

Helmholtz Centre for Environmental Research – UFZ Permoserstr.15 • 04318 Leipzig • Germany

> Contact person: Friedrich Boeing Department Computational Hydrosystems friedrich.boeing@ufz.de

Leipzig, April 2022

Re-submission HESS-2021-402

Dear Dr. Teuling,

we would like to thank you and the reviewers again for the comments which helped to increase the overall quality of the manuscript.

Please find the newly uploaded revised manuscript, incorporating the proposed changes and additions (including the proposed changes and additions to the figures, the discussions and language edits), as well as a marked-up manuscript version (latexdiff file) showing the differences between the originally submitted and the revised manuscript below, as well as the detailed responses to the reviewers' comments. Additionally to the reviewers requested changes, we updated and extended the sections "Code and data availability" (including DOI links) and "Acknowledgments".

Kind regards,

Friedrich Boeing

Helmholtz Centre for Environmental Research – UFZ

Company domicile: Leipzig

Permoserstr. 15, 04318 Leipzig, Germany or PF 500136, 04301 Leipzig, Germany phone +49 341 235-0

info@ufz.de www.ufz.de

Registration court: Leipzig district court Commercial register No. B 4703

Chairman of the Supervisory Board: MinDirig'in Oda Keppler

Scientific Director: Prof. Dr. Georg Teutsch

Administrative Director: Dr. Sabine König

Bank details: HypoVereinsbank Leipzig Sort code 860 200 86 Account No. 5080 186 136 Swift (BIC) code HYVEDEMM495 IBAN No. DE12860200865080186136 VAT No. DE 141 507 065 Tax No. 232/124/00416



Original reviewer comments are in italics, authors' response is in bold.

Referee #1:

SUMMARY:

A recommendation: moderate revision

Using observed SM data collected from 40 locations and in 4 different measurement methods in Germany, the authors evaluate the performance of second generation operational German Drought Monitor in simulating soil moisture (SM). Two major research questions within this paper have been adequately addressed and can be summarized as follows: 1. how well the GDM capture the SM dynamics; 2. will GDM with higher spatial resolution produce SM estimates with higher quality compared with the GDM of former edition. Through the research, it was found that 1. SM dynamics simulations could be moderately improved; 2. higher resolution drought information at the one-kilometer scale can be met.

This research is a report of the improvement in the model performance which is evaluated in the perspective of comparisons between the model simulations and the observations. The article conforms to the journal-specific instructions and is relevant to HESS. The work is appropriate to be published in this journal after some revisions. In the following part, I will state the major arguments in detail.

Authors' response #1: We thank the Reviewer for an overall positive assessments of our work. We paid detailed attention to all comments and have addressed all of them below accordingly.

The comparison between the observations and the model simulations are mainly indicated in the form of Spearman rank correlation coefficient. Nevertheless, the results of the comparison are in fact not so ideal in the perspective of solely the Spearman correlation coefficient, not to mention that significance values have been neglected when the observations and model results are compared. For example, when the observations from all the sites are included, the coefficients are normally lower. In addition, p-value has only been mentioned when both versions of the mHM and observations are compared (table 2). No mentioning of significance values also makes the comparison results not so validated. The significance value should be mentioned in the research to make results more reliable

Authors' response #2: Thank you for this comment. We agree with the reviewer that the significance value is the essential information. Therefore we added the significance values into Figure 4 (see exemplary Figure R1 below) and Figure 5 by indicating locations with p-values < 0.05 and positive correlations.



Figure R1: Same as Figure 4 in the main manuscript, but including information of correlation significance (p-value < 0.05) and positive correlation.

In addition, the analyses are focused on the explanations for the correlation coefficients, few further knowledge regarding the inner uncertainty within the models such as misrepresentations or no involvement of certain natural factors are included to be accounted for the discrepancy between the observations and model results. In this way, the detailed explanations in the conclusion and discussion part of this paper are needed to clarify the differences between observations and model results.

The four kinds of observations seem not to be standardized. In this way, some values of correlation coefficient which concern a specific observation technique, are not guaranteed to be valid in indicating the performance of GDM in simulating SM dynamics. A standardization method should be carried out to ensure a consistent comparison between observations and model simulations.

Authors' response #3: We thank the reviewer for the suggestions and agree that aspects regarding soil moisture measurements need to be considered carefully. Since there are differences between available soil moisture measurement techniques and the spatial simulation scale, the approach of this paper was to include various measurement techniques in the model evaluation. Therefore, we believe that it is one of the strengths of the manuscript to show the comparison of mHM simulations to not only one type, but different soil moisture measurement techniques. Consequently, differences resulting from the measurement techniques are considered and discussed thoroughly in the manuscript.

The differences can be related e.g. to the site conditions, type of measurement devices and spatial or vertical scales. We were not entirely sure to which to which of these differences the reviewer refers. Therefore, we aim to point out in the following how we addressed them in the main manuscript and how to improve it.

The SM sites are not homogeneously distributed over site conditions (e.g.

lysimeters are placed in grasslands). In general, it could be shown that the model performances do not systematically depend on site conditions resulting from different measurement techniques (see also answer #4). We moved Figure A2 to section 3.1 to give it a more central focus.

Regarding the measurement devices, the single profile measurements (SPM), spatially distributed measurements (SDM) and lysimeters (LYSI) uses either Time or Frequence Domain Reflectrometry (TDR, FDR) sensors that do not expect differences in measurement quality. Cosmic Ray Neutron Sensing (CRNS) use a different measurements technique but were validated by spatially distributed measurements using TDR and FDR sensors (lines 200-204). As well, due to methodological reasons, days with snow were discarded from analyses for CRNS as further explained in the method section in lines 200-204 and possible implications in lines 304-307.

The measurements represent different horizontal spatial scales, with the area represented being smallest for SPM and LYSI (\approx point scale), while SDM (resp. the corresponding mean) and CRNS represent much larger areas at 0.1 km² (see lines 275–276). It was shown that SDM shows slightly better correlations to the simulations that support the closer scale match to the simulations (see e.g. 342–345). Those measurements are however still rare (see line 67) and since the paper focuses on SM dynamics the value of including SPM and LYSI measurements increases substantially. Additionally, CRNS data were excluded from the comparison in depth 25–60 cm due to its varying vertical penetration depth that does not allow a consistent depth-wise evaluation (see lines 325-236).

In order to facilitate a better standardization, not only the differences in the methods of the observations, the observation sites' conditions should be considered in explaining the research results, such as the landuse, elevation, precipitation of different sites.

Authors' response #4: We acknowledge the reviewer's recommendations. The main manuscript mentions that the German-wide available information (soil maps, land use) is considered (lines 263-265). No available sitespecific information is used in setting-up the hydrological model, e.g. precipitation is taken from daily available, interpolated DWD-observations and existing observations in the locations of soil moisture measurements are neglected. Hence, a different model setup would be required to meet the proposed standardization.

Correlations of deseasonalized SM between model and observations against sites' conditions (landuse as reported at the site, elevation and average precipitation) are plotted in the appendix of the main manuscript Figure A2 (see also Figure R2). As stated in the results section lines 285-286 "there is no general tendency for lower correlations at forest sites compared to crop and grassland sites (see Fig. A2)". Crop sites show slightly lower correlations than grassland sites, which is expected since anthropogenic activities (e.g., crop rotation) are not represented in mHM. Also, correlations show no clear tendency across the range of elevation and precipitation regimes. A further description is added in the manuscript in lines 295-299 together with moving Figure A2 to the main manuscript as proposed in #3.



Figure R2: Same as Figure A2 in preprint manuscript.

Additionally, the change of the landuse dataset and geology dataset in the two versions of the models has not been verified to have a positive effect on the simulation of soil moisture as the improvement in the second version of the model is not so evident (especially in spring and summer the change of Spearman rank correlation coefficient is negative). Further clarification for the effects of the model setting is needed in this research or in the future studies to be conducted by the authors' team.

Authors' response #5: The increase of model resolution in the second version of the drought monitor was motivated both by the release of a new German-wide soil map [1] and increased user need to higher resolution simulations as motivated in the introductory part of the main manuscript. This resulted in $\approx 1.2 imes 1.2 \, \mathrm{km^2}$ model resolution in the GDM-v2-2021 setup as a compromise between scientific/model perspective (limited by data availability and process representation) and stakeholder/user perspective (see conclusion lines 438-441). Changes in landuse data and geology data (also the change of projection to WGS-84) in the new drought monitor version (GDM-v2-2021) on the other hand were driven by current efforts of increasing the applicability and comparability of mHM to regions other than Germany and outside Europe. See also reviewer response #8 regarding the applicability of mHM outside of Germany. We acknowledge the critical remarks related with these changes but as pointed out in the following that the geology and landcover dataset changes have minor implications for SM drought simulations compared to the change in soil dataset. Currently, mHM takes relatively raw landuse classes. Species specific landcover is currently not accounted for. The difference in the resolution of GLOBCOVER and CORINE landuse dataset are in sub grid scale that influences the subgrid variability (GLOBCOVER resolution: 300 meters, CORINE < 100 m). Differences between the land cover datasets reduce if the land cover data is aggregated to the spatial resolution of the model. For example, at the spatial resolution of 1.2km, over 85 % of the grid cells both datasets agree on the dominant landcover class. This shows that differences stem from differences at high spatial resolution and do not have a large impact on the simulation. We included these aspects in the main manuscript to point out the limitations of the study. We added the following sentence in the main manuscript in line 147: "The changes in landuse and geology dataset can influence the simulations, yet play a minor role for the soil moisture simulations compared to the change in the soil dataset because changes of landuse data are in subgrid scale (resolution GLOBCOVER 300m, CORINE <100m) and no direct feedback of from saturated "groundwater" storage to soil moisture storage is implemented in mHM."

Moreover, the limited length of observed soil moisture data (< 10 years for most locations) causes some uncertainty in the comparison between observations and model results in the whole. More observations are needed to facilitate a more reliable model evaluation using the observation datasets as the existing observations within this paper are only validated in representing some regions of the Germany.

Authors' response #6: We compiled an unprecedented sample of soil moisture observations for hydrological model evaluation in Germany from different state-of-the-art measurement techniques and monitoring networks (FLUXNET and TERENO). We acknowledge the limitations of the length of observational data. In order to investigate the consequences of different time series lengths, Figure R3 shows correlation against length of time series. No trend of deteriorating correlations with length of time series can be detected (this sentence is added to the manuscript in lines 298-299). Nevertheless, we agree on the need to further broaden the observational database for future comparisons of model simulations to observations.



Figure R3: Correlations for the soil moisture observations against simulations (GDM-v2-2021 setup, 0-25cm depth) dependent of number of years of observations.

The authors can try making better simulations of the water cycle including soil moisture, ground water, and precipitation, while a higher-quality observational soil moisture dataset is applied in the future study. Other methods to indicate the correlations between the observations and model results can be used and also other indices apart from SMI can be utilized in indicating the severity of soil moisture drought.

Authors' response #7: Thank you for your comment. We have taken up the suggestion and added a statement in the Conclusion on the importance of "higher-quality" soil observations, especially with regard to future studies."

Last but not least, it would be better to have some discussions on the applicability of the model to other regions other than Germany. Are there some future plans on extending the regional applicability of the model? Besides, some discussions on what can we learn from a small-scale modelling to improve large-scale modelling can be stated.

Authors' response #8: The applicability of the model to multiple scales

and locations is a core element of mHM. The multiscale parameter regionalization (MPR) framework that is unique feature in mHM and now available as a standalone tool yields seamless model parameters at multiple scales and locations in an effective manner [14]. The effectiveness of MPR to transfer model parameters to scales and locations other than those used during calibration was demonstrated first in [11, 4]. The mHM was applied and evaluated in different climatological regions e.g. Europe [15, 8], West Africa [2], India [10] and US [5, 9]. [8] extensively evaluated mHM fluxes in Europe against evapotranspiration, soil moisture, runoff and total water storage (GRACE) (this sentence is added to the manuscript). [10] recently implemented a drought monitoring tool for South Asia based on mHM simulations and SMI. Global mHM simulations are currently conducted and evaluated within ULSYSSES project https://www.ufz.de/index.php?en=47367.

Some specific revisions regarding some parts of the manuscript are listed as follows: 1. The abstract is complete and correctly summarize the content of the paper, but it may need to be reduced a little to be more concise.

Authors' response #9: We reduced the abstract by some information that is also given in the introduction.

2. The process of parameter calibration and optimization needs to be in more detail to facilitate a reproducibility in the future study.

Authors' response #10: To improve reproducibility we extended the description of the calibration procedure with the following tables and put them in the supplements of the main manuscript.

- Table S1 showing the 200 sets of random multi-basin draws.
- Table S2 showing the results of the 201 basins from the final selected parameter set.

3. Some of the conclusions are overstated. The explanation should be in more detail regarding these issues.

(1) Figure 3 shows the time series of both the observations and model simulations. It seems that the coefficient is much higher than those when all the sites were selected. How the sites are selected may need to be mentioned if there are other sites that contain both Cosmic Ray Neutron Sensing (CRNS) and Spatially Distributed Measurements (SDM).

Authors' response #11: The sites have been selected based on the available time series length for different measurement methods. We added relevant information regarding the selection to the manuscript by extending the caption of Figure 3 with the following sentence: "The Hordorf site also contains both CRNS and SDM measurements, but with much shorter time series length. For visualization the stations with longer time series were selected."

(2) There is lower agreement between observations and simulations in winter.

Authors' response #12: This is shown and discussed against other studies extensively in lines 289-302 of the main manuscript. The reasons lie in the variable importance of hydrological processes in different seasons.

(3) There is improvement of second version of the model in representing the upper soil but stagnation in representing the whole soil.

Authors' response #13: We are not sure if we understand this comment correctly. Direct comparison of observed and simulated soil moisture is only possible for the upper soil due to observational data availability. For the total soil column, a comparison of drought intensities was performed showing similar results for both GDM versions. Assessing an improvement in soil moisture simulations between the GDM versions for the total soil column based on SM observations is not possible in the underlying study design.

(4) The values of the Spearman rank correlation coefficient are not high enough to conclude a definite improvement of the first version of the mHM (Table 2).

Authors' response #14: We clearly support this comment, and accordingly we did not conclude a definite improvement. See line 397 in manuscript: "However, the overall improvements were relatively small, partly because the lower resolution model setup (4x4km grid cells) was already capturing the observed SM dynamics well."

4. In general, the authors have given proper credit to related work and clearly indicate their own original contribution. The references are appropriate to the research, but it would be better if some more papers are referenced especially those in which multibasin model calibrations and the SMI were applied.

Authors' response #15: We have extended references with previous works focusing on the multi-basin calibrations, such as [6, 9, 7] (added to manuscript in lines 173-174). SMI was applied a.o. in [16, 12, 13, 3].

The suggestions regarding some minor flaws and typos are described as follows: Page 5, line 125: delete the ".However".

Page 5, line 131: delete the "," between "1.23" and "km".

Page 6, line 138: move "that were used in the analysis" before "are located".

Page 9, line 212, 213: remove "as"

Page 9, line 212, 213: add "," before "including", "the estimating", and "is hampered".

Page 11, line 270: remove "," after the "both".

Figure 3: add (a) to (l) for each sub panel to facilitate a better reference to the figures in the text when making the explanations.

Page 22: change the subtitle to "Conclusions and Discussions".

Page 23, line 436: remove "that".

Page 23, line 438: change "constitute" to "conclude".

The research is sound and fundamental. Some language edits could be good for improving the paper's quality.

Authors' response #16: We agree to the suggestions and have changed

them accordingly. We only kept the section "Summary and Conclusions", Page 22.

References

- [1] BGR. Digital soil map of Germany 1 : 200,000 (BUEK 200) v0.5, 2020.
- [2] M. Dembélé, N. Ceperley, S. J. Zwart, E. Salvadore, G. Mariethoz, and B. Schaefli. Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies. *Advances in Water Resources*, 143:103667, Sept. 2020.
- [3] M. G. Grillakis. Increase in severe and extreme soil moisture droughts for Europe under climate change. Science of The Total Environment, 660:1245– 1255, Apr. 2019.
- [4] R. Kumar, L. Samaniego, and S. Attinger. Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations: DISTRIBUTED HYDROLOGIC MODEL PARAMETERIZATIONS. Water Resources Research, 49(1):360–379, Jan. 2013.
- [5] B. Livneh, R. Kumar, and L. Samaniego. Influence of soil textural properties on hydrologic fluxes in the Mississippi river basin: Influence of Soil Textural Properties on Hydrologic Fluxes. *Hydrological Processes*, 29(21):4638–4655, Oct. 2015.
- [6] J. Mai, B. A. Tolson, H. Shen, [U+FFFD] Gaborit, V. Fortin, N. Gasset, H. Awoye, T. A. Stadnyk, L. M. Fry, E. A. Bradley, F. Seglenieks, A. G. T. Temgoua, D. G. Princz, S. Gharari, A. Haghnegahdar, M. E. Elshamy, S. Razavi, M. Gauch, J. Lin, X. Ni, Y. Yuan, M. McLeod, N. B. Basu, R. Kumar, O. Rakovec, L. Samaniego, S. Attinger, N. K. Shrestha, P. Daggupati, T. Roy, S. Wi, T. Hunter, J. R. Craig, and A. Pietroniro. Great Lakes Runoff Intercomparison Project Phase 3: Lake Erie (GRIP-E). Journal of Hydrologic Engineering, 26(9):05021020, Sept. 2021.
- [7] N. Mizukami, M. P. Clark, A. J. Newman, A. W. Wood, E. D. Gutmann, B. Nijssen, O. Rakovec, and L. Samaniego. Towards seamless large-domain parameter estimation for hydrologic models: LARGE-DOMAIN MODEL PA-RAMETERS. *Water Resources Research*, 53(9):8020–8040, Sept. 2017.
- [8] O. Rakovec, R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schäfer, M. Schrön, and L. Samaniego. Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins. *Journal of Hydrometeorology*, 2016.
- [9] O. Rakovec, N. Mizukami, R. Kumar, A. J. Newman, S. Thober, A. W. Wood, M. P. Clark, and L. Samaniego. Diagnostic Evaluation of Large [U+2010] Domain Hydrologic Models Calibrated Across the Contiguous United States. *Journal of Geophysical Research: Atmospheres*, 124(24):13991–14007, Dec. 2019.

- [10] T. R. Saha, P. K. Shrestha, O. Rakovec, S. Thober, and L. Samaniego. A drought monitoring tool for South Asia. *Environmental Research Letters*, 16(5):054014, May 2021.
- [11] L. Samaniego, R. Kumar, and S. Attinger. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale: MULTISCALE PARAME-TER REGIONALIZATION. *Water Resources Research*, 46(5), May 2010.
- [12] L. Samaniego, R. Kumar, and M. Zink. Implications of parameter uncertainty on soil moisture drought analysis in Germany. *Journal of Hydrometeorology*, 14(1):47–68, 2013. ISBN: 1525-755X.
- [13] L. Samaniego, S. Thober, R. Kumar, N. Wanders, O. Rakovec, M. Pan, M. Zink, J. Sheffield, E. F. Wood, and A. Marx. Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change*, 8(5):421–426, May 2018.
- [14] R. Schweppe, S. Thober, S. Müller, M. Kelbling, R. Kumar, S. Attinger, and L. Samaniego. MPR 1.0: a stand-alone multiscale parameter regionalization tool for improved parameter estimation of land surface models. *Geoscientific Model Development*, 15(2):859–882, Jan. 2022.
- [15] S. Thober, R. Kumar, J. Sheffield, J. Mai, D. Schäfer, and L. Samaniego. Seasonal Soil Moisture Drought Prediction over Europe Using the North American Multi-Model Ensemble (NMME). *Journal of Hydrometeorology*, 16(6):2329– 2344, Dec. 2015.
- [16] J.-P. Vidal, E. Martin, L. Franchistéguy, F. Habets, J.-M. Soubeyroux, M. Blanchard, and M. Baillon. Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite. *Hydrology and Earth System Sciences*, 14(3):459–478, Mar. 2010.

Original reviewer comments are in italics, authors' response is in bold.

Anonymous Referee #2:

Review of "High-resolution drought simulations and comparison to soil moisture observations in Germany" This manuscripts analyses the relationship between soil moisture observations and estimations by models in Germany with focus on drought monitoring. The manuscript is well written and organised. Nevertheless, I would like to include some caveats related to the limitations of the validation approach and the usefulness of the new high spatial resolution data base in order to assess drought severity. I include specific details related to these issues (and others) below (numbers refer to the specific lines of the manuscript):

Authors' response #1: We thank the Reviewer for the assessments of our work. We paid detailed attention to all comments and we have addressed all of them below accordingly.

11- What is "vegetation period"? Is maybe "vegetative active period"?

Authors' response #2: We agree to the suggestion and will change terms in the manuscript.

Table 1- I would like to ask for a technical question. Do you think if the quality of the globcover map is sufficient for the modelling. How is considered the uncertainty of land cover information in the model? I find very high detail of information related to the improvement of the soil maps, map I have the impression that the land cover data is not considered so carefully and it can be strongly relevant to model soil moisture given different water consumption by ecosystem types (even at the scale of species), the role of root structure, root depth, etc.

Authors' response #3: The hydrological simulations of German drought monitor operate at the nation-wide scale with large-scale available information. One of the research questions was to evaluate whether it possible to provide higher resolved information at a satisfying quality.

The increase of model resolution in the second version of the drought monitor was motivated both by the release of a new German-wide soil map [1] and increased user need to higher resolution simulations as motivated in the introductory part of in the main manuscript. This resulted in $\approx 1.2 \times 1.2 \text{ km}^2$ model resolution in the GDM-v2-2021 setup as a compromise between scientific/model perspective (limited by data availability and process representation) and stakeholder/user perspective (see also conclusion lines 438-441). Changes in landuse data (also the change of geology data and projection to WGS-84) in the new drought monitor version (GDM-v2-2021) on the other hand were driven by current efforts increasing the applicability and comparability of mHM to regions other than Germany and outside Europe. Change in these landuse datasets have minor implications compared to the change in the soil dataset. Currently, mHM takes relatively raw landuse classes. Species specific land-cover is currently not accounted for. The difference in the resolution of

GLOBCOVER and CORINE landuse dataset are in sub grid scale that influences the subgrid variability (GLOBCOVER resolution: 300 meters, CORINE < 100 m). Differences between the land cover datasets reduce if the land cover data is aggregated to the spatial resolution of the model. For example, at the spatial resolution of 1.2km, over 85 % of the grid cells both datasets agree on the dominant landcover. This shows that differences stem from differences at high spatial resolution and do not have a large impact on the simulation.

Nevertheless the reviewer's comment is a very relevant one and we extended the manuscript by pointing out this specific limitations of the study. We added the following sentence in the main manuscript in line 147: "The changes in landuse and geology dataset can influence the simulations, yet play a minor role for the soil moisture simulations compared to the change in the soil dataset because changes of landuse data are in subgrid scale (resolution GLOBCOVER 300m, CORINE <100m) and no direct feedback of from saturated "groundwater" storage to soil moisture storage is implemented in mHM."

150- I find very few information related to the meteorological data. There is not information on the number of stations used for each variable, the quality of the data, quality control processes, data gap filling, temporal homogeneity, etc., but also information related to the quality of resulting gridded data (e.g., cross-validation statistics would be useful). Meteorological data can be also an important source of uncertainty in the model outputs...

Authors' response #4: The meteorological input station data that is used for interpolation is provided by the German Weather Service (DWD) through the Climate Data Center (ftp://opendata.dwd.de/climate_environment/ CDC/). It is subject to extensive quality controls [2]. Additionally, quality controls are implemented in the preprocessing steps of the interpolation routine e.g. checking plausible variable range. In [5] describing the mHM simulations underlying the GDM version 1, the interpolation method for interpolating the meteorological data is described and validated in detail. Different approaches to calculate theoretical semi-variograms were tested and evaluated. A cross-validation (Jackknife method) was performed to test the ability of the External Drift Kriging (EDK) to estimate meteorological variables at the measurement locations. According to [5] the average and the standard deviation for the different errors assessments over all stations for daily interpolated precipitation values were 0.01 and 0.15 mm d - 1 for the bias, 0.64 and 5.60% for the relative bias, 0.93 and 0.03 for the Pearson correlation coefficient, and 1.75 and 0.48mm d-1 for the root mean square error. Additionally, a comparison of the EDK interpolations conducted by [5] to the REGNIE gridded precipitation data [4] provided by DWD showed satisfactory results with spatially averaged bias of the daily precipitation fields of 0 with a standard deviation of 0.11mm d-1 within the period 1951–2010. We extended the description in section 2.2.1 in the main manuscript.

151-154- What about uncertainty of the Hargreaves-Samani equation to estimate Potential Evapotranspiration? It is widely known that temperature based methods show uncertainties related to physically based models like the Penman-Monteith equation. For example, wind speed and relative humidity may have large importance on PET, even more in non-stationary scenarios charecterised by decreased relative humidity over land and wind speed reduction.

Authors' response #5: The actual ET is the important water balance component being the reduction term of the potential Evaporation. Comparisons of actual ET estimated with mHM were conducted in [5] comparing to remote sensing data (MODIS) and FLUXNET towers and by [3] over Europe using in situ observations and a gridded product from FLUXNET showing a good overall fit.

We certainly agree on the superiority of Penman-Monteith methods to estimate PET at the field scale if high quality field-scale data is available. The conclusion in Line 436 refers to this comment stating that "we may achieve a more precise estimation of potential evapotranspiration through implementing the Penman-Monteith methods.". Regionalized estimates of physical based PET are still however largely limited by spatial data availability in terms of number of measurement stations and temporal data availability in terms of record lengths. No reliable high quality daily gridded estimates for both wind and global radiation are currently available for full time period (1950-2020) to allow Penman-Monteith ET estimation at the scale of Germany used in the study. A longer simulation time period is prioritized for the German Drought Monitor to obtain a long statistical database for the SMI estimation instead of cutting the simulation period.

172- Figure 1 > Figure 2. 231-235- The validation procedure is exclusively based on correlations. Nevertheless, if the main purpose of the manuscript is related to drought monitoring, I think more relevant to assess model outputs during periods of water deficits. For example, it would be useful to check the capability of models to identify duration and magnitude of the dry periods. High correlation could mask a poor goodness between observations and models during dry periods. I would suggest to include statistics focusing on the drought periods in addition to the non-parametric correlations.

Authors' response #6: The soil moisture index (SMI) is estimated for every grid cell and every day of the year. Hence, the number of data points to estimate the histogram and percentiles to classify drought is equal to the number of years with observational data. Due to the limited observational data time series lengths, it is not possible to estimate drought characteristics as intensities and duration. Therefore, we decided to use the time span 1951-2015 initially in setting up the first version of the drought monitor to ensure statistical stability of the system. Due to these limitations, we came up with our study design: comparison of observed and simulated soil moisture in a first step and comparison of simulated drought intensities in the following step between the two model setups. The correlation statistics are calculated on deseasonalized anomalies that removes the seasonal mean cycle. From a mathematical point of view, having a constant bias between the observations and the simulated soil moisture would have no effect on the drought classification. The percentile based approach of the SMI would remove the bias. We therefore believe that this analysis adequately describes the model performance

under the available data.

170-210- The length of the observation series is not indicated in this section. This information is relevant to assess robustness of the relationship between observations and models. Have the series the same length? How is this considered in the assessment of the signification of the relationships? I think this issue is affecting the validation of the results over the entire section 3.1 since the length of the series affect the degrees of freedom of the correlation analysis. I see in table 3 that the length of the series is between 2 and 5 years, which is too low to provide a robust validation of the model outputs.

Authors' response #7: We generally agree that longer observational time series would support a more robust validation. Nevertheless, in our study we compiled the best possible observational soil moisture data base on the national scale for Germany. We describe the time series lengths more clearly in the manuscript by adding following sentence "Time series lengths of the observations are between 2.8 and 17.8 years with a median (mean) of 6.5 (6.7) years."

In order to investigate the consequences of different time series lengths, Figure R1 shows correlations against the length of time series. No systematic relation between correlations and the time series length can be detected. This statement has been included in the main manuscript in lines 298-299.



Figure R1: Correlations for the soil moisture observations against simulations (GDM-v2-2021 setup, 0-25cm depth) dependent of number of years of observations.

Figures 6 and 7. Under my opinion, I do not think that this information is providing an useful output to determine the goodness of providing additional spatial resolution to assess drought severity. Large scale statistics are aggregating the information, being normal that both databases at 4km and 1 km of spatial resolution provide similar results. I think the relevant information of the 1 km modelling approach is not the general large spatial pattern but the local differences that could emerge given higher spatial resolution. This is something interesting to be analysed (e.g. using spatial statistics: the variance between grid cells, the differences between areas characterised by diversity of land cover/soil characteristics) to determine if higher spatial resolution is providing relevant information for drought monitoring and management. Observing Figures 6 and 7 I would say that the higher spatial resolution is really not needed as it basically identifies the same patterns that 4 km grids.

Authors' response #8: Thank you very much for the suggestions to include spatial statistics to show the regional differences between the model setups.

Drought intensities that are shown in Figures 6 and 7. can reach a maximum value of 0.2. Figure 7 shows that in the absolute differences up to 0.1 occur between both drought monitor versions on the grid scale. This clearly shows a large impact of the new study setup on simulated drought characteristics. The drought intensity color scale in Figure 7 was improved in the revised version of the manuscript.

Following the reviewers suggestions, we conducted an additional analysis that complement the analysis of the drought clusters. In Figure R2 the variance between grid cells for drought intensities during vegetation active period are shown as semi-variograms. In general, the spatial variance is larger in the total soil than top soil. The GDM-v2-2021 setup shows a general larger spatial variance between grid cells in the top soil and larger increase with distance (see Figure R2 (a)). The spatial variance in the total soil is lower at smaller distances in the GDM-v1-2016 setup, but slightly higher at larger distances. Figure R2 (b) showing semi-variance normalized by distance (and log scaled x-axis to to improve visibility of smaller distances) demonstrates that in the GDM-v2-2021 setup the distance-normalized variance of drought intensities is increased especially at small spatial scale in both the top and total soil, indicating larger local differences in response to drought intensities. We included Figure R2 in the main manuscript in section 3.2 and added a paragraph based on the findings above (lines 396 - 403).



Figure R2: Empirical semi-variograms for drought intensities during vegetation active period in upper Soil for GDM-v1-2016 and GDM-v2-2021 setup. The bin size was set to 5 km (nearest larger even km bin size relative to the GDM-v2-2016 modelling resolution). The len scale and nugget of the fitted exponential theoretical semi-variograms are noted in the legend. In Subplot b) the y and x axis are log scaled.

References

- [1] BGR. Digital soil map of Germany 1 : 200,000 (BUEK 200) v0.5, 2020.
- [2] F. Kaspar, G. Müller-Westermeier, E. Penda, H. Mächel, K. Zimmermann, A. Kaiser-Weiss, and T. Deutschländer. Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre. Advances in Science and Research, 10(1):99–106, Aug. 2013.
- [3] O. Rakovec, R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schäfer, M. Schrön, and L. Samaniego. Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins. *Journal of Hydrometeorology*, 2016.
- [4] M. Rauthe, H. Steiner, U. Riediger, A. Mazurkiewicz, and A. Gratzki. A Central European precipitation climatology – Part I: Generation and validation of a highresolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift*, pages 235–256, July 2013.
- [5] M. Zink, R. Kumar, M. Cuntz, and L. Samaniego. A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrology and Earth System Sciences*, 21(3):1769–1790, 2017. ISBN: 1607-7938.