Anonymous Referee #1, published 24 Jul 2022

The model evaluation paper of Naz et al. describes a version of the established ParFlow-CLM model applied over Europe and evaluated its hydrological components.

ParFlow-CLM is an established modeling tool, and a publication of a model evaluation paper that builds a foundation for future scientific use is certainly something I would like to support. Unfortunately, in the current stage, the manuscript does not deliver on this goal and seems to purposely hide model shortcomings. In the current version, I can only suggest significant revisions.
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 certainly wrote the manuscript with that purpose in mind - to provide a foundation for future scientific use, particularly for use in:
1. Studies on the impact of climate change on water resources
2. Coupled Earth system model simulations

Both cases above require large scale integrated hydrology to capture macro scale groundwater dynamics, groundwater-surface water interactions (Condon et al., 2021). Therefore we have strengthened the Introduction section to emphasize the need for large scale hydrological modeling in these use cases and what the trade offs could be when comparing catchment scale versus continental scale model implementations.

Parflow-CLM (i.e., ParFlow hydrologic model coupled here to the Common Land Model) is an established modeling tool, however continental scale modeling at high resolution (<5km) is challenging both computationally (3D finite volume implementation) and also in terms of data sparsity in regions (e.g. geological information, soil classes). There are few studies which have implemented a high resolution, fully three dimensional coupled land surface - groundwater model continentally, most notably there have been similar approaches over the CONUS domain (Maxwell et al., 2015). We have strengthened this point in the Introduction and Discussion sections.

I would like to focus on two aspects that are currently flawed. Firstly, the paper's motivation could be much clearer from the beginning. In the light of the many publications that already exist on ParFlow and CLM, what is the added value of this model evaluation paper? What is the model's purpose within the range of continental and global models? What questions can it help to answer? Outlining this much more clearly from the beginning will be helpful for the scientific community in making this publication a helpful reference for future research.
We appreciate your constructive suggestions. In the revised manuscript, we have made the objectives and research goals clear by expanding 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 that simultaneously solves surface and subsurface processes with lateral-groundwater flow. The lateral groundwater flow is a key model feature - many modeling systems implemented at continental or global scales are one dimensional and contain a parameterised 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 (< 5 km), physically-based integrated hydrological models can better represent groundwater-surface water interactions, and heterogeneities in the representation of the water and energy cycles, because of the higher resolution surface data used. In addition, owing to ParFlow's 3D flow implementation, this model setup provides a more accurate representation of lateral transport of surface and subsurface water movements driven by topographic slopes (Bierkens et al., 2015). While we agree that several studies exist on ParFlow-CLM, they mostly concentrate on the CONUS region. As the CONUS domain consists of a single country, the CONUS region has reasonable coverage in terms of observational network and geological information. Unfortunately given the European (EU) model domain consists of many individual countries, observations across regions are not all of the same quality or coverage. We have emphasized this point in the discussion, and highlighted that this could be a contributing factor for poor model performance in some regions of the EU domain.

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 resolution with lateral surface and groundwater flow representation. 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). 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). We have strengthened this point in the Introduction.

2. Explicitly incorporating hydrological processes that are not included in the existing land surface models (LSMs) can also benefit the land surface modeling community for 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 represented in LSMs in a simplified way. 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. Maxwell and Kollet, 2008; 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 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.

Secondly, I cannot accept the current evaluation of the groundwater component. The authors use groundwater in the title and motivate the model's usefulness with the argument of an active groundwater component but provide a not convincing evaluation. I do not expect the model to be able to perfectly represent the water table. Still, I think we can only progress if we are open about our models' shortcomings and clearly communicate uncertainties. Poor model performance is not a reason for not publishing something as long as there is a proper discussion on the causes. Currently, the paper is not doing that and uses oversimplified

evaluation methods to obfuscate the actual model behavior. Furthermore, existing literature and models are omitted as well.
We have addressed reviewer's concerns by conducting additional analysis of WTD evaluation with in-situ observations, including absolute value comparison of WTD using R and PBAIS and RMSE statistical metrics. Additionally, we have evaluated total water storage (TWS) with GRACE satellite data.

To strengthen our model evaluation as a whole, we included detailed comparison of our results with the PfCONUSv1 implementation described by O'Neill et al. (2021) and discussed our results in comparison to other global scale models. Please see more details in response to your comments below.

Additional notes:

* While I know that it is difficult to find a repository to host a large amount of data, I employ the authors to think about if selected model outputs could be made available in the spirit of OpenScience principles!

We agree with your comment and selected model outputs will be made available on our public repository of https://datapub.fz-juelich.de/, including a dataset DOI.

* Is it really necessary to use the overcomplicated PF-CLM-EU3km as a name? Why not stick with ParFlow-CLM in the paper? If it is a very different model, why is that not the name used in the title?
We agree with the reviewer's suggestion. We now replaced "PF-CLM-EU3km" with ParFlow-CLM throughout the manuscript.

L. 1: How are these large-scale models useful for water resource management? I see how they are helpful for large-scale policy and fostering scientific understanding but are they really useful for management? Please also define what high-resolution means in brackets - people have very different interpretations about that, and it is changing fast.

Thanks for this comment. We have clarified this point in the manuscript. While we agree that catchment scale would be most relevant for the purpose of water management, catchment scale models only capture processes contained within the catchment boundary, whereas large-scale simulation at high-resolution (< 5 km) is necessary to understand changes to water resources from macro-scale processes such as high evapotranspiration rates leading to soil moisture deficits, resulting, e.g., in mega droughts over large area (for example, the 2018 to 2020 European drought; Rakovec et., 2022), water storage deficits and flow regime shifts (hydrological droughts; Hanel et al., 2018), and widespread flooding (e.g. Western Europe floods in 2021; He et al., 2022).

In addition, the influence of climate variability and climate non-stationarity can not be modeled at a catchment scale (Massei et al., 2020). In the revised manuscript, we reformulated the text to avoid confusion.

3: How is the coarse spatial resolution linked to the lateral fluxes and groundwater components - isn't that mixing up things? What small scale processes specifically?
We argue that the issue of model resolution and accurate representation of the surface and subsurface processes are interlinked which has been discussed extensively in the literature (Beven et al., 2015; Bierkens et al., 2015; Melsen et al., 2016; Wood et al., 2011; Fan, 2015).

Land surface models often ignore lateral surface and subsurface water movements, also because these fine scale processes cannot be resolved realistically at coarse resolution (e.g. Clark et al., 2015) - we have made this point in the manuscript. On the other hand, 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 because 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) - we have clarified this point in the manuscript.

4: what does more complex refer to? Complex in what regard?
Here complex refers to more complex models such as the integrated land surface hydrological models such as Parflow-CLM (e.g., as defined in Kuffour et al. (2020)) that solve three-dimensional Richard's equation to simulate three-dimensional movement of subsurface water in a continuum approach with two-dimensional overland flow whereas most LSMs are one dimensional and therefore only solve subsurface water movement vertically and ignore surface routing. This is clarified in the revised manuscript.

11: what is PF-CLM-EU3km? It has not been introduced; quantify good agreement
We now replaced “PF-CLM-EU3km” with ParFlow-CLM throughout the manuscript. Originally we tried to distinguish our specific implementation from others, such as PfCONUSv1.

17: this is the first-time heterogeneities are mentioned. Is it implied that this is a result of the higher spatial resolution? This should be explained
Thanks for this comment - we overlooked this explanation. We have now explained this in the revised manuscript.

Fig. 1 c) WTD in log scale without indicating what red is. Is that deeper than 100 m? How deep is it? Why is the WTD so deep near larger rivers? Why so shallow in mountainous regions? What is the reasoning here why this is plausible? Is it plausible in the light of the performance of other large-scale models?
Thanks for pointing this out. Red color indicates deeper water table with maximum of 51 m depth. 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 more steep, 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.

415: I get the problem of inconsistent WTD elevation data. Still, this should be solvable for at least some regions in Europe. I feel that the authors feared that the model performance would be judged too harshly. Whatever the reason, the solution shown here is not acceptable. Furthermore, you can't simply select only the cells that simulate WTD < 10!! This is the range almost all models do a good job. This is not advancing our science. This is far from ok.

Thanks for your comment. This 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. See our detailed response below.

We did not deliberately set out to obfuscate the model's shortcomings. Please note that we showed an example of ParFlow-CLM performance in the supplementary material (Figure S11 and S12) with highest and lowest correlation (R) values across different regions to highlight model limitations in different regions.

To address the reviewer's concern about not including all the data, we conducted our analysis using all the available data without any filtering for quality control. However, this has resulted in no significant differences, compared to previous results as shown in the following Fig. 1.

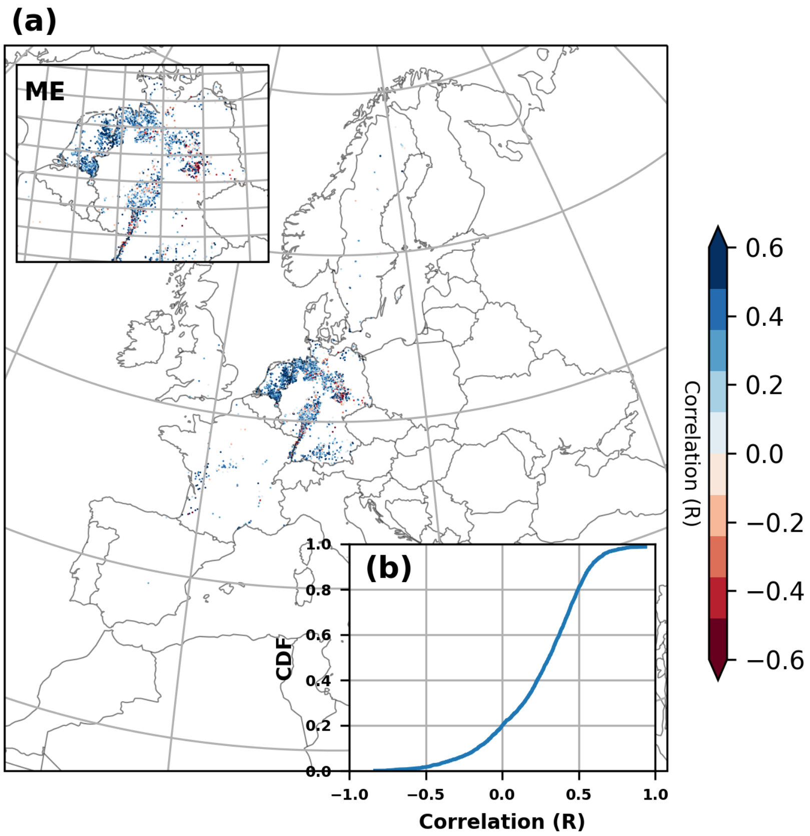

Figure 1: (a) Correlation map between in-situ water table depth (WTD) anomalies and ParFlow-CLM model using all available data (2738 grid cells). (b) Cumulative distribution function (CDF) of correlation coefficient of ParFlow-CLM with observed WTD anomalies. The inset in (a) shows a zoom of the Mid-Europe (ME) region.

We would like to acknowledge that the reviewer's suggestions have led to a more comprehensive analysis and as a result, has strengthened the revised manuscript.

Please show how much the model deviates from observations. You motivate your paper with the statement that representation of groundwater is essential and then skip a proper evaluation of your model.

I suspect it will not perform perfectly - no large-scale model currently can, and you are providing some reasonable answers by referring to Gleeson et al., 2021, which is good but not enough. Please provide a more extensive discussion on how the performance differs from other existing research.

To address this comment, we extended our analysis to make a direct comparison between model and observations for all those locations (720) where WTD data is provided. As explained above, for most locations (i.e. 2018 grid cells), groundwater levels (meter above sea level) data was provided but no reference surface elevation information was given which is needed to convert groundwater levels to WTD or to calculate modeled groundwater levels for direct comparison of absolute values. Therefore, we excluded these locations from this comparison. For the remaining 720 locations, the difference in the observed and simulated WTD is shown in Fig. 2. 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 more 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.

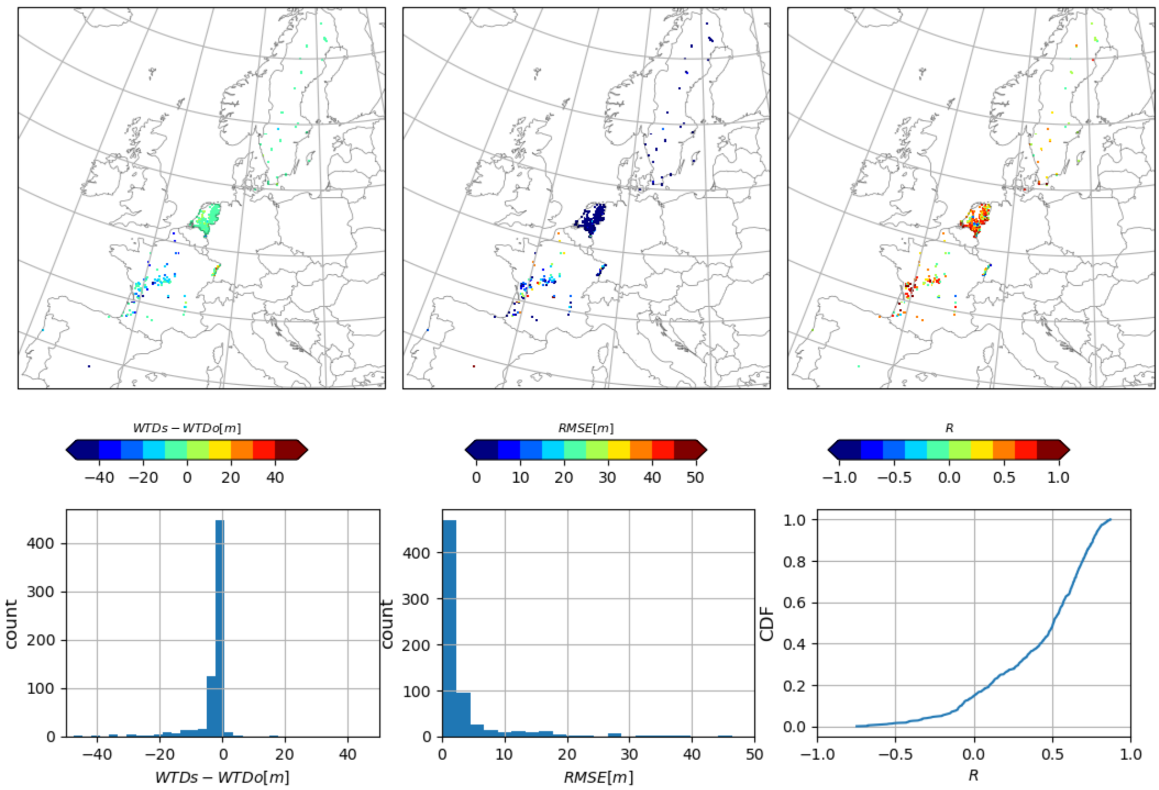

Figure 2: (a) Difference in observed and ParFlow-simulated WTD at filtered locations (N = 720), and (b) RMSE values at filtered locations, (c) Spearman correlation (R) values at selected locations. Histogram plots show the distribution of (d) simulated minus observed WTD and (e) RMSE values. (f) Cumulative 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. 3.

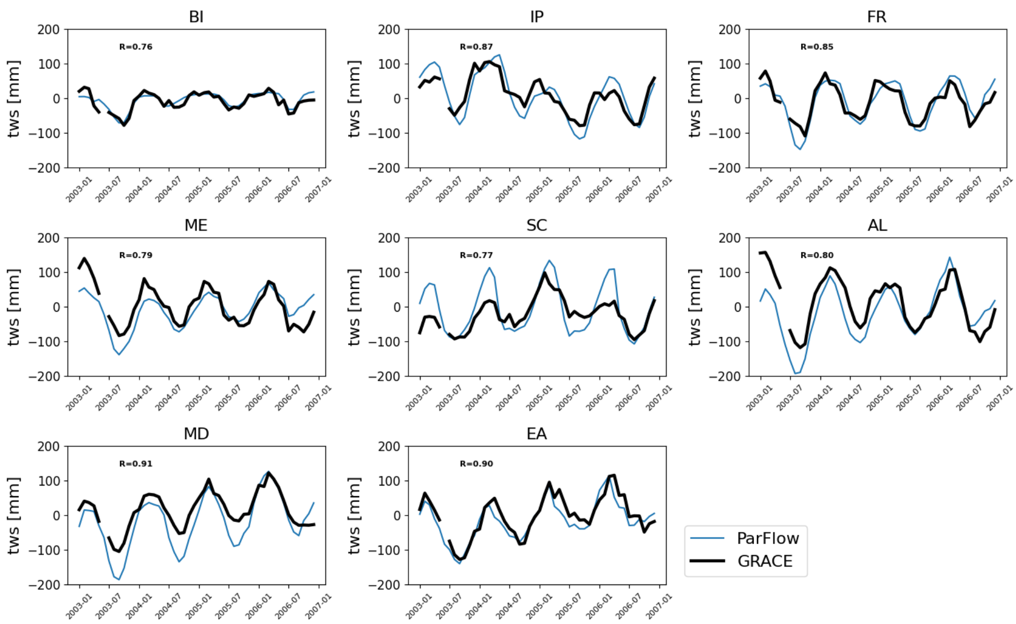

Figure 3: Time series of total water storage anomalies simulated by ParFlow-CLM and its comparison with GRACE products across major regions in the EU-CORDEX domain.

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 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		GRACE dataset (monthly)	ranging from 0.43 to 0.94 for		Both model setups show stronger dry anomalies and overpredict wet anomalies relative to the GRACE data.

		major regions			major basins		
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

417: ?? = Fig. 7

It has been corrected in the revised manuscript.

References:

Beven, K., Cloke, H., Pappenberger, F., Lamb, R., and Hunter, N.: Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface, Sci. China Earth Sci., 58, 25–35, https://doi.org/10.1007/s11430-014-5003-4, 2015.

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.

Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., Fransen, H.-J. H., Verhoef, A., Van Loon, A. F., Sulis, M., and Abesser, C.: Global Groundwater Modeling and Monitoring: Opportunities and Challenges, Water Resources Research, 57, e2020WR029500, https://doi.org/10.1029/2020WR029500, 2021.

Fan, Y.: Groundwater in the Earth's critical zone: Relevance to large-scale patterns and processes, Water Resources Research, 51, 3052–3069, https://doi.org/10.1002/2015WR017037, 2015.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science, 339, 940–943, 2013.

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.

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.

Hanel, M., Rakovec, O., Markonis, Y., Máca, P., Samaniego, L., Kyselý, J., and Kumar, R.: Revisiting the recent European droughts from a long-term perspective, Sci Rep, 8, 9499, https://doi.org/10.1038/s41598-018-27464-4, 2018.

He, K., Yang, Q., Shen, X., and Anagnostou, E. N.: Brief communication: Western Europe flood in 2021 – mapping agriculture flood exposure from synthetic aperture radar (SAR), Nat. Hazards Earth Syst. Sci., 22, 2921–2927, https://doi.org/10.5194/nhess-22-2921-2022, 2022.

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.

Kuffour, B. N. O., Engdahl, N. B., Woodward, C. S., Condon, L. E., Kollet, S., & Maxwell, R. M. (2020). Simulating coupled surface–subsurface flows with ParFlow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. Geosci. Model Dev., 13(3), 1373-1397. https://gmd.copernicus.org/articles/13/1373/2020/

Massei, N., Kingston, D. G., Hannah, D. M., Vidal, J.-P., Dieppois, B., Fossa, M., Hartmann, A., Lavers, D. A., and Laignel, B.: Understanding and predicting large-scale hydrological variability in a changing environment, Proc. IAHS, 383, 141–149, https://doi.org/10.5194/piahs-383-141-2020, 2020.

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., Condon, L. E., and Kollet, S. J.: A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3, Geoscientific Model Development, 8, 923, 2015.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., and Uijlenhoet, R.: Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin, Hydrol. Earth Syst. Sci., 20, 2207–2226, https://doi.org/10.5194/hess-20-2207-2016, 2016.

Mertes, L. A. K.: Documentation and significance of the perirheic zone on inundated floodplains, Water Resources Research, 33, 1749–1762, https://doi.org/10.1029/97WR00658, 1997.

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.

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.

Rakovec, O., Samaniego, L., Hari, V., Markonis, Y., Moravec, V., Thober, S., Hanel, M., and Kumar, R.: The 2018–2020 Multi-Year Drought Sets a New Benchmark in Europe, Earth's Future, 10, https://doi.org/10.1029/2021EF002394, 2022.

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.

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.