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

Referee #3:

Review of Boeing et al., HESS-2021-402 "High-resolution drought simulations and comparison to soil moisture observations in Germany"

This study investigates the performance of the land surface modelling system underlying the German Drought monitor. The authors compare two model versions with different spatial resolution against a comprehensive set of soil moisture measurements obtained through different methods and from different regions across Germany. They find overall similar performance with slight improvements in the cold season, while additionally regional changes in the results can be observed, probably following changes in the underlying soil type quantification.

———————-

Recommendation: I think the paper requires major revisions.

The topic of the study is interesting and timely. Drought monitoring in Germany is of increasing relevance given the recent droughts, and their impacts on agriculture, forestry and water resources. Accurate drought monitoring, as well as respective analyses and prediction, requires accurate land surface models which can represent relevant processes. In this context, the current study is a relevant contribution as it compares the modelling system of the German drought monitor with a comprehensive set of soil moisture measurements obtained through different techniques, and from different regions throughout the country. Further, the study illustrates the advantages of performing simulations at higher spatial resolution which is increasingly relevant for regional- scale drought assessments. I think that the manuscript is a good match match for the readership of HESS, and appreciate the efforts made by the authors to improve the manuscript in response to the previous comments of both reviewers.

However, before it is ready for publication some concerns particularly related to the understanding of differences between modelled and observed soil moisture dynamics should be resolved, as detailed below.

Authors' response #1: We really appreciate your time and efforts in providing detailed and relevant comments to improve the quality of our work. During the revision, we paid detailed attention to all critical comments and we have addressed them with our best efforts. In the following you find all corresponding answers.

———————

General comments:

(1) I agree with both reviewers that additional explanations/analyses are necessary for understanding the differences between the modelled and observed soil moisture dynamics, and feel that the author's responses in this context fall short in several aspects.

(i) Why does the updated model version compare better with observations in fall and winter, but not so much during the growing season?

Authors' response #2: Thank you for this comment. We recognise that several of the changes in the model setup may provide explanations for the improved model performance in the fall and winter. The higher modelling resolution of the 1 km runs may better resolve the sub-grid variability of cold season related processes such as snow accumulation that improves

the simulated SM dynamics. Additionally, the finer spatial soil texture representation possibly contribute to an improved model representation of soil wetting/drying e.g. especially during saturated conditions in the cold season. Nevertheless, we would like to stress that the changes in these seasons are significant in statistical terms, but the differences in absolute terms are still rather small ($\Delta + 0.07$ for fall and $\Delta + 0.12$ for winter over all locations). We added these clarifications to the revised manuscript.

(ii) Following the main purpose of the German drought monitor, the validation of the modelled soil moisture against observations needs to be performed also with an exclusive focus on drought periods in addition to the overall agreement already assessed in the paper. Even though there is less data available during drought periods, I feel it should still be sufficient for a meaningful analysis, particularly given the relatively low threshold used in the definition of droughts, and the fact that the time series length does not systematically affect the determined agreement between modelled and observed time series (Figure R1).
Authors' response #3: We thank the Reviewer for this suggestion. To address this comment, Figure R1 shows the dry anomaly spectrum based on the Spearman rank correlations between simulated and observed deseasonalized SM anomalies that fall below the 20th percentile in the observed SM time series. It is important to emphasize that we do not aim to estimate drought periods here, as its solid calculation requires much longer time series. The estimation of drought is performed using histograms for every grid cell and day of the year (see method section 2.4.1 in the manuscript). Consequently, estimating robust percentiles requires time series lengths of minimum 30 years – this means that the time series length of the observational data is considered insufficient. Figure R1 a) shows a median correlation of 0.61 over all observations in the GDM-v2-2021 (1-km setup). The performance in the two model setups remains similar. However, the comparison separated between the measurements with larger spatial footprint (SDM, CRNS) and point scale measurements (SPM, LYSI) shows that the agreement between simulations and the larger footprint observations increased towards the high resolution setup, but the median agreement to the point scale SM measurement decreased. In general, the measurements with larger spatial footprint display higher agreement to the simulations. Due to the varying day-to-day variability of SM between the SM observation types and the simulations, in Figure R1 b) additionally a statistical smoothing was applied by calculating a running 30 day mean on the daily SM time series before removing the seasonal cycle. This approach is similar to the SM pre-processing for the SMI as proposed in Zink et al. [2016]. Figure R1 b) shows when smoothing is applied, the agreement between observations and simulations during dry periods can be substantially improved to a median correlation of 0.7 over all observations in the GDM-v2-2021 setup ($\approx$ 1-km resolution). Especially the agreement between the point scale measurements and simulations is increased to a median correlation of 0.63 in both model setups. We added this paragraph and Figure R1 to the section 3.2 of the revised manuscript.

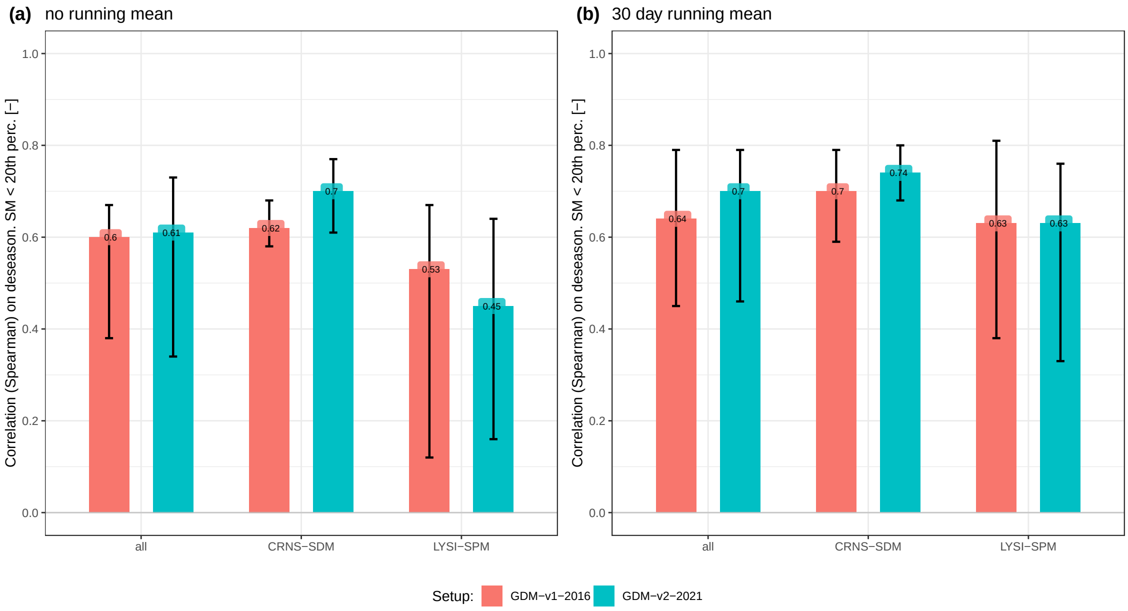

Figure R1: Correlations of deseasonalized soil moisture below 20th percentile (based on the observed SM timeseries) between simulations and observations. In b) additionally a statistical smoothing was applied by calculating a running 30 day mean on the daily SM timeseries before subtraction of the seasonal cycle. The correlations are shown for all observations (n=46) and separated between observations with larger spatial footprint (n=20) including Cosmic Ray Neutron Sensing (CRNS) and spatially distributed measurements (SDM) as well as point measurements (n=26) including single profile measurements (SPM) and lysimeters (LYSI) .

(iii) To better understand the small-scale differences between both model versions it would be helpful to relate the obtained differences (e.g. in Figure 8) to the underlying differences between the soil type maps and land use classifications to establish some cause-effect relations.
(iv) The role of updates in the land use vs. soil datasets, as raised by both reviewers, is not really clear even in the updated version of the manuscript. While the authors state that changes in the land use dataset do not affect the results, it remains unclear how they arrived to this conclusion.
Authors' response #4: We would like to thank the Reviewer for his suggestion. We combined question (iii) and (iv) due to their similarity. The study design was conceptualized to present a comparison of the high resolution 1 km setup to soil moisture observations and secondly a comparison to the previous coarser resolution operational modelling setup. The major motivation to test a newer version of the hydrological modelling system was the availability of the BUEK200 soil dataset, resulting in a 25-fold spatial resolution increase. Also other changes were implemented (e.g. increased grid cell resolution of the hydrological simulations, change of landuse and geology dataset). Additionally, the calibration parameters of both setups were estimated independently. We would like to clarify that our study does not aim to explain methodological differences between model setups (e.g. the effect of solely changing the soil maps on soil moisture), but to provide a comparison of two operational modelling setups used for drought monitoring. To address these important and interesting questions, a different study setup would be required.

However, to eliminate any reader's confusion, we have added these limitations more clearly to the revised manuscript: Figure R2 demonstrates the different roles of the change in SM dynamics between the model setups related to the specific soil and land use datasets by showing temporal correlations between SM from separated model runs fixing all model settings (L1 $\approx$ 1.2 $\times$1.2 km^2 resolution, default mHM parameters) only changing the soil dataset (BUEK200 – BUEK1000) and in a separate step only changing the land use dataset (CORINE – GLOBCOVER). Figure R2 suggests that the change of the soil dataset has a much larger impact on the SM simulations compared to the change of the land use dataset in these specific model setups. The CORINE and GLOBCOVER land use datasets both have already high horizontal resolutions ($\approx$100 m and 300 m, respectively). The differences between the land use datasets mostly lie in the sub-grid scale of the mHM hydrological modelling resolution and have a minor effect on the upscaled hydrological response at the L1 level (here $\approx$ 1.2 $\times$1.2 km^2). This paragraph and Figure R2 has been added to section 3.2 and the appendix in the revised manuscript, respectively.

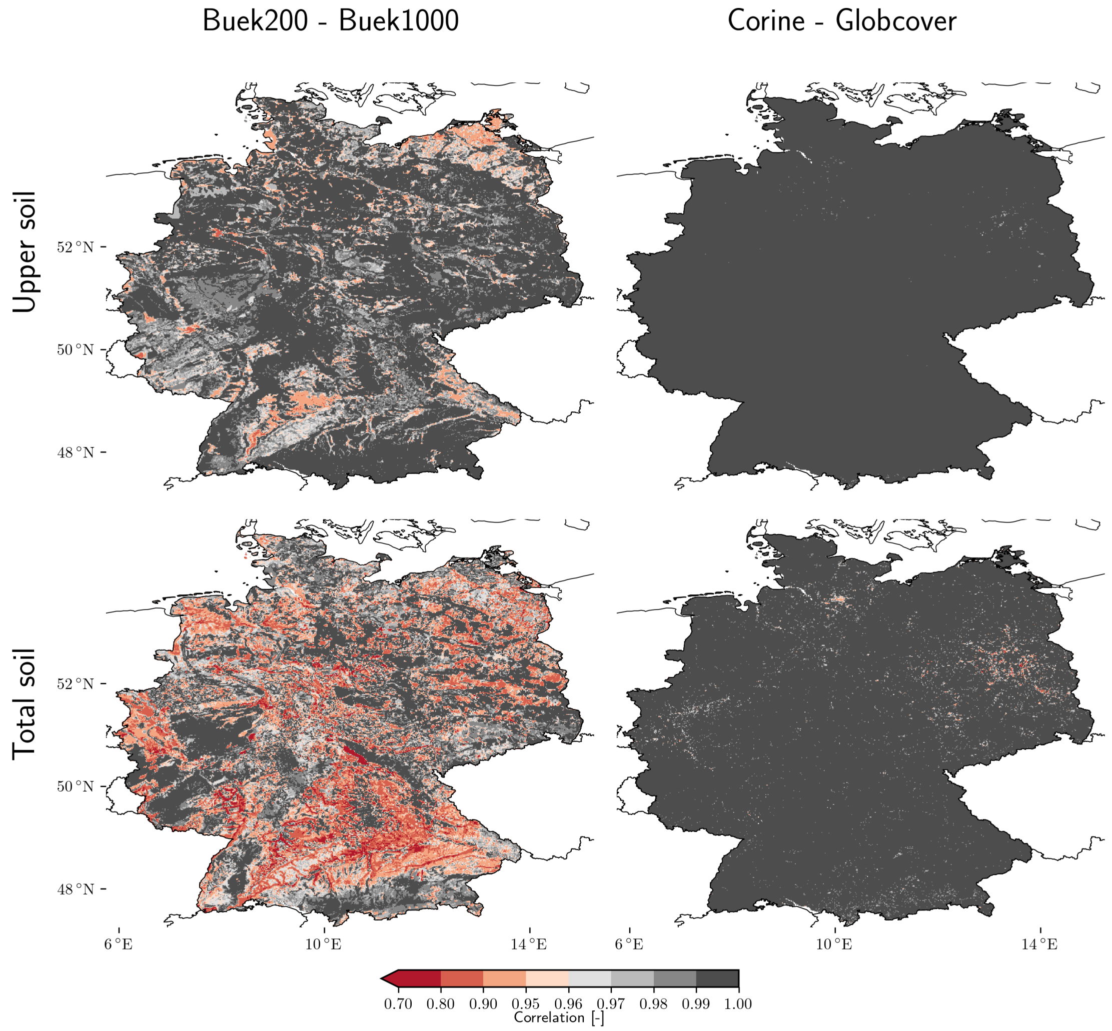

Figure R2: Correlations between simulated daily SM using model runs for the time period 1991-2019 keeping the model settings identical (L1 1.2km^2 resolution, default mHM parameters) except changing soil dataset BUEK200 versus BUEK1000 (left) and secondly changing landcover dataset CORINE versus GLOBCOVER (right).)

(v) Finally, it would ask the authors to add some discussion on the poor agreement between models and observations in winter months (and partly also in May). I realize that this might also be related to poor quality of the measurements in the cold season, but wonder what this lack of validation means for the applicability of the Drought monitor during these months of the year?

Authors' response #5: Thank you for your suggestion. Several aspects related to lower correlations in winter in the manuscript were discussed, e.g.,: (1) "Lower correlations in winter can be related to higher uncertainties in simulations as well as observations with respect to frozen soils and snow cover." (307/308); (2) Figure A3 shows that the spatially distributed measurements typically have very high correlations in winter, while the CRNS measurements show lower correlations. (3) Additionally "Low correlations between simulation and observations are accompanied by low correlations between the measurement methods, especially in winter." (L 308/309). This implies that the low correlations in CRNS to simulations

are likely due to uncertainty in the measurements, yet we also emphasized that only a first order comparison of CRNS to the simulated SM was applied here. Moreover, we mentioned "As snow days were removed from the time series in the CRNS measurements, the anomaly calculation from the remaining data was impacted by a smaller and incomplete sample size" (lines 316/317). Now, we added one more aspect that the sensor quality of SDM, SPM and LYSI in winter can be reduced during frost days. Especially, SPM and LYSI measurements can be affected by sensor failures as they rely only on few sensors compared to the spatially distributed measurements (SDM) with a larger number of sensors.
Uncertainties in the hydrological model mHM related to the cold season are also discussed. "Furthermore, the mHM model does not contain a full energy balance model, which limits the description of soil frost depths." (L 312,313). Regarding the lower correlations in May we hypothesize that they may be attributed to the static vegetation module in mHM. Lower correlations of deseasonalized anomalies in May are detected especially at forest locations (median of 0.63). The timing of leaf unfolding in trees, usually between late April to May [Chen et al., 2018], is subject to annual fluctuations and affects evaporation from the soil and thus soil moisture dynamics. We added this explanation to the revised manuscript.
Finally. regarding the applicability of the GDM in winter time, the data is currently used e.g. by water suppliers all over Germany as an indication for the development of groundwater levels. The daily updated data of the current GDM-v1-2016 is used in several operational federal-state and national information systems in Germany, such as the National Drought Atlas of the Federal Agency for Cartography and Geodesy. No specific winter-time limitations have been reported to the authors.

(2) In this study drought is defined to occur 20% of the time. This seems arbitrary and is not related to actual impacts of drought on ecosystems and/or society. Further, this seems quite a high percentage in my opinion; if vice versa another 20% of the time then would be associated with conditions with too much water, this effectively means that 40% of any time series considered in this study are regarded as extreme conditions.
Authors' response #6: The SMI drought threshold concept used in the German Drought Monitor is based on the D0-D4 classification system for droughts from the US-Drought monitor [Svoboda et al., 2002] that related drought categories to potential impact types. The drought thresholds reflect the occurrence of similar soil moisture conditions in the past and hence indicate the potential impacts of these conditions [Zink et al., 2016]. The 20th percentile is defined as moderate drought conditions, that indicate conditions of "possible damages to crops and pastures". Extreme drought conditions are defined as the 5th percentile indicating "high probability of major losses in crops and pastures".
The drought intensities as depicted in Figures 7 and 8 take into account the degree of negative departure from drought conditions (SMI < 0.2) (hence, the extremer the drought conditions, the higher the intensities) as well the temporal aggregation length and the spatial aggregation area. Madruga de Brito et al. [2020] evaluated a broad range of drought impact types through media impact statements. The SMI based drought

magnitude showed reasonable correlations with drought aid as well as agricultural and livestock impacts.
The resulting impact of soil moisture drought conditions need to be identified for each specific impact type based on the timing within the year and duration of drought conditions. For example, Peichl et al. [2018, 2021] identified specific monthly damage functions between the SMI and different crops using varying statistical methods. The work showed that dry SM anomalies in some months can reduce yield (e.g. August, September for maize), while in other months it may increase crop yield (e.g. May for maize). Impacts of SM droughts can affect a broad range of sectors besides agriculture. Especially, the considered soil depth of the SMI is relevant for different sectors. While the drought conditions in the upper soil (0–25 cm and 0–60 cm) are more relevant to agriculture, drought in the total soil column (up to 2 meters) indicate potential impacts on water resources and the forestry sector. We added explanations related to the SMI drought concept and impacts to the section 2.5 of the revised manuscript.

(3) There are many small language errors (such as missing articles or wrong grammar) throughout the manuscript. This was also pointed out by reviewer #1, but apparently the authors mainly corrected the issues explicitely pointed out by the reviewer. Several errors are left in the manuscript, as stated below in my specific comments. The authors need to take special care of these points, and of an accurate and correct language throughout the entire manuscript when revising the manuscript.
Authors' response #7: We carefully double-checked the manuscript, and on top of that, the final revised version was corrected by the native speaker.

(4) The processing of the in-situ soil moisture measurements is not entirely clear. For example I did not get how the measurements, which represent different soil depths, are aggregated to the layers considered in the analysis.
Authors' response #8: We are sorry for causing any confusion here. Based on the number of available in situ sensors in the respective simulation soil depths (0-25 cm, 0–60 cm and 25–60 cm) as noted in Table 3, a simple weighted vertical average was calculated. Highest weights were allocated when the sensor depth is located in the center of the soil depth range and weights linearly decrease towards the edges of the soil depth range. We added this clarificationon the weighting procedure in the end of section 2.3

Further, I understand that in-situ data are averaged across stations in case they are located in the same grid cell. But what happens if the time series length between the stations is different? This could induce spurious signals in the aggregated time series which may affect the comparison with the model data. I appreciate that this is mentioned as a limitation by the authors in line 443, but feel that this should be addressed by e.g. normalizing the in-situ time series to make them more comparable.
Authors' response #9: The in-situ data was only spatially averaged for the available sensors of the spatially distributed measurement networks (SDM). As pointed out in line 443 not all sensors were active over time

within the measurement networks. We kindly ask the Reviewer for understanding that research on the effect of changing sensor sample size is beyond out of scope for this study. However, as stated it it is currently subject of an ongoing research. An improved spatial average calculation can be included in future studies.

I do not wish to remain anonymous - Rene Orth.

Specific comments:
abstract: it would be good to mention the analyzed soil depth
Authors' response #10: Thanks, this was added.

line 9: replace "older" and "present" with more informative terms
Authors' response #11: We removed those terms.

line 15: here you could cite Orth et al. [2022]
Authors' response #12: Thank you for pointing to this reference. We added it accordingly.

line 36: Why is relative in quotation marks? And I feel that "relatively" would fit better.
Authors' response #13: We have changed wording to "relatively".

line 54: "of" is not needed
Authors' response #14: Thank you for pointing this out, however in this case we could not find the "of" to remove.

line 58: would change "usually" to "sometimes" as many Fluxnet stations do not measure soil moisture at all
Authors' response #15: We agree and have changed accordingly.

line 59: I would recommend to cite here Koster et al. [2019] and to include its findings in the discussion here.
Authors' response #16: Thank you for pointing to this interesting work. We added the findings in the revised introduction. Additionally, we included O et al. [2020] in the introductory discussion of uncertainty in hydrological modelling in line 46/47.

line 67: I guess this is about *reflected* neutrons?
Authors' response #17: During the revision we discussed the proposed addition of the word "reflected" to describe the detected neutrons. However, we concluded to not change the sentence. The term "reflected" would not be fully correct here, because (i) the detector also measures parts of the incoming neutrons directly, and (ii) the neutron interaction physics in the soil is – strictly speaking – not a reflection, but rather a complex combination of processes involving inelastic scattering (capture), energetic excitation of the atomic nucleus, nuclear evaporation, and subsequent elastic scattering. The use of the word "neutrons" here covers the overall process more generally, without going too much into the details.

line 83: "...observed soil moisture that constitute the basis...", please improve phrasing
Authors' response #18: We have rephrased it.

line 85: "...higher spatial resolved soil...", please improve phrasing
Authors' response #19: We changed it from "spatial" to "spatially".

line 105: insert "the" before "US", and explain abbreviation
Authors' response #20: We rephrased into "the conterminous United States".

line 107: "The aggregation has based a set of upscaling rules...", please improve phrasing
Authors' response #21: We rephrased to "The aggregation is based on a set of upscaling rules.."

line 128: Germany does not extent down to 45N
Authors' response #22: Thank you for pointing to that typo. We changed it to 47.25N.

lines 136/137: I do not understand this.
Authors' response #23: We rephrased the entire paragraph to make it more understandable. Please, check the revised manuscript.

lines 145-148: I do not understand this.
Authors' response #24: This is related to Reviewer's question (iii) and (iv). We have extended the description and moved it to the end of section 3.2. of the revised manuscript (see also Author's comment #4).

line 153: insert "the" before "plausible variable range"
Authors' response #25: Text was changed accordingly.

line 154-156: Please provide more information on the calculation of the variograms, including more motivation for doing this.
Authors' response #26: Added information and motivation for calculating the variograms, as follows: "The meteorological station data is subject to extensive quality controls [Kaspar et al., 2013]. Additionally, quality controls such as checking the plausible variable range are implemented in the preprocessing steps of the interpolation routine. Theoretical variograms are estimated based on all available station data to derive seamless fields of hydro-meteorological fluxes and states for entire Germany [Zink et al., 2017]. An exponential model is used for precipitation and spherical models for the temperature variables."

line 156-158: no actual surface radiation is used as an input?
Authors' response #27: Potential evapotranspiration is calculated based on the Hargreaves-Samani equation using average, minimum and maximum daily temperatures. Extraterrestrial radiation is estimated based on the latitude and day of the year. This was added to the revised

manuscript.

line 202: superflous bracket
Authors' response #28: Done.

line 235: insert "the" before "sample size"
Authors' response #29: Done.

line 236: "data were masked to...", please improve phrasing, and provide additional details
Authors' response #30: The observed SM time series have data gaps (see Table 3 for availability). The data points in the simulations at these days were removed in a pre-processing step to allow a comparable calculation of SM seasonality on the available data. We added this description to the Method's section.

lines 244, 263: Why are these reference time periods different?
Authors' response #31: The reference time period of the cumulative density functions is 1951–2015. The SMI is then calculated for 1952–2020 based on these cumulative density functions. We included these details including improved phrasing in the revised manuscript.

line 294: typo in "represented"
Authors' response #32: Corrected.

line 295/296: "model performances does...", please improve phrasing
Authors' response #33: We changed to "model performance does.."

line 296: "correlations with length of time series...", please improve phrasing
Authors' response #34: Improved.

line 331: "generally higher", please explain
A possible explanation is given in line 333/334. We rephrased this part to make it more clear.

line 344: variable rooting depths is not a process
Authors' response #35: We changed to species-specific root water uptake. In this respect, we identified a typo in line 141 where tillage depth was falsely denoted as rooting depth. We changed that accordingly.

line 381: insert "the" before "total soil column"
Authors' response #36: Done.

line 392: "2019 total soil", please improve phrasing
Authors' response #37: We improved accordingly.

line 399: remove "to"
Authors' response #38: Done.

line 433: change "to be determined for" into "across"

Authors' response #39: This was changed accordingly.

lines 448-454: This is a very important point, good to see this here!
Authors' response #40: Thank you for appreciating.

Table 1: What is meant with geology dataset?
Authors' response #41: The dataset defines aquifer properties that govern the base-flow recession rate. This description was added to the Method's section. We changed the term geology to hydrogeology in Table 1 accordingly.

Table 2, caption: please explain abbreviations dfj, jja, mam, son
Authors' response #42: The text was amended accordingly.

Figure 1, caption: "Soil physical properties clay...", please improve phrasing
Authors' response #43: We rephrased accordingly.

Figure 4, caption: panel d) is not mentioned
Authors' response #44: We added description to figure's caption.

Figure 5, caption: "positive correlations are marked with x", I do not understand this as there are also x symbols denoting negative correlations in the figure (?)
Authors' response #45: In the caption it was stated that „data points not fulfilling conditions of significant correlations (p-value < 0.05) and positive correlations are marked with x." The second condition of positive correlations between simulations and observation was included since also negative correlations could be significant, but would not indicate the quality of correlations. We improved the phrasing accordingly.

References

L. Chen, J.-G. Huang, Q. Ma, H. Hänninen, S. Rossi, S. Piao, and Y. Bergeron. Spring phenology at different altitudes is becoming more uniform under global warming in Europe. *Global Change Biology*, 24(9):3969–3975, Sept. 2018. ISSN 13541013. doi: 10.1111/gcb.14288. URL https://onlinelibrary.wiley.com/doi/10.1111/gcb.14288.

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. ISSN 1992-0636. doi: 10.5194/asr-10-99-2013. URL https://asr.copernicus.org/articles/10/99/2013/.

R. D. Koster, R. H. Reichle, S. D. Schubert, and S. P. Mahanama. Length Scales of Hydrological Variability as Inferred from SMAP Soil Moisture Retrievals. *Journal of Hydrometeorology*, 20(11):2129–2146, Nov. 2019. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-19-0070.1. URL http://journals.ametsoc.org/doi/10.1175/JHM-D-19-0070.1.

M. Madruga de Brito, C. Kuhlicke, and A. Marx. Near-real-time drought impact assessment: a text mining approach on the 2018/19 drought in Germany. *Environmental Research Letters*, 15(10):1040a9, Oct. 2020. ISSN 1748-9326. doi: 10.1088/1748-9326/aba4ca. URL https://iopscience.iop.org/article/10.1088/1748-9326/aba4ca.

S. O, E. Dutra, and R. Orth. Robustness of Process-Based versus Data-Driven Modeling in Changing Climatic Conditions. *Journal of Hydrometeorology*, 21(9):1929–1944, Sept. 2020. ISSN 1525-755X, 1525-7541. doi: 10.1175/JHM-D-20-0072.1. URL https://journals.ametsoc.org/view/journals/hydr/21/9/jhmD200072.xml.

R. Orth, S. O, J. Zscheischler, M. D. Mahecha, and M. Reichstein. Contrasting biophysical and societal impacts of hydro-meteorological extremes. *Environmental Research Letters*, 17(1):014044, Jan. 2022. ISSN 1748-9326. doi: 10.1088/1748-9326/ac4139. URL https://iopscience.iop.org/article/10.1088/1748-9326/ac4139.

M. Peichl, S. Thober, V. Meyer, and L. Samaniego. The effect of soil moisture anomalies on maize yield in Germany. *Natural Hazards and Earth System Sciences*, 18(3):889–906, Mar. 2018. ISSN 1684-9981. doi: 10.5194/nhess-18-889-2018. URL https://nhess.copernicus.org/articles/18/889/2018/.

M. Peichl, S. Thober, L. Samaniego, B. Hansjürgens, and A. Marx. Machine-learning methods to assess the effects of a non-linear damage spectrum taking into account soil moisture on winter wheat yields in Germany. *Hydrology and Earth System Sciences*, 25(12):6523–6545, Dec. 2021. ISSN 1607-7938. doi: 10.5194/hess-25-6523-2021. URL https://hess.copernicus.org/articles/25/6523/2021/.

M. Svoboda, D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, and S. Stephens. THE DROUGHT MONITOR. *Bulletin of the American Meteorological Society*, 83(8):1181–1190, Aug. 2002. ISSN 0003-0007, 1520-0477. doi: 10.1175/1520-0477-83.8.1181. URL https://journals.ametsoc.org/view/journals/bams/83/8/1520-0477-83_8_1181.xml.

M. Zink, L. Samaniego, R. Kumar, S. Thober, J. Mai, D. Schäfer, and A. Marx. The German drought monitor. *Environmental Research Letters*, 11(7):074002, July 2016. ISSN 1748-9326. doi: 10.1088/1748-9326/11/7/074002. URL https://iopscience.iop.org/article/10.1088/1748-9326/11/7/074002.

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. ISSN 16077938. doi: 10.5194/hess-21-1769-2017. ISBN: 1607-7938.