Manuscript ID: hess-2021-402
Response to the Reviewers

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 extensively described in the main manuscript. This resulted in $\approx 1.2 \times 1.2\,\mathrm{km}^2$ 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 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 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 sub-grid 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.
We will include these aspects in the main manuscript to point out the limitations of the study. We propose to add 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 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 fields of 0 with a standard deviation of 0.11 $mmd{-}1$ within the period 1951–2010.

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 spatial modelling scales 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.

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 suggest to 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.

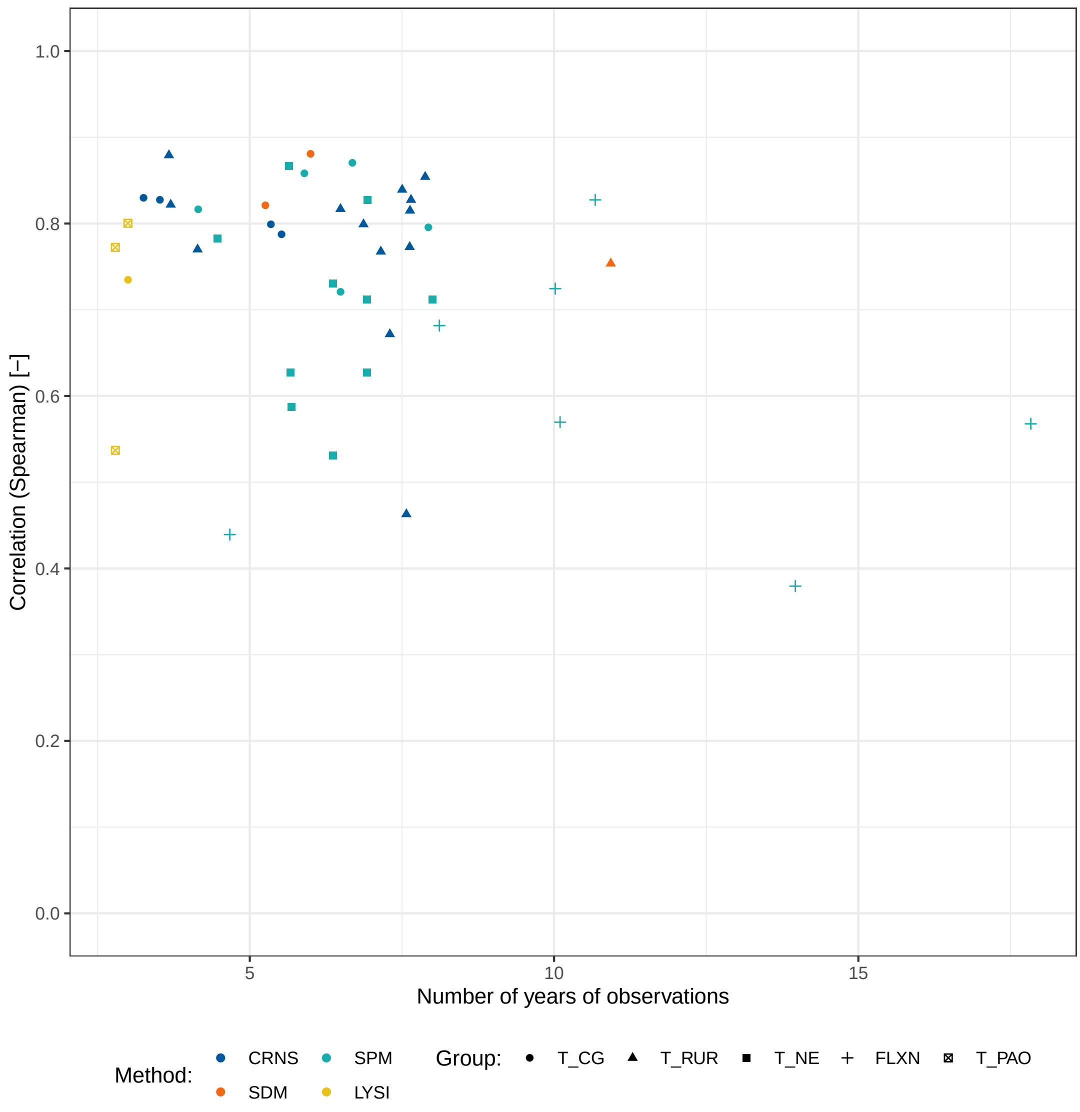

Figure R1: Correlations for the soil moisture observations against simulations (GDM-v2-2021 setup, 0-25cm depth) dependent of number of years of observations.

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 Fig. 6 and 7. can reach a maximum value of 0.2. Fig 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.
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 suggest to include Figure R2 in the main manuscript in section 3.2 and add a paragraph based on the findings above..

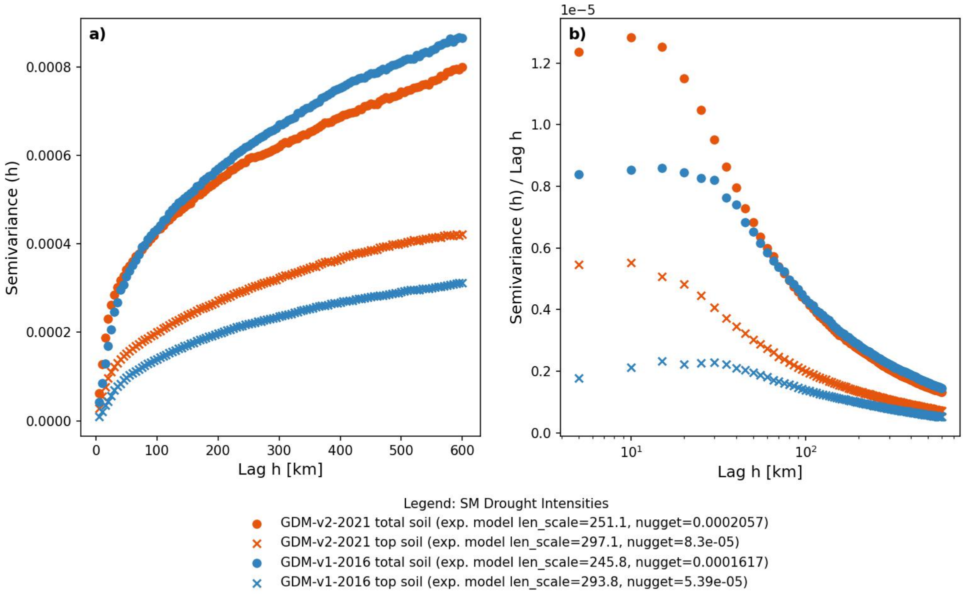

Figure R2: Empirical semi-variograms for drought intensities during vegetation active period in upper Soil for GDM-v1-2016 and GDM-v2-2021 setup. The bin size was set to 5 km (nearest larger even km bin size relative to the GDM-v2-2016 modelling resolution). The len scale and nugget of the fitted exponential 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 high-resolution 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.