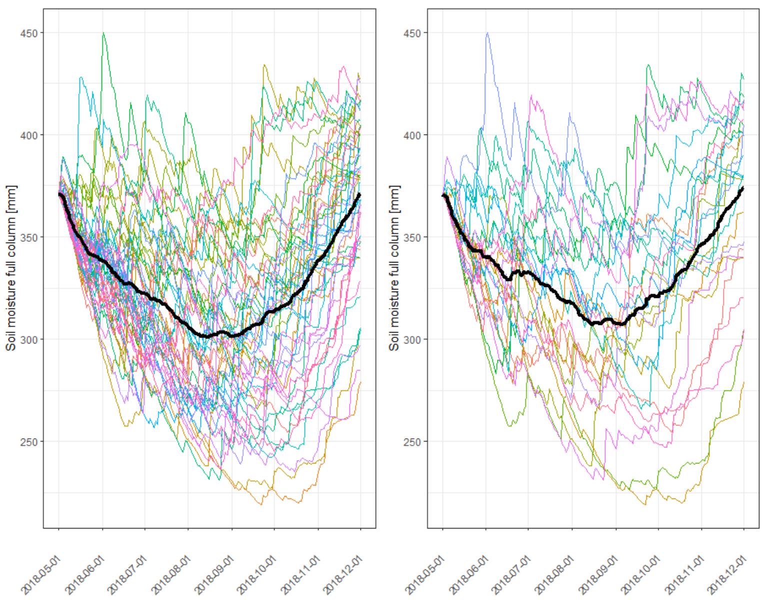
Editor review answers.

SI Editor 1	
But I agree with R2, the science is not novel. From a predictability point of view, it is well known that skill is higher over months 1-3 and decreases with lead. It feels like the scientific questions are somewhat token, and are just included to showcase the tool. If the paper is looking just to demonstrate a new tool, that is fine. Certainly, being more computationally efficient over alternatives is a reasonable justification. But maybe NHESS is not really the right journal for this - GMD maybe better.	Partly agreed. The fact about the decrease of forecast skill over time and 1-3 months as the best skill is known indeed, but not for local scale and drought conditions. Therefore, we believe that our findings are novel in a way that we confirm the abovementioned fact for the stated conditions as well. We elaborated on the research questions.
There is one potential issue with the setup of the system. The authors are running with daily SEAS5 and comparing with runs using hourly ERA5? Then also comparing with 3h SMAP. Maybe the authors are interpolating all to daily resolution first and didn't explain. But if the authors aren't, then this seems to be a flaw in the methodology. There are some parts of the world where sub-daily dynamics have a huge impact on water balance (e.g. https://hess.copernicus.org/preprints/hess-2021-48/). Granted, these are generally semi-arid regions with short-duration rainfall events and high PET. But still, it seems like inadequate scientific design in principle to compare things that have been run at different temporal	We rephrased the setup description for clarification. The BROOK90 itself runs with a variable number of iterations per day, which is automatically determined and dictated by the equilibrium of the water balance equation inside the system. The minimum iteration number is two, or in case of precipitation event, it equals the resolution of the meteorological data input. However, under some conditions (i.e. heavy rain, drought stress, complex soil profile) this iterations per day can be as high as 1000, independent from the input data resolution. Nevertheless, the output of the model always has a daily resolution. Indeed, meteorological input data in the study for the BROOK90 model have different resolutions. However, the BROOK90 model allows accounting to create subdaily precipitation (hourly scale) using daily precipitation value and monthly values of mean event duration in hours (see DURATN parameter in the model documentation http://www.ecoshift.net/brook/b90doc.html#). In the study, we calculated this parameter for each catchment separately from the available ERA5 hourly data and applied it to forecast forcing to improve the results. All model results, as well as SMAP data are aggregated and compared on the same time scale (daily-monthly, mentioned in specific sections/figures). Regarding the influence of the subdaily precipitation

resolutions.	dynamics of the quality of water balance estimations - we agree with the statement in general. However, it was found that e.g. for ET component in Germany, ERA5 with hourly resolution shows worse results than the same data on daily scale (https://hess.copernicus.org/articles/26/3177/2022/)
Also the authors declined R2 suggestion to extend further back in time because SEAS5 was launched in 2017, and “possible extension of the modeling time-period to earlier dates will lead to mixing of system versions (e.g. 4th and 5th)”. That isn’t correct. The hindcast is produced with an identical model version as the forecast. It is set up to be entirely consistent. There is an argument about a different number of ensemble members (25 in hindcast vs 51 in the forecast) but as far as I can tell they are just using ensemble mean anyway. I wouldn’t expect much difference in mean when using the 25 v 51. But if there is, the authors could just use the first 25 members of the forecast and have data from 1993-present to do something more rigorous and scientifically interesting. Not to say that the authors have to, just that their rebuttal excuse not to do so is not based on a correct point.	Subsequently, we have found that hindcasts prior 2017 were produced with the SEAS5 model as well (https://www.ecmwf.int/en/elibrary/81237-seas5-user-guide). Yes, in the sections 2.1–2.3 (now 3.1-3.3) we use the ensemble mean (which however does not mean the application of SEAS5 ensemble mean as forcing, but the calculation of the mean from the 51 ensemble BROOK90 runs). However, in section 2.4 (now 3.4) we show results using all ensembles. In both cases, the potential reduction or mixture between 25 vs 51 ensemble members will be noticeable (see the example figure below). Nevertheless, different ensemble size was not the only reason; the main reason was that we wanted to focus exactly on the extreme drought in Europe in 2018-2019.  *Comparison of full ensemble set (51) vs first 25 members forecast for Natzschung catchment and its influence on ensemble mean. Forecast starts 2018-05-01. Ensembles are shown with thick coloured lines, ensemble mean - with thick black line. Soil moisture for the full column is shown, results for REW or topsoil will show an even higher difference.*
Finally, there is a fundamental mismatch between what the authors are selling and what the authors are evaluating The authors are selling the forecast as able to provide	Agreed, in the main sections (2.1,2.3-2.4, now 3.1,3.3-3.4) we evaluated catchment-mean results. Comparison of point soil moisture observations with BROOK90 model results (considering the presented 1.5D setup) is in our opinion pointless for a number of reasons. In another study we have found and discussed that even with the model parameterisation based on detailed vegetation and soil

information for action down to 100m. Yet as far as I can tell all the evaluation is all based on catchment-average. The authors don't test the 100m skill in any way as far as I can see. To do that is difficult - needs lots of point data ideally. But at least with the data, the authors have the authors could do something - comparing the ERA5-driven run with the SEAS5-driven one, using a metric that keeps the spatial dimensions explicit (e.g. skill maps, although the authors need to make a larger soil moisture reforecast to do this).	profile data with hydraulic functions determined in a lab, the model could still deliver significant deviations compared to measurements from moisture sensors (https://doi.org/10.1127/metz/2023/1155). However, we propose to extend section 2.2 (now 3.2) to include information concerning HRU-scale evaluation. As a direct way to depict spatial grid/HRU evaluation like we did in section 2.3 (now 3.3) is not possible, we are adding three extra figures produced from results in Natzschung catchment. One is a graph where soil moisture for all HRUs will be plotted against time for ERA5 and one SEAS forecast to compare differences in prediction for the most dry period (summer 2018). Second is a series of maps with mean monthly soil moisture for 10/2018 produced with ERA5 and SEAS forecasts with 1, 3, 5 and 7 months lead times. Third is a series of maps with KGE values for full soil columns calculated on a daily scale between ERA5 and SEAS forecasts with 1, 3, 5 and 7 months lead times.
Regarding the author's response to the reviewers, I believe that their response is acceptable. The focus of the study is on a single event (2018 European drought) and on the local scale (hence the 12 small catchments). This should be made very clear, otherwise, the authors create too high expectations	Agreed, as we answered before, changes are made to the title and introduction, to narrow the focus of the study and make it more clear to the reader.
SI Editor 2	
They cannot state ECMWF forecasts. ECMWF has different prediction systems and products. Regarding seasonal forecasts, they should refer to SEAS5 (I guess this is what the authors used), and distinguish between forecasts and forecast systems.	Agreed, corrected throughout the text.
The scientific questions are not well presented. Again the focus is on applying a	Agreed, the research questions have been updated.

global water balance model framework for forecasting the 2018 European soil moisture droughts at the local scale. With this in mind, the scientific questions should be better stated.	
The structure of the manuscript is confusing me. There should be a clear Models and Data section and another Methodology. The first should present the Global BROOK90, the meteorological historical data and forecasts used (ERA5 and ECMWF SEAS5), and soil moisture forecasting. The other section should present the methodology for evaluating the soil moisture performance and forecast accuracy.	Agreed, the structure has been changed accordingly.
Methodologically the authors compare the model performance using SMAP data as reference. This would allow to indicate the model structural and parameterization limitations. Then they compare the ECMWF-based forecast mean to ERA5-based simulation. This means that the evaluation is done in a pseudo-reality and hence the model structural and parameterization limitations are omitted. Hence what did we learn from the two evaluations; this is not well communicated by the authors.	SMAP itself is not a purely observation-based product, as it uses satellite estimations of surface temperature and land model, thus it is not canonical model validation, but rather a comparison with the best available global open-source soil moisture data often used in literature as a reference. Therefore, we would not state that mismatch of SMAP and GBR90 soil moisture indicates structural/parameterization problems of our setup. The major shortcomings of the setup are already mentioned and discussed in https://doi.org/10.2166/nh.2021.150 and https://doi.org/10.5194/hess-26-3177-2022, therefore we do not want to repeat these statements once more. By setting ERA5 as a reference to ECMWF SEAS5 forecast we indeed want to focus on meteorological input data uncertainty of the soil moisture forecast. We elaborated on communication of the results in the conclusion.
I was a bit surprised to read in section 2.3 that there was no clear pattern between behavior of forecasted soil moisture and catchment characteristics. The authors	Agreed, inter-comparison of relative and absolute errors could be affected by catchment-specifics of soil moisture. We calculated proposed modifications of absolute error and put them instead of the current figures and adjusted description. Now few patterns considering catchment clustering are visible and noticed, although in our opinion

need to explain this from a process perspective and compare it with insights that conclude the opposite (Sutanto et al., 2022,https://www.nature.com/articles/s41598-022-06553-5).	there is still no clear correlation with catchment characteristics. We could only assume with caution that the gained results (general underestimation in non-drought and overestimation in drought conditions) are more likely to be driven by meteo-data uncertainty (forecast inaccuracy) rather than different responses from catchments.
I do not believe that the performance metric used for forecast accuracy is the right one to allow inter-comparison between the 12 catchments. The authors use the absolute and relative difference which is very specific to the specific soil moisture average conditions. I believe the authors should calculate the absolute error and divide it by the mean soil moisture for that catchment (a type of standardization) and this would allow comparability with the other catchments. I wonder if their conclusion on (5) could be related to this.	
I believe that there should be a discussion on (1) different approaches (pros and cons) for forecasting and the added value of the proposed Global BROOK90 framework, and (2) elaborating more on the application of the framework to other hydro-climatic regimes and ungauged or poorly gauged systems.	Partly agreed. The review of current available tools/methods for drought forecasting and monitoring is presented and the potential niche for Global BROOK90 is identified in the introduction (second and third paragraphs). Regarding the framework applicability, we upgraded the description of Global BROOK90 framework (section 1.1).