

Changes made in reply to reviewer comments:

Below we detail the changes made to the manuscript in response to the reviewer comments:

General remarks:

1.) *“The applicability and purpose of the dataset is not entirely clear to me. The authors only briefly comment on this in the conclusions section, stating it could be used for exploring land-atmosphere feedbacks, investigating potential model simplifications, and data assimilation testing. Honestly, I do not see how the dataset can be useful in such analyses as (i) land-atmosphere interactions are known to be hardly robustly captured by models in general, and (ii) the testing of model simplifications and data assimilation would require different versions of the dataset with respective different model configurations in my opinion. In this context, I ask the authors to expand and clarify their discussion on the applicability of this dataset.”* -Reviewer 1

“ However, I still am left with the question of what is this data going to be used for beyond the work done by the authors. From my understanding, a publication in ESSD should be a dataset that one could expect to be used extensively by the larger community. If this covered the entire country or Europe then that would be a different story, but I just don’t see it as is. I still believe the work should be published but the paper would greatly benefit by putting it all in a greater context. I suggest the discussion provide a much more detailed overview of what this data could be used for and why it is necessary to have a unique dataset instead of just having each researcher rerun the simulations. In essence, it would be nice to know what this dataset could provide the larger scientific community 5-10 years from now.” -Reviewer 2

Both these raise similar points. We have changed the discussion section extensively, adding two paragraphs detailing further use of this dataset. For more details check our responses to the reviewers.

2.) *“CLM3.5 is used here as land surface model. This is outdated. By now, CLM5 is clearly more advanced in terms of simulating processes related to vegetation and hydrological dynamics at the land surface (Lawrence et al. 2019).”*

We added a paragraph where we explain that using a more up-to-date version would not have had an impact on this simulation due to our setup choices: “Version 3.5 of CLM that is used here is already relatively old. Even though version 5 was not yet available when we started our work, it is now and comparison is warranted. Newer versions of CLM have several major improvements over 3.5. The first one is a more sophisticated routing scheme leading to much improved soil moisture profiles. In our case we replace this part with ParFlow anyway so our older version is not a disadvantage in that regard. Other improvements are the inclusion of carbon and nitrogen cycles as well as more options for crop type vegetation. Here we purposely simplify our setup as we not only have and want static land use but also use a blend type of crop with no sharp changes in LAI due to harvests. Instead, we assume harvest to be an ongoing process all throughout autumn. Thus, all these improvements do not downgrade the simulation results presented and discussed in this study.”

3.) *“Throughout the manuscript there are various inaccurate statements limiting the reproducibility of the model simulation (e.g. ‘increased by about’, ‘increased by approximately’, ‘set considerably higher’,*

'needed to be increased from its standard values', various pedotransfer functions used without explaining criteria, see also respective specific comments below). More information needs to be provided in each of these cases to ensure the reproducibility of the entire analysis, either in the manuscript or in an appendix."

We have changed the appropriate sections to be specific. The details of each change are given in the section below. Since the forcing and namelist files are part of the dataset as supplementary files, the exact setup of our models can be re-created and values checked. We added this information with a link to the forcing files to the text.

4.) *"There are several arbitrary choices made throughout the study which need to be (better) motivated. This includes modifications of the modeling setup of which the purpose is not clear or the magnitude is arbitrary (e.g. 20% increase of sand fraction, ignoring of karst layers, conceptualization of alluvial layers as gravel and bedrock layers including the assumption(s) of values for various involved parameters, modifications of the LAI data). When making these modifications to adapt the model behavior in particular respects (more sandy soils to enhance infiltration) it should be kept in mind that even if the particular purpose is fulfilled, the land-atmosphere system is highly interconnected such that unforeseen side effects can occur. Further, the arbitrary choices include the approach(es) used to validate the modeled dataset (e.g. spatial averaging of model data across 25 grid cells for validation of atmospheric boundary layer characteristics, seemingly random time intervals of the soil moisture, evapotranspiration and runoff validations)."*

We have clarified our reasons as to why we changed some of the values. Most are based on other studies or previous results. Again, for more details as to the nature of these changes please check the detailed response we gave.

5.) *"Soil types are an important ingredient for hydro-climatological model simulations. The downscaling-based derivation of soil types in this study is (i) difficult to understand and (ii) contains several assumptions which are not motivated, among which are the amount of considered 1995 locations, the 20% increase of sand fraction (see above), the choice of an exponential model, the choice of conditional co-simulation versus kriging, the focus on first three soil horizons (first means uppermost I guess?). I wonder what is the impact of the choices made here on the final dataset?"*

I hope we have sufficiently answered this in our comment. We still have updated the related section to be clearer regarding the process and the impact of our choices. We have some more details below where they are mentioned again.

6.) *"The validation of the model simulation in terms of evapotranspiration is very limited. While it is reassuring that the ET and groundwater dynamics are broadly coupled according to expectations this is not a quantitative assessment. The modeled ET could instead (or additionally) be compared with state-of-the-art evapotranspiration datasets such as GLEAM (Martens et al. 2017) or FLUXCOM (Jung et al. 2019) at larger spatial scales."*

We have added an additional paragraph in detailing the reason for this very limited analysis. The various reasons are detailed in our response. The change is this: "We want to point out that in this region ET is almost always limited by atmospheric demand which is why we limit his analysis to bare-soil and evaporation only. Since the upper-most layers can dry quickly the resulting drop in evaporation can

be seen which is not the case for ET if there is an extended root zone as we have even for crop and grassland. These bare-soil areas are not a feature of the real catchment and as such cannot be compared to real measurements.”

7.) “I like the comprehensive validation of the dataset in terms of several variables - an overview table summarizing the determined strengths and weaknesses would be helpful for users I think.”

We have added such a table as table A.3.

8.) “There are too many figures in my opinion, diluting the main messages. Figures 3-5, 7, 14, 17 could be moved to supplementary, and Figures 9 & 10 could be combined.”

We have removed former figure 7 and moved former figures 2,4,5,14 and 17 to the appendix. In addition, we updated several of them in accordance with the more specific comments below.

Specific comments:

“line 35 and throughout: ‘simulated’ would be more straightforward than ‘virtual’, using such terminology the term ‘real’ (line 34 and throughout), referring to observations, can be removed from the manuscript”

We have used simulated instead of virtual throughout the text now.

“line 56: test a disaggregation method”

In the cited paper this was done for soil moisture. Doing this for this coupled system and all related variables and parameters would be a complete study on its own. The paper was cited to highlight the uses of simulated datasets rather as an example we wanted to follow.

“line 75-78: you do not aim to reproduce to observed catchment dynamics but still validate the model in some respects - this seems contradictory to me; what is the aim here if not validating the model against observations? how useful is a modeled dataset for the community if is not resembling observations?”

We have added a sentence in section 4 explaining why we validate our simulation even if we have purposely deviated from reality in some cases: “Even though we do not aim to be as close to reality as possible, we feel it important to show that the model system is behaving as expected and is thus suitable for the various use cases we discussed.”

“lines 139 & 145: the chosen time period and simulation catchment/area are not motivated”

We have added a sentence to detail that the time period limit was simply due to forcing availability. The choice of the catchment was arbitrary and it was only chosen because of size and because it fits in a roughly square computational domain: “... as 2007 was the first full year where high resolution atmospheric forcing was available and nine years was the maximum possible length with our granted compute time.”

“line 172: please give more information on the ‘software restriction’”

The restriction was somewhat unique with the system we were running on. We have added this information but it will likely not be useful any more as the system (and other like it) have been replaced over the last couple years and this restriction no longer applies.

“line 183: please give more information on the location of the grid cells and the artificial elevation modification to ensure reproducibility, here or in an appendix”

We have added a remark linking to the provided forcing data where the full elevation map is available: “All these changes are part of the forcing files that are provided with the full dataset making it easy to reproduce our simulations (https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=Neckar_VCS_v1_FORCING)”

“lines 195-196: ‘about 20%’ & ‘about 3.3’, please be more accurate”

We have clarified that 20% is an area average depending on PFT. The 3.3 change is due to the general LAI value of needle-leaf trees in that region which we clarified as well.

“lines 194-197: in the abstract of the Tian et al. 2004 paper I found “On average, the model [...] overestimates FPAR over most areas in the Northern Hemisphere compared to MODIS observations during all seasons except northern middle latitude summer.” “The MODIS LAI is generally consistent with the model during the snow-free periods...” which makes me wonder why the authors modify LAI in summer? Further this could create jumps in the LAI time series from May-June and August-September. More importantly, you state here that LAI is used “for the year 2008”. Does this mean there is no interannual vegetation variability? This would affect evapotranspiration and thereby many related variables and would need to be stated as a serious shortcoming.”

We have added several lines detailing that we do indeed ignore interannual variability on purpose and why (again see our detailed response): We also clarified that the changes in summer are small and related to the aggregation of shrubs and proper forests and that there are no jumps for LAI possible due to the way CLM treats LAI: “As a result, interannual variability is not considered in this simulation as we have the same LAI curve for each PFT each year. While this somewhat limits the comparability to ET observations in spring it is somewhat lessened by the fact that CLM does always use the values from two months and interpolates based on the date (for instance on the 1st of April the values from March and April would have almost equal weight, while on the 14th of April the April weight is clearly dominant).”

“line 204: please give the spatial resolution instead of the scale 1:1000000”

We added this. (~1km, not constant due to map projection)

“line 221: ‘approximately 20%’, please be more accurate; further, and more importantly, please motivate this modification and its magnitude”

We have added that is 20% except for cases where sand% was already very high to avoid unrealistic cases of >90% sand. We clarified that this is due to earlier simulation results and the resolution that we use: “...increased by 20% (for clay-rich areas that previously had practically no sand, less for areas that already had high sand content to avoid cases with >90% sand) resulting in a slightly higher hydraulic conductivity because previous simulations yielded too shallow unsaturated zones as was expected for the resolution of this simulation. The change to sand content was a direct response to this issue as it fixed most of the emerging biases”

“lines 229-231: please give detailed information on where which pedotransfer functions have been used”

We have added more detail, the text now reads: “The pedotransfer functions of Cosby et al. (1984) is used to estimate saturated hydraulic conductivity based on soil texture, the one from Rawls (1983) is used to estimate soil bulk density based on soil texture and organic matter and the one from Tóth et al. (2015) is used to estimate van Genuchten parameters based on soil texture and bulk density. These have been selected based on data availability, applicability of the particular approaches, and previous evaluations conducted in the area (Tietje and Hennings, 1996).”

“lines 239 & 242: repetition of the information that karst is not considered”

Mentioned it only the first time.

“lines 239-240: ‘to avoid the manifold hydrological challenges related to its modeling’, please be more specific here, also please comment on the impact of this simplification of the approach on the final dataset”

As mentioned in our detailed reply, karst is very hard to model correctly, especially with only 100m of depth of a model. Because of that we just disregard it. We added a sentence to make clear that while overall accuracy of groundwater suffers, effects on the near-surface soil moisture in these areas is limited: “While this can have significant impact on groundwater equilibrium and dynamics on the decade time scale, for the rather short time period considered here we would simply be left with a bias.”

“line 245: if these alluvial bodies are so relevant, why does this study use datasets which do not include them?”

The datasets we used are the most detailed ones for that region and even those do not include them for smaller rivers. They are very important for accurate subsurface flow which is why we added them. We made it clear in the text that they are not resolved in the dataset: “(not part of the soil and not large enough to be resolved in the geological map)”

“line 249: evapotranspiration errors in models can be significant and might be underestimated here”

We already assume a 30% error here and because we are energy limited this is very conservative: “...as in this climate we are almost always energy limited and therefore ET differences are reduced to LAI differences”

“section 4: please discuss for the performed validation analyses how the determined performance of the study dataset compares generally with the performance of other regional climate models in similar hydro-climatic regions”

As mentioned in our response, the goal of this study is not to be in line with climate models. Indeed, we show a more static picture, purposely ignoring climate change and accompanying effects to concentrate on compartment interaction instead. We added in the discussion the possibility for investigating this as a separate piece of work: “One of these topics is considering more transient changes (LAI and climate change) rather than the more static picture we present.”

“line 277: ‘simulated realistically’, please give more details here on how this is quantified”

Added more detail: “(timing, strength of wind gusts, change of wind direction, change in temperature and pressure)”

“line 301: 1km2, is this referring to the spatial resolution being 1km x 1km?”

It is 1x1km, made this clear in the text.

“line 323: ‘quite well’, please be more specific and objective. Further, in Figure 9a between 6h-17h the pattern in the model is actually opposite to that of the observations, I would not refer to this as fitting ‘quite well’.”

We have added a section explaining this in more detail and pointing out that afternoon rain is underestimated, which is normal with the model we use: “especially in late afternoon and night while it overestimates precipitation during the late morning while underestimating it in early afternoon in summer. In Winter this effect is much less pronounced but still there. It is connected to convective showers that are still too small and parametrized. These parametrizations were not originally designed for the km scale and have issues at this resolution and start producing precipitation too early.”

“line 331: the potential (dis)agreement of simulated and actual land cover can be checked using high-resolution land cover datasets such as provided by ESA CCI line 343: how are the temperature standard deviations determined?”

While the mismatch can be checked we still have to account for it when comparing results. We made this clear in the text. We also explained how the temperature deviations are calculated.

“line 361: ‘very well’, please be more specific and objective”

Changed wording and added more detail: “... are very close to observations during the day and at heights above 10m.”

“line 367: why not using the ESA CCI soil moisture dataset derived from observations of various satellites for this validation?”

Since we have different land-use and soil properties a comparison would not give any conclusive result.

“lines 394-395: I do not really understand why this daily matching is applied here? Also it is not clear how this is done.”

This means that for each day a factor is calculated and the actual result is corrected by this factor: “...and applied as a correction factor.”

“lines 391 & 396: I guess you are referring to Figure 15 here, not Figure 16 as stated.”

Figures have changed greatly, all figures should now be referenced correctly.

“line 415: ‘adequate agreement’, please be more specific and objective”

Changed wording: “... similar to the observations showing the same responses to rain events as well as similar behavior during dry periods, which is noteworthy...”

“line 422: ‘will always be replaced’ needs to be toned down in my opinion”

We now say “often” instead of “always”.

“line 440: ‘good distribution’, please be more specific and objective”

Changed line: "...reasonable split between shallower and deeper (5 meter and below) groundwater tables compared to expected values from observations with shallower levels overall."

"lines 447-450: I do not understand how the "fluctuations" are "scaled". Do you divide by the inter-annual standard deviation to obtain normalized anomalies (or z-scores)? If so, please name it this way as the term "fluctuations" is rather unclear."

In this case fluctuation merely means that the yearly average is removed for both time series: "(subtracting the yearly mean from the respective time series)"

"line 450: I guess this should be "according to Figure 19b" and not 19c?"

Again, figures have changed, should be correct now.

"lines 451-452: How is this trend computed?"

This was poor wording. We changed it to "fluctuations" to indicate that the same time-series as before is referenced.

"line 454 and following: I like this discussion of limitations and issues"

Thanks.

"line 458: to me it seems three challenges being discussed here (?)"

This was an error left over from a previous version. It is corrected.

"lines 502-512: As there are multiple concrete ideas to improve the model setup and consequently the dataset, why not implementing them before publishing this dataset? comment if this will be game-changers"

We have added a comment giving our opinion on this: "While these changes would show improvements, they are likely marginal or very specific (river discharge characteristics) and would therefore not warrant the great computational cost to re-run for such a long time. Future model developments of TerrSysMP may enable this option and it would be interesting to compare resulting datasets and quantify the simulation speed increase by using GPU compute technologies."

"Figure 2: This figure is the same as in the Gasper et al. 2014 paper, with the reference given in the caption. I think it is uncommon to use figures from previous papers, so I would remove this and only refer to the figure in the reference paper in the main text."

Moved to Appendix

"Figure 3: Maybe I missed that but what is domain 1?"

Again, left over from an older version, it is now simply the "simulated domain"

"Figure 4: "e+00" can be removed"

removed

"Figures 4-6: Please label the color bars."

done

“Figure 6: Please harmonize “evaporation” and “evapotranspiration” in the caption and the axis label. The same applies for the main text in section 4.1.”

We changed everything to “evaporation” in the text and figures.

“Figure 8: Values are quite far apart from color bars. Also, it would be nice to also express the difference as percentage.”

We removed the large spaces between values. A percentage difference would show almost the same picture, just with lower values in the mountains. We therefore left this out.

“Figure 11: y-axis label missing Please explain what is meant with the “temperature standard deviations””

Added label and explained standard deviations in the text: “(mean absolute difference between daily and monthly mean profile)”

“Figure 13: Please specify from which times the reference radiosonde observations are taken. Further, please explain how the standard deviation is derived.”

Same as last figure, also the times are in the figure already. Also added them to the caption.

“Figure 15: It would be insightful to quantify the agreement of the temporal dynamics with e.g. a correlation.”

This option is discussed in the text instead.

“Figures 16-18: Please use the station names throughout instead of the position numbers.”

Changed to station names.

“Figure 19: Panel a is not labelled, as well as color bar and axes therein The terms “model” and “reality” are not consistent with the terminology used throughout the manuscript.”

Changed the figure accordingly.