Dear Editor, Dear Reviewers,

Thank you very much for your helpful comments on our manuscript. Your suggestions and remarks have improved our manuscript considerably. We hope that you find your comments sufficiently considered and we would like to thank you for your support. Please also find below our detailed response to each reviewer comment (in red).

Yours sincerely,

Heye Bogena on behalf of the co-authors

Reviewer #1

General comments

The data paper is a great example of what can be achieved when disparate datasets are brought together and processed in a consistent manner. I like the use of normalised quantiles to facilitate widespread comparisons. The dataset will be a great resource for a variety of studies. I have no problems with recommending this paper for publication and only have a few minor comments which are outlined below.

We thank the reviewer for the positive evaluation and the comments.

I think the potential for a global version of the analysis and processing undertaken in the paper is exciting prospect.

We fully agree.

Specific comments

L43-44 – Unusual wording. Do you mean "For example, in the last decade Europe has suffered from drought events unlike any seen since the beginning of weather recording."

We corrected this.

L44 - cosmic-ray (not Cosmic-ray)

We corrected this.

L83 – "cosmogenic" in different font

We corrected this.

L86 - delete "prone"

We corrected this.

L118-119 – suggest "The Weisssee data only shows limited and short snow-free periods where soil moisture data is available…"

We agree and changed the sentence as suggested.

L248 – is something missing here are "More specifically, " ?

We corrected this.

L274 – it might be worth mentioning that the 24 hr moving window is good for drought studies but may obscure the signal from wetting events which typically occur much faster

We agree and added this information.

Table 1 – is it possible to get the column headings on each page

Table 2 – as for Table 1

We agree and added column headings on each page.

Fig 2, 3 and 4 – my versions look quite blocky – perhaps it is lower resolution that final will be

ALL Appendix figures – text in figures (labels/legends/etc) not crisp

All figures have been revised.

Reviewer #2

This is an exciting advancement to see a majority of the COSMOS-Europe data in one place. It will be a great resource for many disciplines to use. I don't have any additional comments beyond the technical revisions raised by the other reviewers. Excited to continue to see the European and Global COSMOS datasets come together for a great communal resource. Nice work!

We thank the reviewer for this very positive review.

Reviewer #3

General comments

I agree with the other reviewers in that the presented network will be a great unified source and a reference in soil moisture studies over Europe.

Thank you for the appreciation of the presented work.

As for the manuscript review, it is my impression that the processing method employed, which is overall well documented throughout, is based on well established practices and sufficiently corroborated using scientific sources. My comments here are mostly of form, or requirements of clarification. Please consider my suggestions as I think the experience of a reader will benefit from some improvements in the writing and transparency of the paper.

Thank you very much for your detailed comments, which will improve the paper. With very few exceptions, which are explained in more detail below, the change requests will be implemented.

In the analysis of drought events, the stress is placed on the use of the normalized quantiles, and my impression is that this was quite distracting from the actual aim of this section, which is the suitability of COSMOS-Europe to study droughts. Arguably other methods, for instance a simple calculation of the anomalies, could yield similar results by highlighting the soil moisture extremes. Have you considered for instance a comparison of the COSMOS-Europe anomalies with those of ERA5 (which is used elsewhere in the processing), or a correlation between the normalized quantiles and e.g. the Palmer Droughts Severity Index calculated from the ERA5 precipitation and temperature data sets?

We fully agree with the reviewer that the data set can be interpreted using more than the method used in the manuscript. Please note that due to the requirements of the ESSD manuscript type "data paper", we cannot expand on the data analysis. The presentable analysis is limited to showing the potential of the provided data and does not allow for further comparisons with other datasets. However, we will be happy to include references to the suggested methods in the revision.

Specific comments

Abstract section

The abstract starts by stating the impact of droughts in Europe; however, drought analysis is rather a possible application of the data, which you demonstrate in your study. The network can serve a vast array of other applications, which you also mention further on (e.g. EO data validation, model assimilation). Please consider rephrasing the abstract to place the stress on the novelty (i.e, harmonized processing) and characteristics of this data set, rather than on the analysis you carry out in the discussion.

Thank you for the comment. We agree that the COSMOS-Europe-Network data can be of great use for various applications. On the other hand, we also think that in particular the harmonized processing is of great use for the analysis of climatic extremes of continental relevance. We have extended the corresponding description in the introduction accordingly:

LL54-58: "The data of the presented COSMOS-Europe network and the harmonized processing framework open up a manifold of potential applications for environmental research, such as remote sensing data validation, trend analysis, or model assimilation. The data set could be of particular importance for the analysis of extreme climatic events at the continental scale. Due its timely relevance in the scope of climate change in the recent years, we demonstrate this potential application with a brief analysis on the spatiotemporal soil moisture variability."

We will also shift the focus more towards the novelty of this work at the beginning of the abstract, while keeping climate change driven drought monitoring as one of the main motivations for this dataset:

LL44-47: "Climate change increases the occurrence and severity of droughts due to increasing temperatures, altered circulation patterns and reduced snow occurrence. While Europe has suffered from drought events in the last decade unlike ever seen since the beginning of weather recordings, harmonized long-term datasets across the continent are needed to monitor change and support predictions."

Methods section

Lines 139-143 seem to me more of a general level introduction of the processing. I would state this under 3. Methods rater than 3.1 Data pre-processing, as it also regards 3.2, 3.3, …

Thank you. We changed this accordingly.

On line 161, can you please provide the rationale behind the assumptions?

Thanks for this justified question. In fact, the relevant wording in the text was misleading. We have revised the wording accordingly.

LL165-166: "Missing data on soil texture, porosity and organic carbon where taken from the global raster-based soil dataset SoilGrids (Hengl et al. 2017)."

How are the calibration errors affecting the uncertainty estimate (N0 in line 289), how is it dealt by in the cited reference? For instance, what is the impact of the regression with ERA5 (Figure A1)? Often it seems like the residuals are quite large even for R squared larger than 0.7, is this effect negligible?

Iwema et al. (2015) found that a higher number of calibration dates could provide higher confidence for the N0 estimate. However, to our knowledge, there is no study that takes into account the uncertainty of N0. We tried to transparently display the number of calibration days in Fig. 2 and Fig. A3, such that the reliability of the N0 parameters can be individually assessed. Regarding ERA5 data, we agree that its use to fill data gaps could introduce additional uncertainties due to the deviation from local conditions. However, this only applies to the times when ERA5 was used instead of local data. As the number of gaps varies greatly between the CRNS sites, general uncertainty statements are difficult to make. However, we added a statement on additional uncertainties involved in the process of data gap filling with ERA5 data:

LL298-299: "Please note that additional uncertainties may have occurred when filling gaps in the air pressure and humidity data with ERA5 data."

L298 - The use of normalized quantiles is quite relevant in your analysis. Do you plan to distribute this data with the network data or in the publication?

A complete list of the data provided can be found in A5. The normalized quantiles are not part of the dataset as they are intended as an example of data use. Since we provide the necessary calculations, the user can easily calculate the normalized quantiles himself.

L328 – I imagine that a reader (or user) will be interested to know how the proposed data set varies from the 'outdated COSMOS scheme' (L252), even considering that some station data are distributed in other places than your data repository (e.g., the Rietholzbach station can also be found here: http://cosmos.hwr.arizona.edu/), presumably with a separate processing scheme. What differences can be expected between these?

The basis of our processing methods leans heavily on the commonly used method presented by Zreda et al. (2012). However, there are some important differences to the COSMOS-HWR-Arizona database dated in 2012, since we follow latest insights on CRNS research from the last 10 years. For example, the Arizona database does not contain corrections for air humidity, which were suggested by Rosolem et al. (2013) and Köhli et al (2020). A main differences lies also in the weighting of the in situ calibration data according to neutron transport theory (Schrön et al. 2017). We therefore expect soil moisture deviations that can be substantial depending on the date of previous processing schemes. We now mention this aspect more prominently in the revised manuscript:

LL408-415: "Finally, we would like to point out that there are still alternative data products to some of the CRNS stations used here, which were processed using less comprehensive methods. Although our processing methods closely follow the commonly used method presented by Zreda et al. (2012), there are some important differences as we make use of the latest CRNS research findings from the last 10 years. For example, the US COSMOS database (http://cosmos.hwr.arizona.edu/) does not contain corrections for air humidity, which were suggested by Rosolem et al. (2013) and Köhli et al (2020). Another main difference of our processing scheme is the weighting of the in-situ calibration data according to neutron transport

theory (Schrön et al. 2017). We therefore expect soil moisture deviations that can be substantial depending on the date of previous processing schemes."

Results and Discussion section

L360 - A more robust trend analysis could be used here - for instance, how well can it be differentiated between the trends of individual stations (e.g. based on a Mann-Kendall test)? Would we see patterns as aggregated by e.g. climate zone? The previous analysis was based on individual stations and climate classes, therefore it would perhaps be more consistent (and insightful) to use this approach still.

As argued above, data analysis is very limited as ESSD focusses on the publication of research data and any in-depth interpretation of data is outside the scope of regular articles (https://www.earth-system-science-data.net/about/aims_and_scope.html). Therefore, our analysis can only be a starting point and we hope that there will be many users of the data who will do a more in-depth data analysis.

L390 - "This indicates that the COSMOS-Europe data could be beneficial for model applications at the continental scale despite the limited coverage in some areas of Europe". This is more of a speculation than a logic conclusion based on the result, or is in any case general applicable to any in-situ soil moisture network. What differs from the CRNS data in Baatz et al. (2017) and here?

In the study by Baatz et al. (2017), individual CRNS stations were used in a mesoscale catchment to improve an area-differentiated land surface model by means of data assimilation. Similarly, the COSMOS-Europe data could be used to update a continental land surface model even though the measurements are not area-wide (e.g. as remote sensing data). We have added a better argumentation in the revised version:

LL396-400: "For example, Baatz et al. (2017) assimilated measured soil moisture data from a CRNS network into the area-differentiated land surface model CLM 4.5 (Oleson et al., 2013) and showed that updating states and hydraulic parameters leads to better regional hydrologic predictions. This indicates that the COSMOS-Europe data could be beneficial for model applications at the continental scale despite the limited coverage in some areas of Europe, i.e. even though the measurements are not area-wide such as remote sensing data."

Figures and Appendix

Could you please provide a better resolution of the base map in Figure 1?

Yes.

In the lower sub plots of Figure 4, some of the soil moisture absolute values exceed 0.6 for several sensors, while looking at the time series above, this does not seem to happen. What is the reason for this?

This is because the plot above shows monthly data, while the lower ones show daily data.

Figure A3: please add the x-axis label in the plot. Otherwise, the image is complete and provides a very good simple graphic illustration.

Thank you very much. We added the x-axis label.

Suggestion: In Figure A4, do the colors refer to the same date across different subplots? If so, could be worth adding the sampling date associated to the colors.

Since some sites have more than 20 sampling dates this suggestion is difficult to realize. However, the sampling dates are already indicated in Fig. 2.

Figure A5: do you plan to distribute the uncertainty estimates with the soil moisture time series? I would expect to see it listed somewhere here under b).

The uncertainty estimates are part of the dataset and already indicated by the suffix "_std" in Figure A5, e.g. "NeutronCount_Epithermal_Cum1h_corrected_std" or "SoilMoisture_volumetric_MovAvg24h_std".

Technical corrections

L49 – Repetition of "information". Proposed rephrasing: "in addition, the uncertainty estimate is provided with the dataset, information that is particularly useful for remote sensing and modelling applications."

We changed this accordingly.

L58 – Summers are not years – I proposed the phrasing: "The years … are considered the most notable of the 21st century in terms of summer drought .."

We changed this accordingly.

L59 – The use of "but" suggests a contradiction, which is not present.

We corrected this.

L67 – In the sentence "…, with most of the ecosystem response occurring indirectly as a feedback between soil moisture and the atmosphere, …" it is not clear how the ecosystem is related to the soil moisture-atmosphere interaction, or how it amplifies anomalies (in the following sentence). Please simplify/clarify.

We clarified this section:

LL70-72: "Recently, Humphrey et al. (2021) have shown that soil moisture variability explains 90% of the interannual variability in global carbon uptake. The corresponding feedback between soil moisture and the atmosphere amplifies temperature and moisture anomalies and intensifies the direct effects of drought and soil water stress."

L100 – Suggestion: the text would benefit here by the use of an impersonal form, e.g. "… and how the data is processed in a harmonized way."

We changed this accordingly.

L108 – Suggestion: "The key environmental and soil-related physical properties at the sites …"

We changed this accordingly.

L115 – Where is the information on the land cover types originated?

We will added the following:

LL116: "According to site owner information, … "

L141 – The SM signal is not extracted, rather modelled. I would rephrase to "… accurate modelling of the soil moisture signal from ...'"

We do not think that modelling is the right term. In fact, extraction of the SM signal fits better because we apply several corrections to the CRNS signal that can be denoted as filters.

L143 – (Also later in the document) URLs should be cited according to the APA rules. See e.g.

https://www.lib.sfu.ca/help/cite-write/citation-style-guides/apa/websites

Thank you, we will change this accordingly.

L144 – Aggregated how? Resampled, averaged, …?

We will add this information (averaged).

L146 – "Reduce the measurement uncertainty" is not entirely correct; you can say that the relative uncertainty of the aggregated estimate is smaller, but the uncertainties of the individual measurements remain the same

We will change this accordingly.

L147 - How does it improve consistency? In my opinion, the sentence is not complete/correct as data quality is different from data consistency (which refers to something specific, e.g. temporal consistency, …)

We think that improving data consistency also includes the elimination of outliers in a harmonized and consistent way.

L154 - ERA5 has a specific citation guideline, please make sure it is respected:

https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation

We added the following sentence to the acknowledgements: "The ERA5-Land data are provided by ECMWF (Muñoz Sabater, 2021)."

And we added the following citation:

LL620-621: "Muñoz Sabater, J.: ERA5-Land hourly data from 1950 to 1980, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.e2161bac, 2021."

L221 – Missing symbol "and [] is". This happens also later on in the text, please check the mathematical symbols rendering

We corrected this.

L229 – Where is the biomass information originated (data set)?

Biomass information is not currently used, with the exception of the Wuestbach1 site, for which site owner data is used.

L248 – Missing symbol "More specifically, [] is integrated"

We corrected this.

L280 – Missing symbol "the propagated uncertainty [] is highly ..."

We corrected this.

L286 – Missing symbol ", [] is its Gaussian uncertainty, ..."

We corrected this.

L319 – is “TEreno” supposed to be with uppercase “E”?

Yes.

L321 – can you provide a reference to the NodeRed documentation? As it is mentioned as a processing tool.

We added a citation for the Node-RED documentation:

L628: “Node-RED. Node-RED documentation, https://nodered.org/docs/user-guide/. (2021, December)”

L356 – Please change “these occurred predominantly in climate zone Cfb,” to “these occurred predominantly in the climate zone Cfb,”

We corrected this.

L560 – Reference title has a mistake: “Error estimation with for soil moisture…” should be “Error estimation for soil moisture…”. Please check all references used and correctness of citation.

We will correct this.