Dear Editor(s) of the Earth System Science Data (ESSD) journal,

Dear Dr. Sibylle Hassler,

We would like to thank you for considering our manuscript 'Twelve years profile soil moisture and temperature measurements in Twente, The Netherlands' (essd-2022-90) for publication in the well-established ESSD journal and offering for the second time the opportunity to revise and resubmit it. Included in this submission is a clean copy of the revised manuscript, a marked-up version of the manuscript showing the changes made and detailed point-by-point responses to the comments of referee 1.
We have carefully considered the comments, questions, and suggestions for improvement made by referee 1 and applied them where we found appropriate. The major changes can be summarized as follows:

- Textual improvements have been made throughout the manuscript.
- The text with respect to analysis of the spatial representativeness has been updated to emphasize that the representativeness of the top 5 cm soil moisture is discussed.
- The second paragraph of the abstract has been rewritten to make better readable standalone.
- Figures 8, 9 and 10 have been updated according to the referee's suggestions.

The comments of the referee 1 have again helped us to improve the overall quality of the manuscript, for which we are grateful.

Overall we trust that we have addressed all the referee comments, and have further improved the quality of the manuscript to the level expected from publications in the ESSD journal. We would like to thank you for taking the time to handle this contribution and look forward to hearing from you.

Yours truly,

Rogier van der Velde
On behalf of the authors.

Authors' response to Referee 1

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.

Authors' response:
We thank the referee for the positive and detailed comments. Again the level of detail and provided suggestions are much appreciated. In the text below, we provide our point-by-point response to the comments.

The referee's suggestions have been implemented in the revised manuscript. Our response to the referee's comments is structured as follows:

- The native referee comment is labelled and written in black.
- The authors response is written in blue
- Text from the manuscript is written with the Times New Roman font whereby the native text is in black and the changes are in red.
- The line numbers refer to the clean copy of the revised manuscript.

Comment R1C1
ll. 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.

Authors' response:
We removed the reference to specific R^2 and RMSE values, avoided the use of the words 'network-scale' and summarise the main findings more precise and in a more concise manner.

Comment R1C2
l. 21: VSM - acronym not explained in the abstract.

Authors' response:
We replaced VSM by field averaged soil moisture content

Comment R1C3
ll. 18-25: Overall, I find these newly added lines difficult to read. Please try to make this more concise.

Authors' response:
We have rewritten the text as follows,

L18-24:
An indication for the spatial representativeness of the permanent monitoring stations is provided through comparisons of the 5 cm station measurements with the top 5 cm field averaged soil moisture content derived from the field campaign measurements. The results reveal in general reasonable agreements and root mean squared errors that are dominated by underestimations of the field averaged soil moisture content, which is particularly apparent for the grass fields and strong after heavy rain. Further, we discuss the prospects the datasets offer to investigate i) the reliability of soil moisture references that serve the development and validation of soil moisture products, and ii) the water and energy exchanges across the groundwater-vadose zone – atmosphere continuum within a lowland environment in a changing climate.

Comment R1C4
ll. 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, his fragment does not bear any information. Please delete.

Authors' response:
done

Comment R1C5
l. 156 should read "In the site selection, care was taken to evenly distribute the SENSOR LOCATIONS across [...]"

Authors' response:
We changed this to monitoring locations to be consistent with the used terminology.

Comment R1C6
l. 173: should be "section 7", now, I suppose.

Authors' response:
done

Comment R1C6
l. 180: Why "soil layer" instead of just "soil"?

Authors' response:
We changed this to soil-water-air mixture

Comment R1C7
l. 183-184: Where can I find the information which locations have a limited coverage of measurement depths? Should this information go into Tab. S2?

Authors' response:
We have added this information to a newly created Table S3 and added to the main body of the manuscript the following:

L182:
Table S3 provides for each station the installed sensor types and installation depths.

Comment R1C8

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

Authors' response:
The two sentences have been replaced by

L210-211:
The leave-one-out cross-validation procedure is adopted for calculating the performance metrics because of the limited sample size and to provide an uncertainty estimate for coefficients *a* and *b*.

Comment R1C9
l. 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.

Authors' response:
Done. We also mention in the abstract that the top 5 cm soil moisture content is measured during the field campaigns, and swapped and rewrote the second sentence at the start of section 4 with the first sentence of section 4.1 as follows:

L225-227:
The objective of the campaigns was the validation soil moisture retrievals from satellite observations via estimates of the spatially aggregated top 5 cm soil moisture content, hereafter referred to as surface soil moisture.

L234:
Sampling took place at up to five fields near a monitoring station with in total of 28 sampled fields near 12 monitoring stations.

Comment R1C10
l. 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, [....]"

Authors' response:
done

Comment R1C11
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).

Authors' response:
We renamed section 5.2:
Spatial representativeness of observed surface soil moisture.

Rewritten the sentence around L331-332 as follows:
Field averages derived from the surface soil moisture measurements collected during the campaigns (see section 4) have been used to assess this issue.

And refer to 'surface soil moisture' in various parts of this section.

Comment R1C12
l. 336: you replaced "representativeness for the field" by "representativeness of the field" which is not correct, in my opinion.

Authors' response:
We have changed this to, its representativeness for the field.

Comment R1C13
ll. 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 in between remain undisturbed. This might have fundamental implications for soil hydraulic properties in the upper 30 cm. Please discuss this briefly, if you agree.

Authors' response:
To address this issue, we have modified the sentence as follows,

L328-332:
Large differences in the meteorological inputs, e.g. precipitation and incoming solar radiation, are not expected, but small-scale topography, spatially variable soil texture, different land covers and degrees of soil compaction as a result of agricultural management practices, and field-specific drainage infrastructure may cause discrepancies between the VSM at the border and inside of the field.

Comment R1C14
l. 347: "which can be attributed to edge effects": this is just a hypothesis, so I suggest not to make the statement that absolute.

Authors' response:
We have changed the word 'can' to 'may '.

Comment R1C15
ll. 349-350: not only higher interception losses, but also higher transpiration, wouldn't you agree?

Authors' response:
We agree and changed this to 'transpiration and interception of precipitation'

Comment R1C16
l. 351: "majority" - why so unspecific? Couldn't you just state the number of profiles which fall into that range?

Authors' response:
We have modified the sentence as follows,

L341-342:
The R^2 values range for six out of the nine stations from 0.516 to 0.793, while R^2 values of 0.36 and 0.38 suggest that stations ITCSM_05 and ITCSM_18 are less representative of the fields.

Comment R1C17
l. 361: "this may be argued for" - please rephrase

Authors' response:

We have replaced this may be argued for by this may be explained by

Comment R1C18
l. 364: I would not use "performance", but rather "agreement"

Authors' response:
done

Comment R1C19
l. 365: "antecedent precipitation" - antecedent over which period before the campaigns?

Authors' response:
We have modified the sentence as follows,

L352-354:
Further analysis shows that a large part of this weaker agreement stems from two days (19 October 2016 and 28 June 2017) with exceptionally large mismatches, which have in common that on average more than 27 mm of precipitation was recorded in total on the day itself and the day before.

As the text states on average more the 27 mm of precipitation was recorded on average by summing the volume of the day itself and the before. The 27 mm is derived as the average of those sums recorded by the three automated weather stations. The temporal precipitation distribution are a little bit different for the two days. Prior to 18/19 October 2016 it was relatively dry and the majority of the rain fell on October 19. We took the field measurements during/after a down poor. Prior to 27/28 June 2017 it had rained already on five consecutive days and on 27/28 extra rain fell.

For the sake of brevity, we did not include this detailed explanation because we find that this information does not add to the narrative.

Comment R1C20
l. 366: But did you systematically sample, within the field, in local depressions? Otherwise, this would not explain the systematic underestimation, right?

Authors' response:

It is true that the Netherlands is flat but it is not as flat as a pancake as people may think. We sampled during those day and I (the first author) remember very well that parts of the field were completely submerged whereas other parts remained clear of standing water.

To clarify this, we have modified the sentence as follows,

L356-358:
The large rain volumes on those days most likely led to overland flow that accumulates in local depressions and led on those days to the partial flooding of fields as a result of small scale topography, whereas the instrumentation at ITCSM_10 was installed in a slightly higher and, therefore, drier part of the field.

Comment R1C21
l. 377: "inflated" is not an adequate term, here. Use "high" or "large" instead (if you in fact think it is large).

Authors' response:
We replaced 'inflated' by 'large'.

Comment R1C22
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?

Authors' response:
The purpose of the figure is to evaluate how the biases for individual fields propagate when aggregated over a number of fields. This provides a best estimate for the bias of the entire network since we have not been in the position to sample near all our monitoring stations on any day in a systematic manner. The mean and standard deviation of the 5 cm VSM of the entire network are shown in figure 8 to illustrate how representative the sampled fields are for the mean 5 cm VSM measured by the entire network. The text has been rewritten as follows,

L360-364:
The metrics labelled 'sampling day' are based on matchups between the mean values of all field-averaged surface soil moisture and corresponding 5 cm station VSM measurements collected on a specific day. They show how the biases found for individual fields propagate when aggregated over a number of fields and provide an indication for the bias of the entire network. In support, Fig. 8 shows the mean of the field-averaged surface soil moisture and the matching mean of the 5 cm station VSM for the years 2015, 2016 and 2017 along with the mean 5 cm VSM of the entire network plus and minus the standard deviation.

- On what basis do you state that the campaigns measurements match the station measurements "very well" (l. 398).

Authors' response:

We make this statement based on the R^2 value of 0.770 reported in section 5.2. In analogy with the terminology used in that section, we scale 'very well' down to 'fairly well' and clarified the context by rewriting the text as follows,

L390-394:

Figure 8 shows that the mean values of all field averaged surface soil moisture and corresponding 5 cm station VSM measurements collected on a specific campaign day match fairly well with each other given a R^2 of 0.770 as well as the mean of 5 cm station VSM of the entire network (solid lines). These results provide an indication for the bias of the entire network, but the results presented in section 5.2 also demonstrate that further investigations should address the effect of spatial heterogeneity at field-scale.

- 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.

Authors' response:

We agree that the term 'temporal representativeness' is not appropriate. We refer here to a mismatch between the station VSM measured with probes installed at a depth of 5 cm and the top 5 cm soil moisture (surface soil moisture) measured during the field campaigns through soil sampling and with impedance probes.

We have rewritten the text as follows,

L396-401:

In addition, the presented data enables research into the representativeness of station VSM measured with probes installed at a depth of 5 cm for the top 5 cm soil moisture measured during campaigns that are typically considered as reference in validation studies. We carried out a preliminary analysis and found that the best match between the surface soil moisture measured during the field campaigns is found with the 5 cm station VSM that is recorded several hours up to two days later. The presented datasets provide an opportunity to investigate this and the physical processes that affect the near surface soil moisture profile, in particular infiltration and evaporation.

- 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.

Authors' response:

The figure has been modified according to the reviewer's suggestions and the caption is updated to explain the precipitation data source:

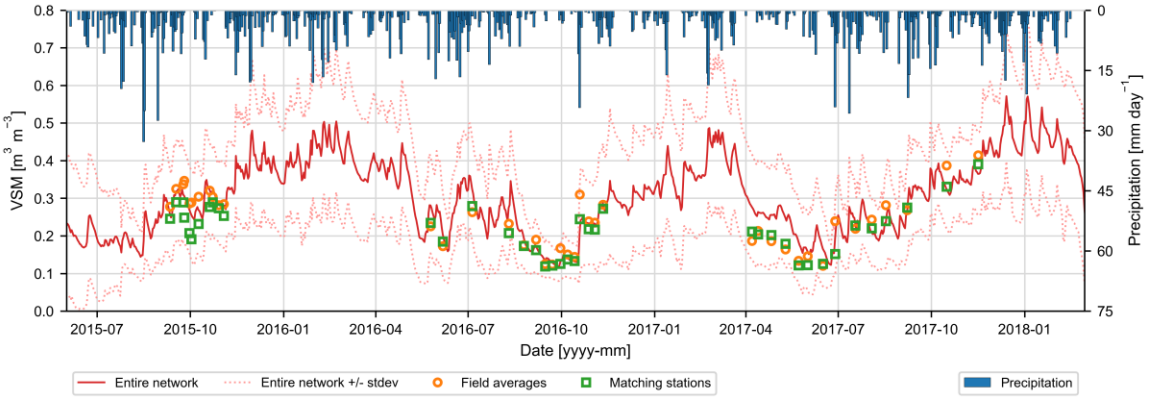


Figure 8: Mean values of field-averaged surface soil moisture measured during the 2015, 2016, and 2017 field campaigns (marker: circle) and of 5 cm VSM measured at the matching monitoring stations (marker: square). The solid and dotted lines represent the mean 5 cm VSM of the entire network +/- the standard deviation. The precipitation shown on the secondary axis is derived as the arithmetic mean from the data collected by the three KNMI AWSs.

Comment R1C23

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.

Authors' response:

We update Figure 9 by adding a new panel depicting hourly values of air temperature and precipitation measured at the KNMI AWS Twenthe airport

This figure displays the nexus between precipitation-soil moisture and temperature. For example, the rainfall events spread over 11 and 13 July had a peak intensity of 3 mm h^{-1} at an average air temperature of 16.5 °C and showed an effect on the 5, 10 and 20 cm soil moisture content. The following days a smaller amplitude in the diurnal air and soil temperature cycle can be noted and the soil dried out gradually. The rainfall event of 20 July had a lower peak intensity of 2 mm h^{-1} and fell when the air temperature was higher, on average 19.9 °C, and did not led to a change in the soil moisture contents

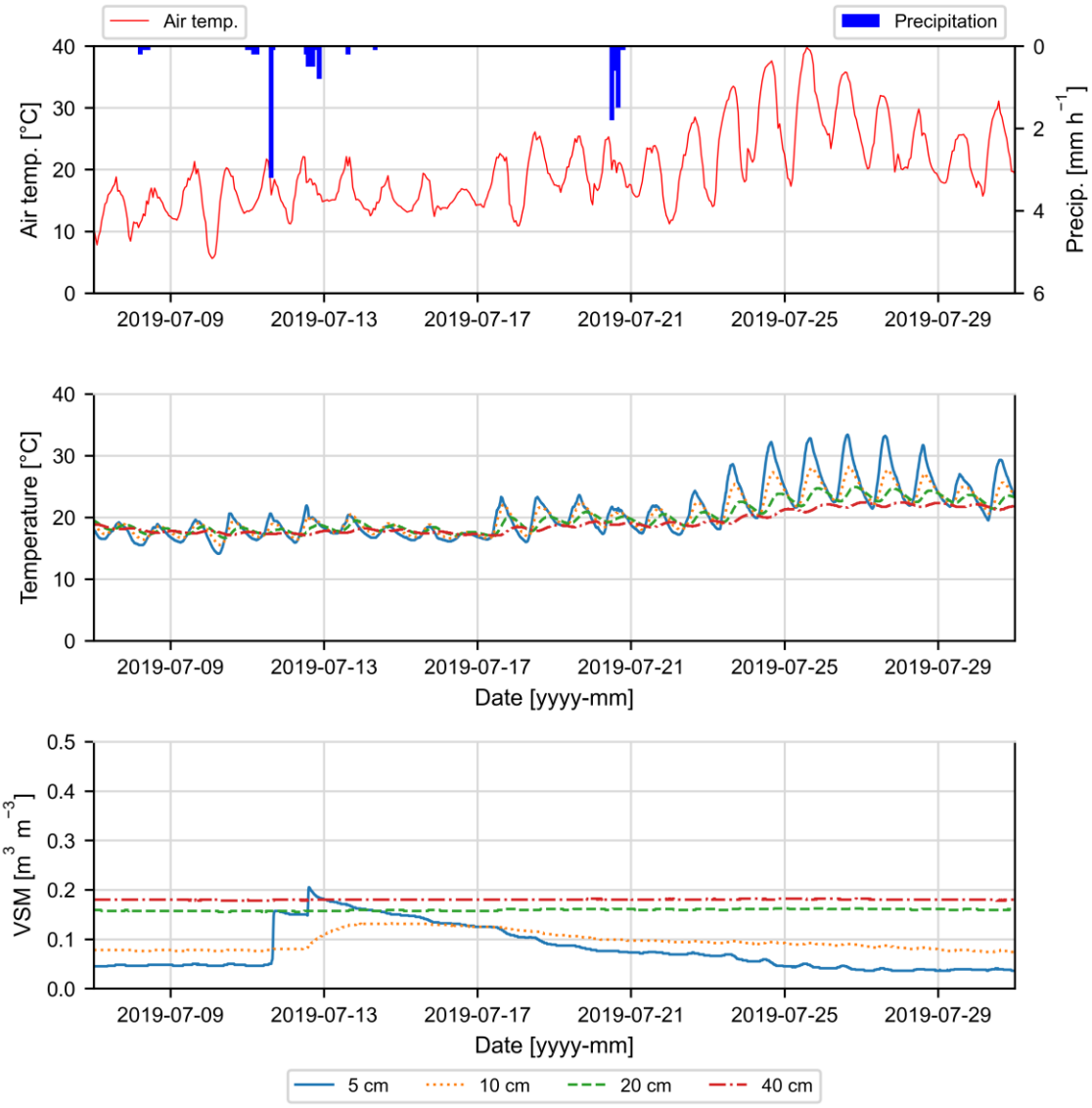


Figure 9: Soil moisture and temperature depth profiles measured at ITC_SM07 from 7 till 31 July covering a 2019 heatwave in Northwestern Europe. The upper panel shows the hourly precipitation and air temperature measured the KNMI AWS Twenthe airport about 12 km southwest of ITC_SM07.

Comment R1C24

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 m³/m³, even close to 0.8 m³/m³ on a location with sand / highly loamy sand (ITC_SM14, see Tab. S2, and also ITC_SM17). I find this quite spurious.

Authors' response:

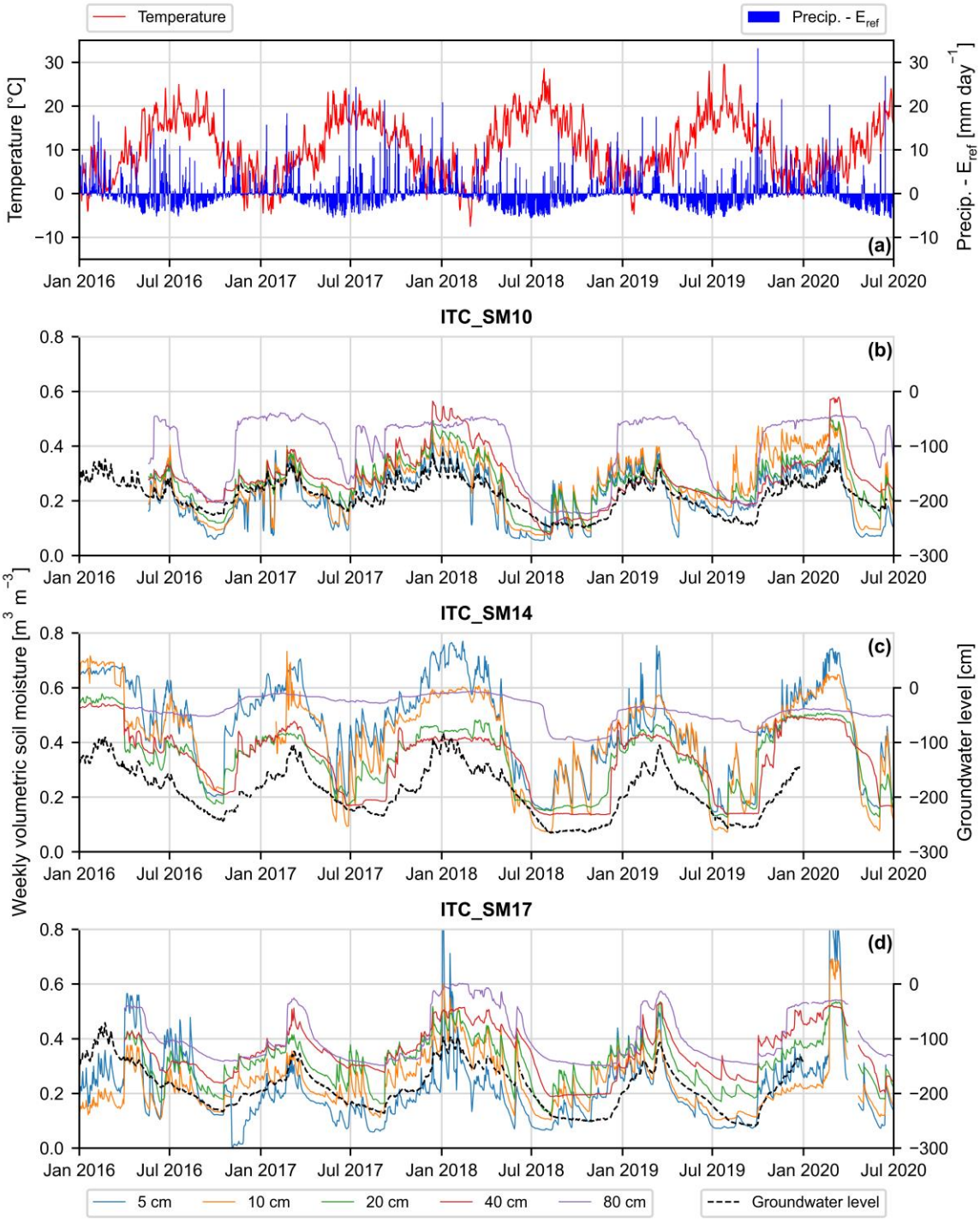
For the revised manuscript we decided to plot the daily precipitation minus the reference evapotranspiration instead of the accumulations as the referee suggests. We find that the daily precipitation surplus provides a better indication for period when we may expect a rise in soil moisture content in comparison to the accumulations.

With respect to the second point the referee made, we found that it was based on an earlier version of the dataset that was based on old calibration. For the figure in the revised manuscript, we have used the published calibrated datasets as is, which include significantly lower the soil moisture content.

Still, high soil moisture values are found, but those are from the Winter/Spring periods of the top 5 cm and 10 cm. Under those conditions, the soil will be completely saturated possibly even with standing water. Moreover, the fields where these high soil moisture values are measured in grasslands that have a high root density and organic matter content near the soil surface. This is not considered in the classification map from which texture class reported in Table S2 has been derived. The following sentence has been added to section 3.1,

L154-155:
It should be noted that near the surface the organic matter content is higher than one would expect based on the texture class and that grasslands have a dense rooting system.

The modified figure is shown on the following page.



Comment R1C25

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

Authors' response:

We have changed the text as follows,

L424-425:
Specifically at a depth of 80 cm, the effects of 2018, 2019 and 2020 droughts are visible, while …

Comment R1C26
l. 434: replace "measurement" by "level" and "increments" by "increases"

Authors' response:
done

Comment R1C27
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.

Authors' response:
We agree that assuming a linear relationship between soil moisture and groundwater is an oversimplification and not suitable to draw conclusions. However, a more detailed analysis, as the referee suggests, goes indeed beyond the scope of this article. We have, therefore, remove the table and corresponding text.

Comment R1C28
ll. 450-452: It is unclear what the authors mean by "[...] make it possible to address sub-catchment scale applications."

Authors' response:
We have modified the text as follows:

L443:
make it possible to study smaller scale applications than were addressed before.

Comment R1C29
ll. 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.

Authors' response:
As suggested we have replaced 'network scale' by 'for the entire network'

Comment R1C30
ll. 502: To what does the "hence" refer?

Authors' response:
We have removed the word 'hence' and let this sentence be the start of a new paragraph.

Comment R1C31
ll. 505: "network scale" - see above.

Authors' response:
We have replaced 'field and network scale' with 'footprints and grid cells' to connect with the satellite and model-based products referred to earlier in the sentence. The sentence is modified to read,

L494-495:
… to develop and validate satellite and model-based soil moisture products at the scale of footprints and grid cells.

Comment R1C32
ll. 505-508: How can your dataset be valuable for upcoming (future) SAR missions if it only spans until 2020??

Authors' response:
We intended to convey that new insights into the uncertainties of soil moisture reference may be helpful for the development of calibration/validation plans of upcoming satellite missions, such as the upcoming NISAR and ROSE-L mission. We have modified the text as follows,

L495-498:
This may, for instance, be relevant for the development of calibration/validation plans for upcoming Synthetic Aperture Radar (SAR) missions such as the NASA-ISRO SAR mission (NISAR; Kellogg et al., 2020) and the Radar Observing System of Europe L-band (ROSE-L; Davidson & Furnell, 2021), which have both soil moisture included as part of their mission objectives.

Comment R1C33
Please provide all figures (except 1 and 3) in vector format in order to allow for lossless zooming.

Authors' response:
We have all figures (except 1 and 3) in .pdf format that will be included in the publication package.

Comment R1C34
Check for consistent use of tense throughout the manuscript, specifically in the newly added parts.

Authors' response:
We have checked the entire manuscript and made modifications throughout.