Comment on essd-2021-325
Anonymous Referee #3

Referee comment on "COSMOS-Europe: a European network of cosmic-ray neutron soil moisture sensors" by Heye Reemt Bogena et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-325-RC3, 2021

COSMOS-Europe: A European Network of Cosmic-Ray Neutron Soil Moisture Sensors. Bogena et al, 2021.

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.

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.

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?
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.

Methods section

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

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

How are the calibration errors affecting the uncertainty estimate (\(N_0\) 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?

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?

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?

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.

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?

Figures and Appendix

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

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?

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

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.

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).

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.”

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 ..”

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

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.

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.”

L108 – Suggestion: “The key environmental and soil-related physical properties at the sites ...”

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

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

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

L144 – Aggregated how? Resampled, 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

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

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

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

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

L248 – Missing symbol “More specifically, [] is integrated”

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

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

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

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

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

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.
