

Review on essd-2022-90 by van der Velde and colleagues

I would like to thank the authors for addressing and responding to the comments.

I still have some comments which should be addressed before publication. Many of these comments apply to the new sections 5.2 and 6 which are interesting, but still a bit unwieldy. I would like to ask the authors to carefully go through these sections once more. I understand that with my comments in the interactive discussion, I motivated the authors to put more emphasis on these parts, and I appreciate the reaction. But as these sections are more oriented towards scientific interpretation (instead of the description of the dataset), we have to be more careful and rigorous with regard to formulating hypotheses. I also found one or two issues that were already present in the original preprint, but which slipped my attention. I apologise, but would still ask to address these issues as well.

II. 18-20: Without reading section 5.2, these lines cannot be understood: How can the spatial representativeness be measured by the R^2 or RMSE? What is meant by network scale? Please find a more concise way to summarise your findings on representativeness in the abstract.

I. 21: VSM - acronym not explained in the abstract.

II. 18-25: Overall, I find these newly added lines difficult to read. Please try to make this more concise.

II. 147-148: "[...] while typically less than 50 mm were recorded per day." Given that the most extreme daily rainfall depths were reported as 50, 142 and 106 mm, it is pretty obvious that the other days had less rainfall. Hence, this fragment does not bear any information. Please delete.

I. 156 should read "In the site selection, care was taken to evenly distribute the SENSOR LOCATIONS across [...]"

I. 173: should be "section 7", now, I suppose.

I. 180: Why "soil layer" instead of just "soil"?

I. 183-184: Where can I find the information which locations have a limited coverage of measurement depths? Should this information go into Tab. S2?

II. 211-213: I suggest using standard terminology to refer to this procedure (leave-one-out cross validation).

I. 227: I suggest using "surface soil moisture" and also label the section "Field campaigns to observe surface soil moisture", so that it becomes clearer to the reader that this is not about SWC profiles.

I. 289-290: "agreement difference" sounds weird. I suggest to replace the entire sentence "Factors that could have contributed to this agreement difference are the deployed instruments [...]" by "This could be explained by the deployed instruments, [...]"

Section 5.2: In the beginning of this section, you should again highlight that any of the following analysis only refers to the agreement at the upper 5 cm. It does not tell us anything about what's happening below (in terms of representativeness).

I. 336: you replaced "representativeness for the field" by "representativeness of the field" which is not correct, in my opinion.

II. 337-339: the factors you mention here apply to most soil moisture measurements. What is most important, in my view, is that the soil management between the fields is different from in the field, namely that the fields are usually ploughed and harrowed while the stripes inbetween remain undisturbed. This might have fundamental implications for soil hydraulic properties in the upper 30 cm. Please discuss this briefly, if you agree.

I. 347: "which can be attributed to edge effects": this is just a hypothesis, so I suggest not to make the statement that absolute.

II. 349-350: not only higher interception losses, but also higher transpiration, wouldn't you agree?

I. 351: "majority" - why so unspecific? Couldn't you just state the number of profiles which fall into that range?

I. 361: "this may be argued for" - please rephrase

I. 364: I would not use "performance", but rather "agreement"

I. 365: "antecedent precipitation" - antecedent over which period before the campaigns?

I. 366: But did you systematically sample, within the field, in local depressions? Otherwise, this would not explain the systematic underestimation, right?

I. 377: "inflated" is not an adequate term, here. Use "high" or "large" instead (if you in fact think it is large).

Fig. 8 and section 6.1:

- I appreciate the motivation to combine the campaign measurements with the continuous measurements. Still, I am having some difficulties to understand the figure and its purpose. The red lines represent the "network", so all soil moisture profiles in Twente? But averaged over all depths? Or just at the surface (upper 5 cm)? And the markers represent any profile/field for which a campaign was carried out on a given day? This needs better explanation.

- Apart from comprehensibility, what can we actually learn from contrasting the means of selected subsets of the data with the overall network mean? You state that the figure reflects the network's "overall performance" - what is meant by that?
- On what basis do you state that the campaigns measurements match the station measurements "very well" (l. 398).
- ll. 403 ff.: I am quite hesitant about the presented concept of "temporal representativeness": "[we] found that the least differences between the values measured during the field campaigns and stations' data records do not necessarily occur at the same time of measurement." To be honest, I do not understand what is implied here and which "physical processes" you refer to. I am not doubting the stated fact, but I am wondering about any explanation beyond "random effect". Please elaborate.
- Technical remarks: (i) do not use filled markers, but wider edgelines for the markers instead. (ii) And which precipitation observations are shown on the secondary axis? Or is this an average of all rain gauges? If yes, how is it averaged/weighted? (iii) The first legend should use four columns and one row instead of one column and four rows.

Fig. 9: In order to adequately interpret the figure, precipitation and air temperature need to be shown, too, on secondary axes and/or an additional panel.

Fig. 10:

- To better understand the effect atmospheric drivers, I usually find it helpful to display the cumulative sum of the daily difference between precipitation and reference evapotranspiration. That way, you can typically see a clear relationship between increasing and decreasing parts of that curve and the drying and wetting of the topsoil. This is just a suggestion, since showing daily air temperature and precipitation for such long time series is difficult to interpret.
- How can the volumetric SWC be higher than $0.6 \text{ m}^3/\text{m}^3$, even close to $0.8 \text{ m}^3/\text{m}^3$ on a location with sand / highly loamy sand (ITC_SM14, see Tab. S2, and also ITC_SM17). I find this quite spurious.

ll. 430: Instead of "Specifically in the 80 cm soil moisture content [...]" better "Specifically at a depth of 80 cm, soil moisture content [...]"

l. 434: replace "measurement" by "level" and "increments" by "increases"

ll. 435-438: In my view, care needs to be taken with such correlations. I understand that this is just a data description paper, so that in-depth analyses are unwarranted. Yet, when correlating SWC time series with the groundwater level, the delays between the signals, corresponding to the travel of the water from the soil down to the aquifer, should be taken into account. At least for many soils without predominant bypass flow, there should be such a clear delay. Before correlating the time series in order to identify which SWC signal best explains groundwater level dynamics, the delay should be accounted for by shifting the series in time, finding the shift which yields the maximum correlation. As I guess such an analysis is taking things too far for this paper, the authors might consider removing the table and the corresponding text fragments.

II. 450-452: It is unclear what the authors mean by "[...] make it possible to address sub-catchment scale applications."

II. 498: You use the term "network scale" which I find insufficiently defined. Maybe rather "for the entire network" if you refer to specific statistical metrics.

II. 502: To what does the "hence" refer?

II. 505: "network scale" - see above.

II. 505-508: How can your dataset be valuable for upcoming (future) SAR missions if it only spans until 2020??

Please provide all figures (except 1 and 3) in vector format in order to allow for lossless zooming.

Check for consistent use of tense throughout the manuscript, specifically in the newly added parts.