these inconsistencies in reporting water table depth data, we compare the anomalies. For example, groundwater levels (meter above sea level) data was provided for most groundwater monitoring wells (i.e., 2018 grid cells out of 2738 located mostly in Germany) but no reference surface elevation information was given. This makes it difficult to convert groundwater levels to WTD or to calculate modeled groundwater levels for direct comparison of absolute values. We complied however with the reviewer's suggestion, to extend our analysis to show the difference in WTD absolute values for the remaining 720 grid cells where WTD data was provided. For these locations, the difference in the observed and simulated WTD is shown in Fig. 1. For these grid cells, we found a good agreement between the ParFlow-CLM and observed WTD with 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.

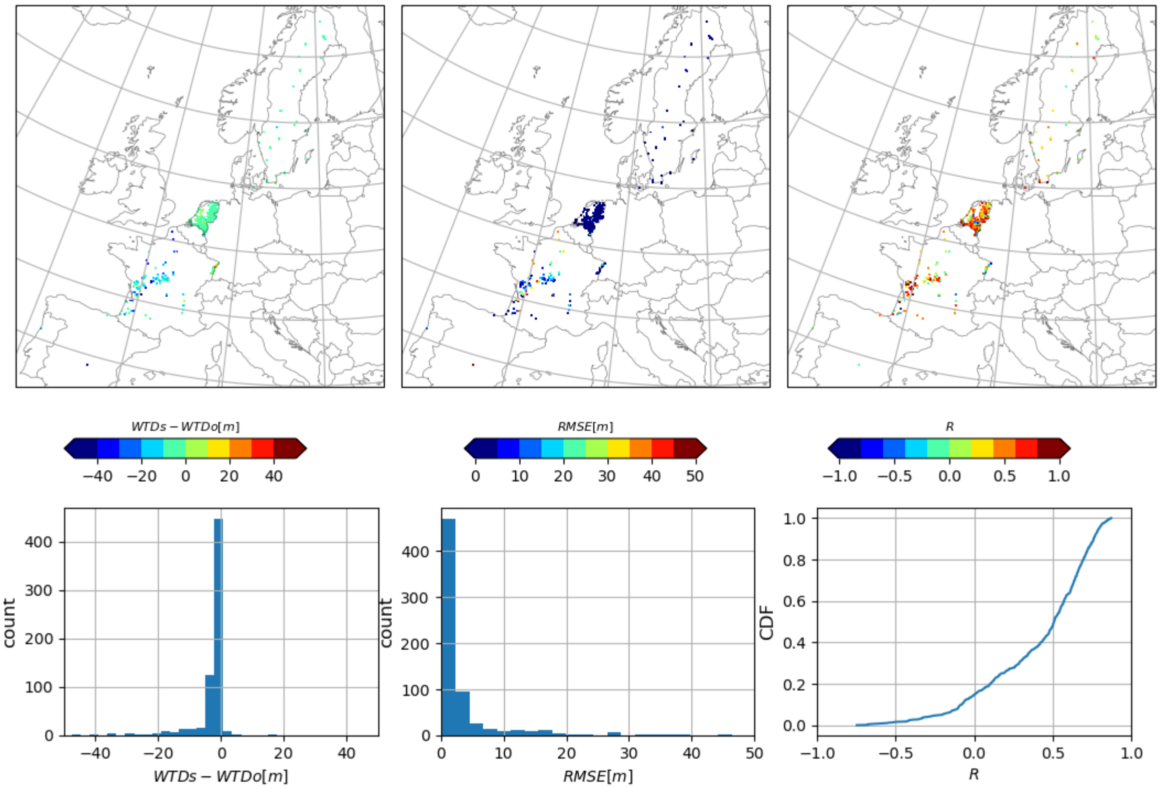

Figure 1: (a) Difference in observed and ParFlow-simulated WTD at filtered locations (N = 720), and (c) RMSE values at filtered locations, (c) Spearman correlation (R) values at selected locations. Histogram plots show the distribution of (d) simulated minus observed WTD and (e) RMSE values. (f) Cumulative distribution function (CDF) of Spearman correlation of ParFlow-CLM with observed WTD monthly data.

In addition to this analysis, we included comparison of total water storage (TWS) simulated by ParFlow-CLM with GRACE satellite data for the time period of 2003-2006 as shown in the following Fig. 2.

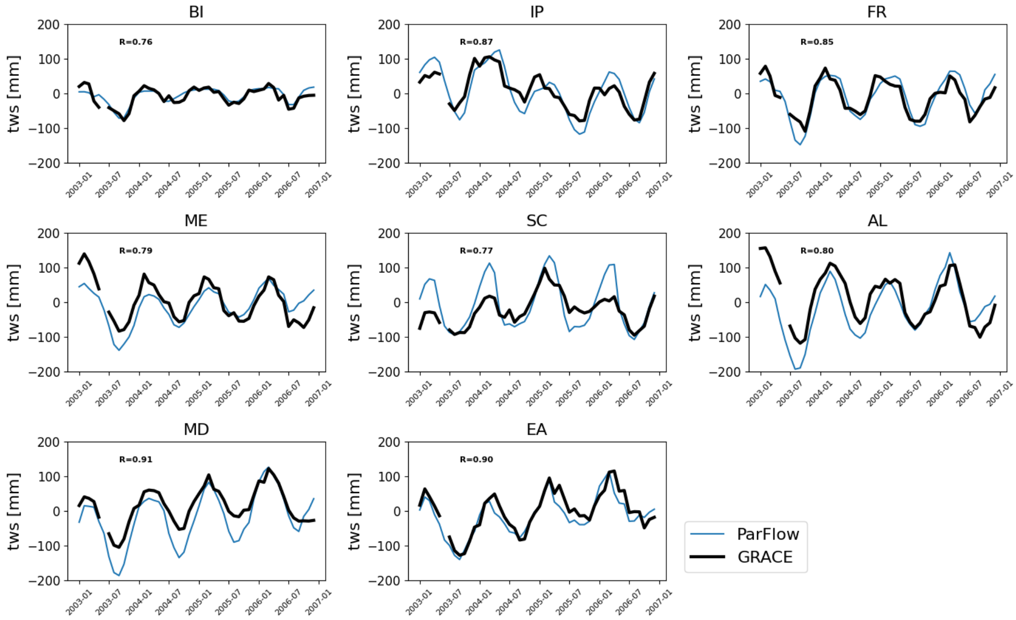

Figure 2: Time series of total water storage anomalies simulated by ParFlow-CLM and its comparison with GRACE products across major regions in the EU-CORDEX domain.

However, my main concern is that I found it somewhat difficult to connect the results to the motivation outlined in the (very well written) introduction of the paper. A large part of the latter is focused on the shortcomings of LSMs and GHMs and their -- admittedly extremely simple – representation of (subsurface) processes. So I would have welcomed a comparison between ParFlow-CLM and a CLM version without ParFlow – possibly the one that is part of ParFlow-CLM -- or with a LSM that includes some simple parametrization of ground water flow (e.g. CLM5 [Felfelani et al., 2021]). Furthermore, the authors indicate that the resolution of the model is important, which I am very willing to believe. Yet they do not show how this affects the simulations in case of their model. Here, a convincing case may have been made by comparing their simulation to the 12km runs in Shastra et al. (2021). If the authors do not want to include an inter-model/-resolution comparison maybe they could think about a different approach to the paper: E.g. as an alternative, the authors could have referred to the study of O'Neil et al. (2021) from the beginning and then set up the paper as a comparison of ParFlow-CLM simulations of Europe and of the CONUS region?

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. In addition, previously published 12 km version of ParFlow-CLM (e.g. Keune et al., 2016; Keune et al.,2018; Furusho-Percot et al., 2019; Hartick et al., 2021) which has been employed over the European CORDEX domain 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, however, the model performance was not rigorously evaluated for all water balance components. In the current study, we employed the ParFlow-CLM model at 3 km resolution over the same domain driven with offline forcings from COSMO-REA6 dataset. Because of two different European modeling setups (i.e. 3 km stand alone ParFlow-CLM and 12 km fully-coupled COSMO-CLM-ParFlow), it is difficult to identify sources of uncertainties that are only caused by model resolution.

As suggested by the reviewer, we summarized the comparison between the European model setup in the current study with CONUS implementations by O'Neill et al. (2021) in the Table 1 as shown below and discussed the main differences in performance between the two models. throughout the manuscript.

Finally, it is not always easy to estimate whether the model really captures a given variable well or not. E.g the authors state that the model "appropriately captures the seasonal cycles" of the WTD (l. 419). However, with only 20% of the investigated cells exhibiting an R > 0.5, it is debatable whether or not this is appropriate. Again, it would have been much more straightforward if the simulation had been compared to a different model / resolution and the question would have simply been about better or worse than XYZ. Without such a comparison, I am not sure that all of the claims made by the authors – e.g. "the added value of capturing heterogeneities for improved water and energy flux simulations in physically-based fully distributed hydrologic models over very large model domains" (l. 16 ff) – are substantiated by their results.
We appreciate your constructive suggestions. As we explained above, we extended the WTD analysis by comparing with absolute values of WTD and compared our results extensively with the CONUS implementation of ParFlow-CLM model (O'Neill et al., 2021) as shown in Table 1 below. Please note that 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 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 implementations 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

Additional comments:

l. 144) (Annoying detail, but) I think that here CLM refers to Community Land Model, while CLM was defined in l. 121 for the predecessor Common Land Model.
Sorry for the confusion. In the revised manuscript, we defined Community Land Model (v3.5) as CLM3.5 and Common Land Model as CLM.
l. 171) Why do you loop a single year to force the model? Doesn't that include the risk of running the model to a non-representative equilibrium state? Also, how did you decide that a 9-year spin-up is enough and how were the states initialized, that a 9-year spin-up is sufficient?
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 spin up over several years owing to the absence of lateral flow and parameterisation 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 year 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.

l. 218) What specific data was assimilated?
The daily SM data at 0.25° resolution from the European Space Agency Climate Change Initiative (ESACCI) were assimilated into CLM3.5 model using an ensemble Kalman filter (EnKF) data assimilation method to produce the 3 km European SSM reanalysis (ESSMRA) dataset. More details on the data assimilation are given in Naz et al., 2019, 2020. In the current simulations, data assimilation was not used.

l. 226) I think it could also be really interesting to compare SM profiles at the stations in addition to the top layer SM.
We agree with the reviewer's comment, however, for the simulation time period (1997 – 2006), soil moisture data were only available for a limited number of stations (19 grids cells). For most stations, the data is available after 2007. More details about the data used in this paper can be found in Naz et al., (2020). Therefore, we refrained from the suggested comparison.

l. 269 ff & Fig1) As you indicate a strong dependency on topography, could you maybe include a plot of the topography in Fig. 1. Also, why is the SWC so low and the WTD so high right next to the river?

We revised Fig. 1 in the manuscript to include topography information as shown below.

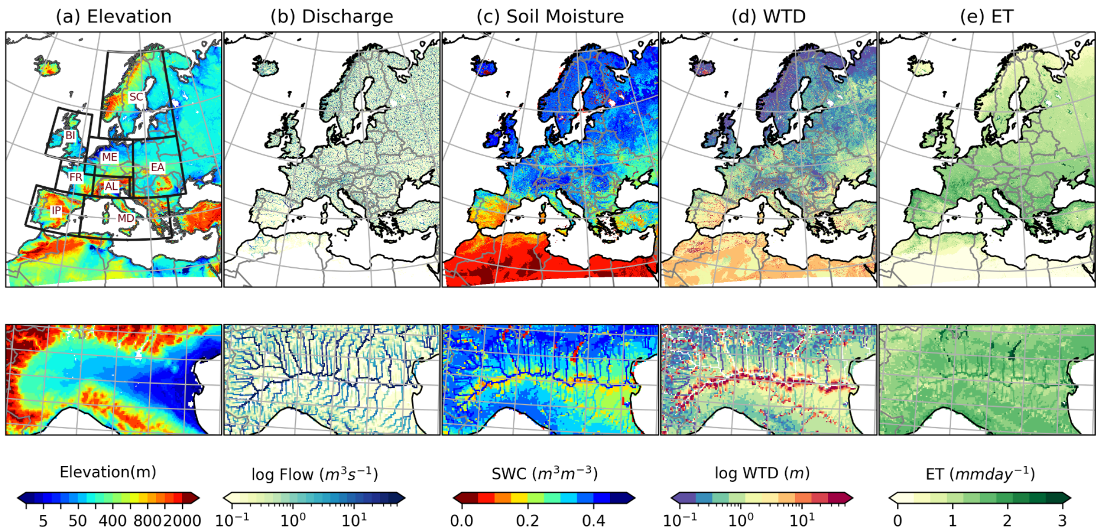

Figure 1: (a) Maps of the EU-CORDEX domain at 3 km resolution (1544 x 1592 grid cells) showing the spatially average distribution of (a) Elevation (b) discharge, (c) surface soil moisture, (d) water table depth, and (e) evapotranspiration (1997–2006) and close-up of Po river basin in Alpine (AL) region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

The deeper WT near the large rivers is probably due to the fact that large rivers were burned into the digital elevation model data in order to hydrologically correct the topographic slopes and ensure European river network connectivity. Burning of rivers appears to make the valleys steeper, resulting in a deeper WTD near the rivers. We have made this point in the manuscript, describing that this was a limitation of the current model setup implementation, that owing to the coarse resolution of the digital elevation model (DEM) (3km), topographic highs were smoothed and in order to get accurate river connectivity we needed to “burn” or imprint the rivers or rather river corridors into the DEM. This limitation is acknowledged in the discussion section along with recommendations for improvement.

l. 289 f.) In case of the Rhine (gauges 2-5) the model appears to underestimate the discharge quite a bit, would this still be explainable by human impacts? Or could it not point to an underestimation of P-ET?

As explained in the manuscript, the underprediction might be related to a (still too) coarse river channel resolution in the model, human impacts on discharge regimes – particularly for highly regulated rivers through reservoir regulations, and power generation or groundwater extraction (e.g., in the case of the Rhine, Elbe and Danube rivers). A 3 km grid cell size might still be too coarse to represent realistic stream networks of smaller rivers and convergence zones along river corridors. In addition, ParFlow-CLM allows for a two-way overland flow routing potentially causing more water losses under dry conditions from channels to groundwater or overbank flow. This may lead to a complete drying of some rivers during summer, further exacerbated by the (comparatively) coarse resolution of the model.

l. 290) I am not sure that everyone is so familiar with the KGE as to immediately know what the range of values indicates. Could you maybe add a very brief explanation here?
This has been explained in the revised manuscript.
Fig2.) I found it a bit hard to identify the gauges in subplot a, do you think it would be possible to zoom in over the center of the first subplot?
We appreciate the suggestions. In the revised manuscript, the figure has been modified to zoom in over the center of the map.
l. 298) I think something went wrong referencing the figure.
It has been corrected in the revised manuscript.
Fig3) Could you clarify that the color-code in panel c is the same as in b?
It has been clarified in the revised manuscript.

l. 339 ff) How can you be sure that the differences are a result of the different treatment of the lateral groundwater flow? I thought that between CLM3.0 and 3.5 there were also major changes in the terrestrial hydrology – e.g. a TOPMODEL approach to runoff generation and changes to the evaporation calculation?
CLM3.5 applies a simple approach to simulate groundwater recharge and discharge processes via the connection of bottom soil layer and an unconfined aquifer as described by Oleson et al., 2008 and Niu et al., 2007, without accounting for lateral groundwater flow. On the other hand, ParFlow-CLM is an integrated coupled surface water-groundwater model which solves the three-dimensional Richards equation to account for variably saturated soil and lateral surface and subsurface flow movements.
To best address this comment, we compared, as an example, the spatial variability of surface soil moisture simulated by ParFlow-CLM for January and August months, 2000 for two regions (Alpine and Mid-Europe) with the ESSMRA dataset (Naz et al., 2020), which is the assimilated soil moisture simulated by CLM3.5, in order 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 the evolving river network and topography. Please note that both models use identical surface information (topography, soil, and vegetation) and atmospheric 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.

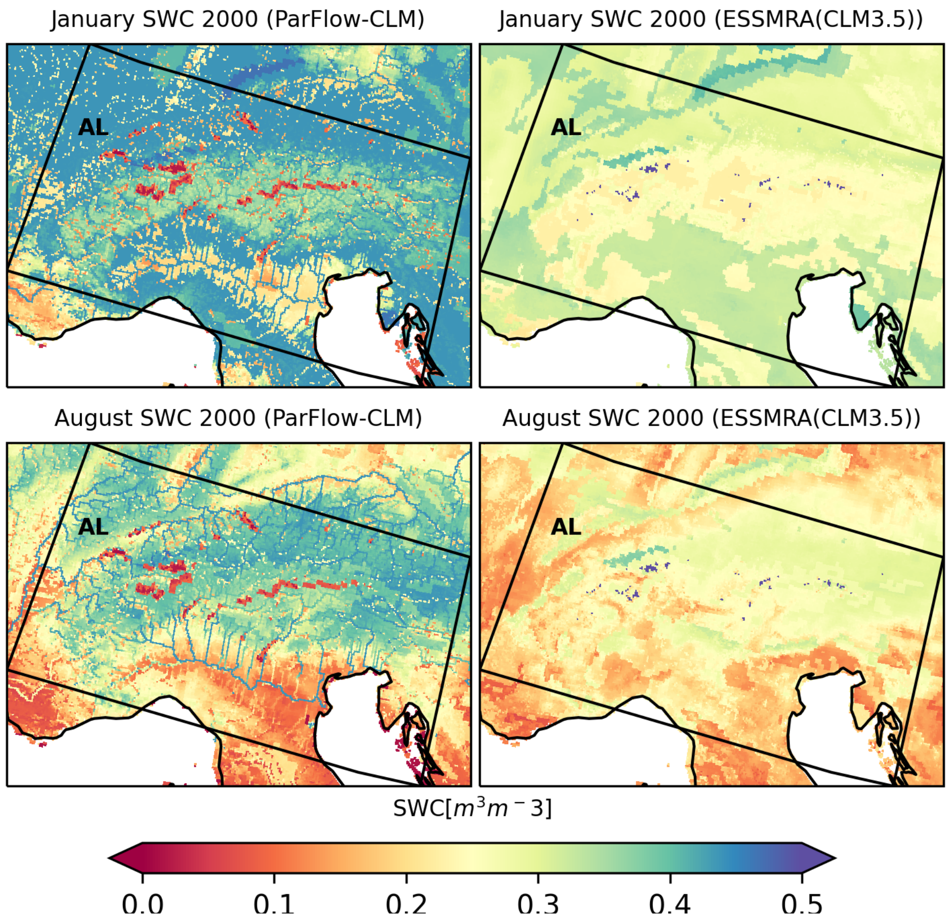

Figure S1. Spatial variability of the surface soil water content (SWC) simulated by ParFlow-CLM and CLM3.5 at the surface soil layer for January and August months of year 2000 over the Alpine region. Please note that glacier areas were not simulated by ParFlow-CLM and soil moisture values are zero at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

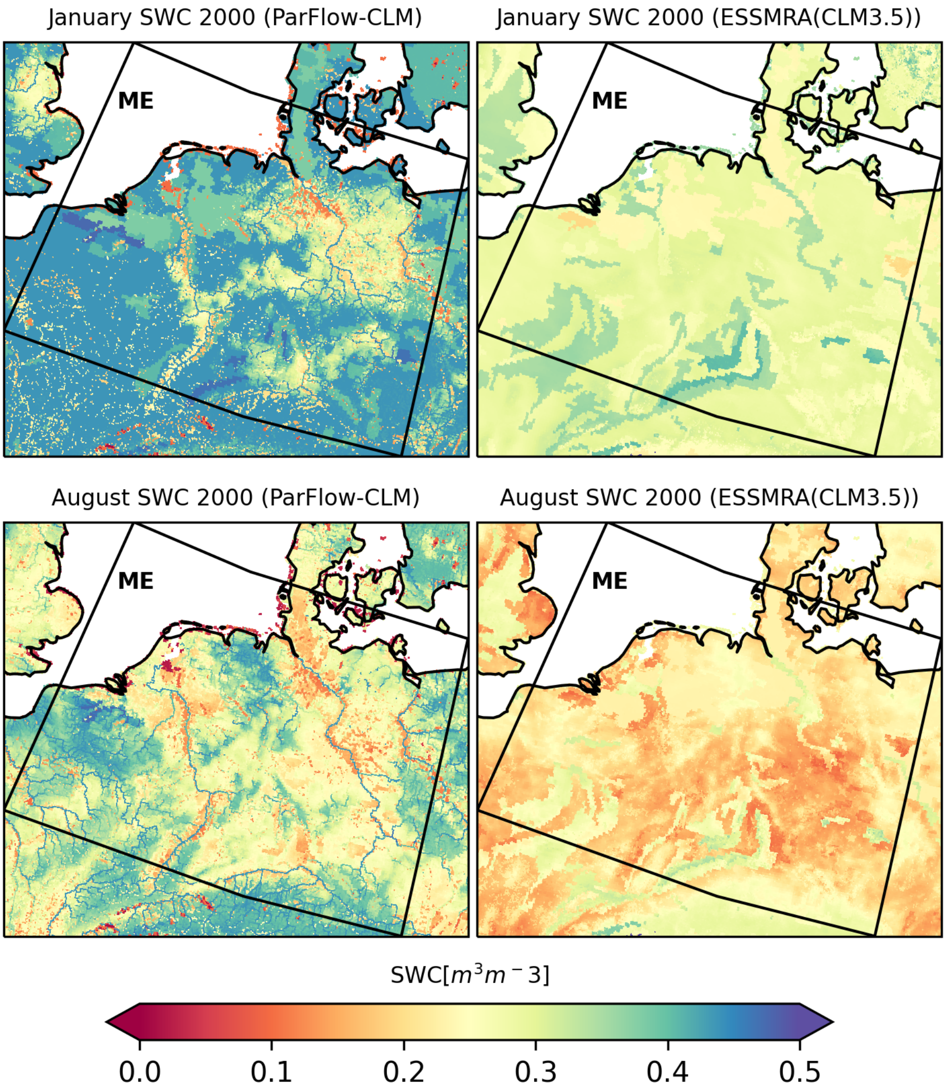

Figure S2. Spatial variability of the soil moisture simulated by ParFlow-CLM and CLM3.5 at the surface soil layer for January and August months of year 2000 over the Mid-Europe region. Please note that glacier areas were not simulated by ParFlow-CLM and soil moisture was set to zero.

l. 352) Not Fig. 4c?
It has been corrected in the revised manuscript.

l. 353) The R values in subplot 4c go beyond this range.
We appreciate the reviewer's comment. It has been corrected in the revised manuscript.
Fig 4.) When comparing ESACCI and ESSMRA in subplot b, these seem to agree much better than ParFlow-CLM agrees with any of the two datasets. As ESSMRA is the closest to a second model that is shown in the study, one could come to the conclusion that the added complexity of the explicit treatment of groundwater fluxes in PArFlow-CLM does very little to improve the near surface soil moisture. Thus, it would be very helpful if the authors could describe in more detail what was assimilated in ESSMRA, because if it was soil moisture directly then the good agreement between ESACCI and ESSMRA is not very surprising. Otherwise it would be very interesting to understand why the ESSMRA appears to be so much closer to ESACCI.

Thanks for pointing this out. Surface soil moisture from the ESA CCI dataset was assimilated into the CLM3.5 model to generate the ESSMRA dataset as described in details by Naz et al., 2020 which is why both ESSMRA and ESACCI are very similar. We used ESSMRA dataset to compare with ParFlow-CLM because both models use identical surface information (topography, soil and vegetation) and forcing datasets and any differences in SM are results of different treatment of groundwater processes. As explained in the previous comment, that despite the assimilation of CCI, CLM3.5 simulates much larger spatial patterns of SM which are mostly related to the soil properties (e.g. soil texture information), while ParFlow-CLM simulates more spatial variability which can be attributed to the effects of river network and topography.

l. 387) Could this overestimation of ET also be a reason for the underestimation of streamflow in the Rhine?
As mentioned above to your earlier comment, and explained in the manuscript we think that the underprediction of streamflow might be related to the following: a (still too) coarse river channel resolution in the model, human impacts on discharge regimes – particularly for highly regulated rivers through reservoir regulations, and power generation or groundwater extraction (e.g., in the case of the Rhine, Elbe and Danube rivers).

l. 417) I think something went wrong referencing the figure.
Thanks for pointing this out. It has been corrected in the revised manuscript.
l. 419) Here I was a bit surprised at the comparatively low R values. Given that precipitation is prescribed based on observations and that both streamflow and ET show a much better correlation with the observations, does this indicate that the model is missing something important in the representation of the groundwater dynamics?
We believe that low values of R for WTD evaluation might be related to uncertainties in aquifer parameterization used in the ParFlow-CLM or the limitations in model resolution such that local aquifers in areas with complex topography cannot be captured. Additionally, model evaluation can be hampered by the challenges associated with groundwater monitoring. For example, the observations might be biased if they are located towards rivers, in low elevations, in areas with confined or perched aquifer systems or in coastal areas. In addition, the comparison of the resolved simulated pressure head, averaged across 3 km, with the point scale observation pressure head, which is highly governed by local surface elevation, can bring about misleading results and amplify inaccuracies. Water table depth observations can also be impacted by pumping which may not be known for many locations.

However, to address the reviewer's concern, we now compared the total water storage (TWS) anomalies simulated by ParFlow with GRACE satellite data as shown in Fig. 3 (shown in response to earlier comment) which shows a good agreement with GRACE data with R values ranging from 0.76 and 0.91 for major regions.

References:
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.
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.
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.

Naz, B. S., Kurtz, W., Montzka, C., Sharples, W., Goergen, K., Keune, J., Gao, H., Springer, A., Hendricks Franssen, H.-J., and Kollet, S.: Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation, Hydrology and Earth System Sciences, 23, 277–301, https://doi.org/10.5194/hess-23-277-2019, 2019.

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.

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, Journal of Geophysical Research: Atmospheres, 112, 2007.

Oleson, K. W., Niu, G.-Y., Yang, Z.-L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., Stöckli, R., Dickinson, R. E., Bonan, G. B., and Levis, S.: Improvements to the Community Land Model and their impact on the hydrological cycle, Journal of Geophysical Research: Biogeosciences, 113, 2008.

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.

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.