Response to Reviewer 2 Comments (RC2) for ESSD-2024-26

We thank you for your comments on our manuscript and suggestions for improving our work. We have addressed all the comments. Our response (AC) to each reviewer comment (RC) are shown in bold text below.

Best regards,

Anatol Helfenstein, on behalf of all authors

General comments:

This is an outstanding contribution, not only the work itself, but also the open datasets and the comprehensive and thoughtful explanation of all the choices made in the construction of this dataset, as well as limitations and suggestions for use. I have some suggestions for improvement. But first answering the specific review questions from the journal.

-- Is there any potential of the data being useful in the future? Most certainly! Well explained in the article.

-- Are methods and materials described in sufficient detail? Yes, although anyone trying this in another country would need to be quite familiar with DSM already.

-- Are any references/citations to other data sets or articles missing or inappropriate? No

Is the article itself appropriate to support the publication of a data set? Yes, very much so.

-- Is the data set accessible via the given identifier? Yes

-- Is the data set complete? Yes

-- Are error estimates and sources of errors given (and discussed in the article)? Yes, also a very nice discussion of limitations.

-- Are the accuracy, calibration, processing, etc. state of the art? Yes, a probability sample was used for accuracy assessment, to compare with internal accuracy.

-- Are common standards used for comparison? Yes

Is the data set significant – unique, useful, and complete? Yes.

-- Consider article and data set: are there any inconsistencies within these, implausible assertions or data, or noticeable problems which would suggest the data are erroneous (or worse)? No, it is quite consistent. Problems with the source data and produced maps are well-discussed.

Is the data set itself of high quality? Very much so.

-- Is the data set usable in its current format and size? Yes

-- Are the formal metadata appropriate? No formal metadata are provided with the dataset. The user will refer to this paper to infer metadata.

-- Is the length of the article appropriate? Yes, it's long, but all is interesting.

-- Is the overall structure of the article well-structured and clear? Yes.

-- Is the language consistent and precise? Yes.

-- Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? Yes.

-- Are figures and tables correct and of high quality? Yes.

Is the data set publication, as submitted, of high quality? Very much so.

AC: Thank you very much for your positive feedback and for addressing and answering the specific review questions from the journal.

Specific comments:

1. Neither the paper, the linked 4TU webpage description, nor the readme.txt at that site mention the coordinate reference system (EPSG:28992 Amersfoort/RD New). Yes this is given in the properties of each .tif but since this is a little-known CRS outside of NL, maybe a mention in these three places (or at least the last two) would be useful to alert the user.

AC: Very good point. We will make sure to add the coordinate reference system in the paper and the paper assets (the three datasets and code).

2. "soil texture" are better termed "soil particle-size separates"

AC: We agree that soil texture can be a somewhat vague term. This is why we write "Note that clay, silt and sand content are particle size fractions (PSF) which together constitute soil texture" (L134). We henceforth always refer to PSF when speaking about clay, silt and sand together, unless we mean soil texture in a more general sense. We think that adding a third term, "soil particle-size separates", as you suggest, would create confusion and there is no real need for it. Hence, we prefer to keep "particle size fractions" (PSF) and soil texture, which are both used in numerous other papers in the field of soil science and digital soil mapping.

3. The entire first and second paragraphs seem unnecessary in a "Data" article. Everyone knows soil data is important, this is not summary of the uses of soil data. The specific case of demand for soil data in NL is relevant, the paper could start there. The history of soil survey and databases in NL is relevant and quite interesting, it's good to see all these references collected here.

AC: We agree that perhaps the first two paragraphs could have been shortened slightly. However, given that ESSD spans over a wide range of earth system science topics and only a small portion of these are related to soils, we think these paragraphs are key to setting the stage and help the readers understand a) how difficult it is to quantify soil information in space (and time) and b) the increasing demand for spatial soil information.

4. L100, 101 "Wadoux et al., 2021b, challenge 5" and other references to this. Explain the "10 Challenges in Pedometrics" and where the DSM challenges come in, otherwise the reference to "challenge 5" is obscure.

AC: We agree with the reviewer that this is now somewhat obscure. Instead of increasing the length of the introduction further, we have decided to remove the references to the specific challenge number and instead reference Wadoux et al., 2021b as a whole. That way no additional explanation is necessary.

5. Table 1: no method is given for bulk density. Probably known-volume cores.

AC: You are right, we only include a description for how the weight is measured, but not how density is obtained (for that the volume also needs to be known). As you suggest, it was indeed done using known-volume cores and we will add this to Table 1.

6. There is no discussion of the point geolocation method (obviously, the older ones were not with GPS) nor the geolocation accuracy. I think the surveyors marked their locations on the 1:50k (?) topographic maps, but these were estimates, although with fairly small field sizes maybe not so uncertain. At a certain point GPS came in -- was it used? Also, when the DPOP was not so accurate in the early days of GPS?

AC: We agree with the reviewer that it is important to include this. We will add the following text at the end of the first paragraph of Sect. 2.1 (L133):

"As the majority of the soil point data were collected before modern Global Positioning Systems (Table 2), soil surveyors marked the point locations on a 1:25'000 topographic map."

In addition, we will add text in the discussion (Sect. 3.3.1) explaining that positional uncertainty related to soil point data most likely contributed to the overall uncertainty of BIS-4D maps. On L475, we will add the following:

"Positional uncertainty due to marking locations on a 1:25'000 topographic map most likely also contributed to overall uncertainty of the BIS-4D maps, as investigated in other studies (Carré et al., 2007; Grimm and Behrens, 2010)."

Carré, F., McBratney, A.B., Minasny, B., 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma 141, 1–14. doi: 10.1016/j.geoderma.2007.01.018.
Grimm, R., Behrens, T., 2010. Uncertainty analysis of sample locations within digital soil mapping approaches. Geoderma 155, 154–163. doi: 10.1016/j.geoderma.2009.05.006.

7. L275 "In addition to rigorous quantitative accuracy assessment, we also evaluated the spatial patterns of BIS-4D prediction maps qualitatively by comparing them to existing soil maps in the Netherlands ... and based on expert judgement." This comes then in \S3.1.2, but there is no discussion of detail patterns. For example looking at BD50 around Wehl (GLD) I see some fields well-delineated but others with some in-field detail, which don't seem to follow obvious drainage lines. Center the map at (209850, 440850) to view. I am sure there are many other locations the authors could choose to comment on the detailed pattern.

AC: We thank the reviewer for this comment and the detailed observations of the maps. As expressed in the above comment, there are countless additional locations we could have chosen to comment on the detailed patterns in the maps. As the paper is already quite lengthy, we had to restrict ourselves and chose to comment the clay, silt and sand content maps in an area with perhaps the most diverse soil geography in the Netherlands (Fig. 4). These patterns are discussed in L323-325, L343, L347, L372 and in even more detail in L385-386. Furthermore we chose to visualize and discuss detailed patterns over depth and in a novel way to visualize soil variation over depth

(Fig. 5). Discussing even more examples of detailed patterns from the potentially 54 produced soil property maps (9 soil properties, 6 depth layers) was beyond the scope of this article. In terms of the quality assessment method, the focus was on properly presenting and discussing the accuracy based on statistical validation techniques and prediction uncertainty, as this can be quantified.

8. L487, 614. What is the nature of the privacy agreement? Can these points be shared under certain circumstances/agreements?

AC: We do not know the details of the privacy agreements and whether there are options to share topsoil data collected during agronomic surveys, as we did not use these ourselves (L487). Perhaps the companies who collect such data can provide more information, which to the best of our knowledge is mostly Eurofins Agro. As for the LSK and CCNL data (L614), more information can be obtained from Wageningen Environmental Research.

9. The produced maps have blank areas -- most are water and sealed urban areas (building footprints) but there are others, e.g., throughout the Montferland push moraine and along the Utrechtse Heuvelrug. This should be discussed under "artefacts" L384ff.

AC: We agree with the reviewer that we should include a remark about some areas with "no data" in the BIS-4D maps that were not water or buildings. We will do so in the suggested Sect. 3.1.2 starting at L389 with the following text:

"In addition, a few 25m pixels are contain no data even though they were not water or buildings. This may be due to no data values in some covariates but should be explored further in an updated version."

10. \S3.5 "BIS-4D user manual" is hardly that -- more like "tips to the user" or "guide for proper use".

AC: We agree. We will rename Sect. 3.5 "Best practices for proper use".

Do the GlobalSoilMap.net specifications include P? Not according to the latest version I have, from 2015.

AC: The reviewer is correct. Oxalate-extractable P is not a standard GSM soil property, but we also do not state that it is in the manuscript. On L119-121 we state that the nine target soil properties were chosen based on GSM, end-user needs and data availability. Hence, the decision to include oxalate-extractable P was motivated by end-user needs and data availability, even though it is not a specified GSM soil property.

Technical corrections:

1. In typography clearly differentiate - (minus) and - (from/to), e.g. in "(MEC=-0.11-0.38)" the first - is minus, the second - from/to but they look the same

AC: Thanks for the suggestion. We will adjust it based on LaTeX best practices. There should be a space before and after minuses, no space after a negative sign, and the "en dash" (--) should be used to indicate ranges. We will correct this throughout the article.

2. L100 "In addition, there are numerous challenges relating to the accuracy of soil maps" should start a new paragraph (new idea = new paragraph).

AC: We agree and we will start a new paragraph and remove the "in addition".

3. L304, 606, 615 URL not fully visible; make these into references? e.g., I imported to Zotero and exported the reference (can be done to BibTeX also). Same with the GitHub code repository.

Helfenstein, A., Mulder, V. L., Hack-ten Broeke, M. J. D., van Doorn, M., Teuling, K., Walvoort, D. J. J., & Heuvelink, G. B. M. (2024). BIS-4D: Maps of soil properties and their uncertainties at 25 m resolution in the Netherlands (Versie 2) [GeoTIFF (.tif)]. [object Object]. https://doi.org/10.4121/0C934AC6-2E95-4422-8360-D3A802766C71

Helfenstein, A., Teuling, K., Walvoort, D. J. J., Hack-ten Broeke, M. J. D., Mulder, V. L., van Doorn, M., & Heuvelink, G. B. M. (2024). Georeferenced point data of soil properties in the Netherlands (Versie 3) [Tabular (.csv); text (.txt)]. [object Object]. https://doi.org/10.4121/C90215B3-BDC6-4633-B721-4C4A0259D6DC.V3

Helfenstein, A., Mulder, V. L., Hack-ten Broeke, M. J. D., van Doorn, M., Teuling, K., Walvoort, D. J. J., & Heuvelink, G. B. M. (2024). Spatially explicit environmental variables at 25m resolution for spatial modelling in the Netherlands (Versie 2) [GeoTIFF (.tif); tabular (.csv); text (.txt)]. [object Object]. https://doi.org/10.4121/6AF610ED-9006-4AC5-B399-4795C2AC01EC

AC: In the ESSD journal recommendations, they recommended to have the link to the data paper assets (datasets and code) in the text followed by the citation and therefore we decided to follow these guidelines. We hope that the journal publication and type-writing team will take care of it. If not, we will only include the references.

4. L615 "that were openly available" -> "that are.."

AC: We will correct it to "that are…".

5. Are units with the solidus (e.g., mg/kg) acceptable? Standard scientific notation uses negative powers when needed, e.g. mg kg^{-1}.

AC: You are correct (https://www.earth-system-science-data.net/submission.html#math). We will change all unit notation with denominators accordingly.