Referee report on van der Velde et al. (2022), submitted to ESSD

Overall evaluation

In their manuscript *"Twelve years profile soil moisture and temperature measurements in Twente, the Netherlands"*, R. van der Velde and colleagues present a corresponding dataset which is supplemented by several soil moisture measurement campaigns which add more points around the long-term measurements in terms of snapshots.

I congratulate the author team for a precious and useful dataset. The overall approach and design (a sparse network of profile probes) is not specifically innovative, yet it allows for interesting analyses and should be a useful asset for the scientific community. Most importantly, the author spent a lot of effort in the (re-)calibration of the soil moisture sensors by comparing the instrumental readings to gravimetric soil moisture measurements - an effort that is well-spent, and certainly nothing to be taken for granted.

The paper itself is mostly well-written, the accompanying dataset well-organised, complete, and contains both raw readings and processed data, conforming with various user requirements.

I still have some specific comments which you might classify as major, but which I am confident the authors will be able to consider in one way or the other. Most of them should be understood as suggestions to improve the manuscript.

Specific comments

Structure and balance of the manuscript

I think in the final parts of the manuscript, the structure should be adjusted to make it more straightforward. My suggestion is to drop section *"5.1 Maintenance monitoring stations"* (which is, in my opinion, not relevant in the context of this paper), and to combine section *"5.2 Data processing and flagging"* with the section *"Data availability"* (please note that the section numbering is wrong after section 5). Section *"5.3 Known issues"* deserves, in my view, a dedicated single subsection, but should be renamed to *"Data uncertainty"*. Please use this section to provide a more comprehensive synthesis on the two most important sources of uncertainty: the sensor calibration (most importantly the relationship between permittivity and volumetric soil moisture, but maybe also some words on the relationship between the primary measurement variable, typically raw voltage, and permittivity, which is assumed to be set by the manufacturer), and the spatial representativeness (see comment below). The section title *"Time series"* is not really adequate. Still, I think it is an important section as it presents an opportunity to highlight the scientific potential of your dataset (see comment *"Underlying research questions and prospective application of the data"*).

Apart from the structure, parts of the paper could be more concise. Please reconsider whether some details on the instruments' construction design and the calibration procedure could be removed for the sake of readability.

Underlying research questions and prospective application of the data

In ll. 44-49, you describe previous research that was based on the presented dataset, and which already started more than 10 years ago. While this exemplifies potential use cases of the dataset, it also leaves the reader with a feeling that this dataset was only published after it had already been comprehensively used.

Of course, that is not true! All the more, however, it is important for this paper to highlight and to exemplify open research questions and application cases related to this data. I think the section that is now entitled "Time series" is a good opportunity to highlight interesting properties of your dataset - as you already started to do. Still, I am wondering why you only show 3 out of 20 stations and 4 out of 12 years. Of course, this paper should not be about a scientific analysis, and of course you should not just show time series of all stations at all depths for all years. But maybe you could demonstrate the potential of your dataset along some exemplary research questions, maybe also at different time scales? This is just a thought, but I think this section could really increase the value and impact of your paper. The section *"Summary"* could then be extended to *"Summary and outlook"* and hence summarise the perspectives of prospective research based on this dataset, and maybe how it could be combined with other datasets.

Horizontal and vertical representativeness

The supplemental campaigns are a great feature of this dataset, yet the importance of the campaign data is, in my opinion, not adequately represented. Only in ll. 364-367, you briefly address the issue of spatial representativeness of the permanent sensors, based on campaign data used by Benninga et al. (2020). However, that publication only looked into this issue for a small part of the fields, and also not in detail. I think it is worth providing a more comprehensive assessment of the uncertainty arising from a lack of horizontal representativeness, by using the campaign data. It does not need to be exhaustive, but at least I don't see why you cannot provide such an uncertainty estimate for all stations.

It is a bit unfortunate, of course, that the campaigns were limited to the upper 5 cm. It is of course more tedious, but still feasible to carry out snapshot campaigns for vertical profiles by successively drilling to the next measurement depth. With the data limited to the surface, it is impossible to assess horizontal and vertical representativeness in a uniform framework. I assume that the limitation of the campaigns was partly justified by resource constraints, and partly by the motivation to use the data for ground-truthing remotely sensed soil moisture products. However, it should not be me speculating… so I would like to ask the authors to be more transparent about that study design and its justification, as well as about the resulting limitations and uncertainties with regard to soil hydrological applications.

Calibration

I appreciate the effort put into the calibration of the probes by using gravimetric measurements. Still, I am wondering whether the presentation of the corresponding procedures and the results could be more concise.

Furthermore, I think that you usually calibrate the relationship between permittivity and soil moisture (which is non-linear), while you took the volumetric soil moisture reading and linearly "re-calibrated" them. Please elaborate briefly on possible implications, if this makes sense to you.

Integration into ISMN

Why are these data not integrated into the ISMN? I would be interested in hearing your opinion on this. Many of the stations included in the ISMN are very similar types of profile measurements, and it should be helpful to put your data into an immediate context with these.

Homogeneity

You pointed out that some of the monitoring locations had to be changed during operations. Is there any way to assess the effect on data homogeneity? Are there periods of overlap for the new and the old locations for such an assessment?

Missing data

Please specify, maybe in a table, the percentage of missing data for each station.

Data repository

I found the data repository (https://easy.dans.knaw.nl) very slow. Even for the small files, I sometimes had to wait for a long time for the download to start. I do not know the EASY repository, and it also does not offer much information. I assume it is ok, but maybe ESSD team members could check whether it meets the standards of ESSD? In the end, I could download the data, so it's fine for me.

Other specific comments

- *Introduction*: at the end of the introduction, could you briefly outline the structure of the manuscript, specifically since the sections are not related to a conventional manuscript structure?
- *Ll. 32*: Gravimetric measurements are also unsuitable for monitoring purposes due to invasive and destructive nature of the measurement
- *Ll. 36*: why "*relative* electric permittivity" ?
- *Ll. 73-80*: Why is this required? I suggest removing this paragraph.
- *Comments on maps in Fig. 1-3:* these should be made more clear and less redundant
    - In my view, Fig. 1 is not required as a standalone figure, it could be included as an inset map in the other maps

- Fig. 2: Why are specific land cover types removed from the DEM? Apart from that, the topography map could be combined with Fig. 3 (soil map), e.g. by using contour lines and/or hillshading
- Fig. 3: Please use hollow markers so the reader can better see the land use behind a location marker; the colour for sand is unnecessarily dominant which makes the figure kind of fidgety
- Fig. 4: again, the choice of colours is quite glaring. More importantly, though, the choice is not suited for readers with color vision deficiency.
- I would find it very helpful to have the monitoring sites and the precipitation gauges in one map. Precipitation is crucial to interpret the soil moisture observations, and it is unfortunate that there is not a closeby rain gauge for each profile probe. At least on the map, the reader should get an impression about the distance to the next rain gauge. Maybe that information could also be included in the location metadata?
- You might also mention the open radar composite data. This data also has its issues, but it provides full coverage and hence provide useful information specifically in the case of convective rainfall in the summer season (https://dataplatform.knmi.nl/dataset/radar-corr-accum-03h-1-0, https://dataplatform.knmi.nl/dataset/nl-rdr-data-rtcor-5m-1-0).
- *Ll. 153-155*: *"The instrumentation is [...] typically placed at the border of fields and preferably several tens of metres away from disturbing features (i.e. trees, roads or watercourses) [...] to minimize disturbance from recurring farming practices and optimize its representativeness for the adjacent fields."* How does the site selection, as described here, affect the representativeness for the agricultural fields? As pointed out above, it would be great if this issue was addressed in more detail in a dedicated section on uncertainty.
- *Ll. 157-159*: Are the site properties (specifically soil!) only derived from the soil map or were they also specifically determined for the exact point (texture, SOM, horizons, …)?
- *Table 1:* I think that the table is useful, but in its current form, it might also go to the supplement, specifically since parts of the legend (asterisks) are also explained there. I think that the column "soil type class" should rather be "texture class".
- *Ll. 169-171*: I think not all these details on sensor construction are required, please replace by adequate reference.
- *L. 238:* *"The sampling strategy during campaigns aimed at characterizing the top 5 cm soil moisture content of fields."* Stating the target variable (SM at 5 cm) is, in my view, not really a sampling strategy. Please also see my above comment on limiting the campaign to surface SM.
- *Section 4.1:* How was the position of the sampling locations measured (GPS or DGPS / accuracy)?
- *Ll. 238-240:* The criteria of selecting a sampling location remain unclear. Which role plays the distance between locations and the parcel size, as compared to a random placement? Fig. 8. does not really help me to understand the samling design.
- *Ll. 242-245:* Which design was used for locations that were neither grassland nor maize?
- *L. 246:* What do you mean by *"The collection of soil samples for GVSM determination [...] stopped [...]"*? Does that mean that after a specific campaign day, you decided that more gravimetric samples were not required?
- *Fig. 9:* I find the figure different to read because all the information is represented by the color code. Please consider a different approach: you could e.g represent the year by a

colored circle and the gravimetric measurement by a marker inside the circle. I would also use a different marker for the maize stems.

- *Fig. 10:* the labels are too small, the resolution insufficient.
- *L. 278:* *"[...] to develop the calibration functions for the ThetaProbe on a field campaign basis [...]"* - please be more specific about what you mean here.
- *Ll. 285-286:* The agreement for Fig. 11 seems higher, but it should be noted that the points are mainly from a different set of stations: so is the better agreement due to the probe, or due to the different locations? Of course, that is difficult to say, but maybe the issue could be briefly mentioned?
- *L. 286:* replace "little" by "small"
- *L. 287:* "distributions" is not the adequate term here, I think. Isn't it the relationships between probe readings and GVSM that you mean?
- *Ll. 296-297:* *"The performance metrics presented in Tables 4 and 5 show that the matching probe ('site') and GVSM measurements generally led to better performance except for the 2009 field campaign [...]"* - I don't agree entirely: in Tab. 5, the "mean" variant performs better for ITC_SM02.
- *Ll. 297-299:* *"Of the field campaign calibrations, the calibration developed for the HydraProbe (2016-2017) led to the best results with a RMSE of 0.032 m^3 m^{-3} versus 0.041 m^3 m^{-3} for 2015 and 0.048 m^3 m^{-3} for 2009."* This sentence is ambiguous, please change to: *"Of the field campaign calibrations, the calibration developed for the HydraProbe (2016-2017) led to the best results with an RMSE of 0.032 m^3 m^{-3} in comparison to RMSEs of 0.041 m^3 m^{-3} for 2015 and 0.048 m^3 m^{-3} for 2009, as obtained for the ThetaProbes."*
- L. 325: readme *file*
- *L. 347:* *"Long-term operation of in-situ monitoring networks goes hand in hand with measurement uncertainties."* - please remove this sentence.
- *Ll. 350-353:* I understand this is too late, but you could also recalibrate the relation between the prime measurement variable (raw voltage) and permittivity yourself, by using pure air and water, for each sensor - correct?
- *Table 8* should also contain, in a dedicated column, the main variable(s) represented by the corresponding third party datasets

Technical comments

- *Dataset:* In the files in `0_measurement_campaigns/1_processed_calibrated`, the combination of field IDs and location IDs are not unique. E.g. for file `Twente_fieldwork_ITCSM03_pd_cd.csv`, the first six rows show, for the field column, the values `1, 1, 1, 1, 1, 1` while the location column shows `1, 2, 3, 1, 2, 3` while I would have expected `1, 2, 3, 4, 5, 6`. Could you elaborate?
- The header of your data repository is "Ten years profile soil moisture and temperature measurements in Twente (version 2022)" - that is inconsistent with the title of your paper.
- *Ll. 22 ff.:* *"conversion of water into vapour via evapotranspiration at the expense of solar radiation"* - I find that weirdly phrased
- *Punctuation* requires revision throughout the paper
- *L. 114:* Is heath actually cultivated in that context? I found it surprising to list it under agricultural land use.
- *L. 175:* replace "circumstances" by "conditions"

- *L. 175:* replace "influence zone" by "footprint" or "spatial representativeness", if that is what is meant
- *Fig. 8:* the north arrow appears unnecessarily large
- *L. 283:* I suggest to remove a) and b) from the text, this is included in the figure caption
- *L. 405:* measurement**s**