

Referee 1 review

Response to the recommender's assessment.

Changes in the text

essd-2025-418 "Soil information and soil property maps for the Kurdistan region, Dohuk governorate (Iraq)"

Bellat et al.

Review by D G Rossiter 29-Oct-2025

Summary: This exceptionally-thorough and well-explained data paper presents details of the soils in the named region based on survey and models. It used modern methods (inference from MIR spectroscopy) as part of the soil properties determination. From this dataset a standard modern digital soil mapping (DSM) exercise was carried out to produce property maps over the study area. The maps were compared to the global SoilGrids v2.0 maps and, not at all surprisingly, had significantly better point evaluation metrics. All results and workflows are available under the FAIR concept. This paper can be a reference for how such a study can be carried out.

Major Comments:

1. *I appreciate the thorough review of previous mapping efforts in the region, it is good to have these listed for reference. The brief review of major pedogenetic processes is also appreciated. Similarly for the tectonic development, it places the study within context. The study's motivation is clear. Adherence to FAIR standards is appreciated. The entire workflow, all sources and products, are available, with DOI, and explained.*

Thank you so much for your positive feedback on our study. As you highlighted our aim was to have a fully thought-out workflow and process and data set accessible according to the FAIR principles.

2. *The Conclusions mainly repeat the Abstract and sections of the Discussion. I would appreciate a broader conclusion about the success of this study, the applicability of this kind of study to similar regions, the issues of global vs. local models, the main limitations to this kind of study, the importance of reproducibility and FAIR, etc. That is, after doing all this work, what do you conclude about the project?*

This comment is most welcome as I am sure the entire revision of the conclusion will help to improve the quality of the paper as suggested by reviewer #1.

1. "We developed a complete workflow for digital soil mapping at a regional scale in the Dohuk Directorate of the Kurdistan Region of Iraq, combining cLHS-driven sampling, MIR-based soil property prediction, and several machine-learning models to produce 50 soil property maps at 25 m resolution, as well as regional soil depth and soil class maps. Compared with *SoilGrids 2.0* and earlier local products, our models offer more locally relevant predictions and improved spatial detail, while also covering a broader set of soil properties and depth increments than previous regional studies. The soil class map further aligns with current WRB standards and benefits from greater observational density than earlier exploratory works.

Beyond these technical achievements, the study highlights the importance of integrating local measurements with models tailored to regional environmental gradients. Global products provide consistent baselines, but they cannot fully capture the geomorphological and topographic contrasts that drive soil variability at fine scales. The superior performance of our regional models demonstrates the complementarity between global and local approaches: global datasets remain essential for broad-scale comparisons, whereas locally calibrated workflows are crucial for operational land management, agricultural planning, and resource assessments.

The proposed workflow is fully transferable to other regions of similar size (~2,000 km²). Areas with comparable environmental conditions—such as western Iran, northern Syria, or parts of the Mediterranean basin—represent suitable candidates for direct methodological transposition. In addition, the time investment required for the entire process, from sampling design to final DSM production, is relatively modest: in our case, approximately one year (235 person-days). The datasets produced in this study also offer broader reusability, for example, as calibration material for MIR/FTIR spectral libraries (Safanelli et al., 2025; Viscarra Rossel et al., 2016) or as part of a regional or global soil profile archive database (Lachmuth et al., 2025).

A further contribution of this work is the provision of the first regional FAIR-compliant soil dataset (Crystal-Ornelas et al., 2022). Soil science research remains highly geographically imbalanced, with five countries producing more than 80% of global output (Cherubin et al., 2025). Southwestern Asia is sparsely represented, except for Iran, and no soil profiles from the Kurdistan Region of Iraq appear in the WoSIS database (Batjes et al., 2020) used by *SoilGrids 2.0*. (Poggio et al., 2021). Such data gaps reinforce global inequalities in environmental knowledge (Allik et al., 2020; Sonnenwald 2007) and limit the capacity of data-poor regions to benefit from international modelling initiatives. By openly sharing our dataset and workflow, we help reduce this imbalance and contribute to greater transparency and reproducibility in regional soil information systems.

Overall, this project demonstrates that high-resolution, locally informed digital soil mapping is both feasible and highly effective in data-poor regions. The workflow presented here substantially advances soil knowledge in Dohuk and provides a generalisable and reproducible model for improving soil information in other regions facing similar environmental and data constraints.”

3. Did you consider DSM for soil classes? Perhaps using a DSMART-like approach with your additional observations? This could be compared with Fig. 6. Obviously that is not to do in the paper, but was it considered and if so, why not attempted? Related to this, it is not clear how the soil class (not classification) map (Figure 6) was created. It's implied that this was expert judgement supplemented by observations, but it's not explicit. Also see comments below re: L443.

You raised a very interesting point. Following the generalisation of DSM for quantitative soil information, it is likely that this will spread to qualitative information, such as soil classes. However, we had several reservations regarding automatic soil classes classification:

- The first and maybe most prominent is the nature of soil classes. As described in the IUSS WRB (2022), soil classes can have multiple identities, with major and minor soil types present in the same landscape. Therefore, prediction models do not seem the most adapted tools – in our opinion – to discern the major and minor patterns of soil. However, it is true that with a random

forest model, you can assess the likelihood of a prediction and then select it manually based on your own expert opinion.

- Second, the low influence of some physical variables on soil patterns in certain cases. Vegetation from Landsat or Sentinel bands can “trick” the model into predicting well-developed soil where there is just a low vegetation cover of Regosol, for example (in the foothills or the Little Khabur Valley). A good example is provided by Taghizadeh-Mehrjardi et al. (2014), who used a decision tree to predict soil classes. The predicted soil map (Figure 6) was poorly linked with topographic or geomorphological variables, which is uncommon for soil maps where these criteria are usually strongly linked (Graham 2006).
- Third and last, our aim was not to conduct soil classes automatically from the beginning. The time spent in the field (3,5 months) was part of a general exploration of the region to understand the landscape as a whole (vegetation, hydrology, human installations, topography, geology, geomorphology). Therefore, we considered that an “expert-based knowledge” mapping is more suitable and reliable than an automatic one when time is allowed for the exploratory part of the survey.

The mention of DSMART-like is most welcome. We did look at the Easher et al. (2023) paper, and their results look promising. The question is whether this method is applicable in an arid or semi-arid context, whereas it is used here in a continental climate.

To clarify the use of “expert-based knowledge” mapping, we added a sentence.

1. “This resulted in a new regional soil map built on an expert-based knowledge method (Figure 6).”.

References:

- Easher, T. H., Saurette, D., Chappell, E., Lopez, F. de J. M., Gasser, M.-O., Gillespie, A., Heck, R. J., Heung, B., & Biswas, A. (2023). Sampling and classifier modification to DSMART for disaggregating soil polygon maps. *Geoderma*, 431, 116360. <https://doi.org/10.1016/j.geoderma.2023.116360>
- Graham, R. C. (2006). Factors of soil formation: Topography. In G. Certini & R. Scallenghe (Eds), *Soils: Basic Concepts and Future Challenges* (pp. 151–164). Cambridge University Press. <https://doi.org/10.1017/CBO9780511535802.012>
- Taghizadeh-Mehrjardi, R., Sarmadian, F., Minasny, B., Triantafilis, J., & Omid, M. (2014). Digital Mapping of Soil Classes Using Decision Tree and Auxiliary Data in the Ardakan Region, Iran. *Arid Land Research and Management*, 28(2), 147–168. <https://doi.org/10.1080/15324982.2013.828801>

4. *Can you comment on the realism of patterns as seen in Figures 9–14? We have the point evaluation statistics, but the map shows a landscape. Do the elements we see there correspond to reality, of course by expert judgement? Are the fine details revealed by the 25 m resolution realistic or artefacts?*

The question of the realism of the patterns is closely linked with the soil map classes. Indeed, in a first place the presence of nudlithic Leptosol (l.155), so to say bare soil make any interpretation of

soil variables difficult to assess. Similar can be said based on the soil depth map where < 5 cm soil are observed (Figure 14). Although a filter could have been applied, we did not want to add this filter as the purpose of the paper was to provide a raw data set at a regional scale.

An example is the low presence of C_t , in all of the anticlines, and of $CaCO_3$ from the eastern mountainous are, at the lower depth increment of the soil (Figure 10 – 13), while most of the Bekhair anticline mountains are limestone. It is likely that these areas do have great concentration of $CaCO_3$ (Mustafa and Benni, 2014) but due to the lack of soil samples – and basically any *solum* – the model is unable to predict correctly the values.

For others as the MWD (50 – 70 cm) or EC (70 – 100 cm), the distribution pattern is directed by the model used. CART models are based on regression trees which make them very sensitive to any changes in the variables. Therefore, distribution of the variables does not entirely fit – to our opinion – the reality of the landscape at some point. The question remaining is weather you should keep a better model (in term of accuracy) but less reliable in explaining the variability of the landscape or a model with lower prediction results but potentially more fitted for landscape prediction (e.g. QRF)? Our choice was the one of the better model accuracies.

Regarding the other variables they do represents a reality of the landscape this is already discuss in Part 4.1.

Regarding the soil depth map (Figure 14) yes indeed it fits entirely the landscape observations and this is already discussed 1.362 – 364.

One element also has to be taken into account it is the colour scale used in the figures 9 – 13. While they range from blue to red, the variables have in certain case very low SD or range (EC, MWD), which can make it more difficult to read. We did change the colour to the “classical” *viridis* palette.

References:

- Mustafa, M. M., & Benni, T. J. (2014). Mineral ressources of the High Folded Zone. *Iraqi Bulletin of Geology and Mining*, Special Issue : 6, Article 6.

Detailed Comments:

The WRB 2006 has been replaced by WRB 2022: IUSS Working Group WRB: World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps, 4th ed., IUSS, Vienna, Austria, 234 pp., 2022. However, I think the definitions used in this paper have not changed.

Thank you for this comment. Indeed, we did change the reference of the IUSS WRB 2006 into 2022 publication.

L179-80 RUSLE: how were the parameters calibrated? Were they from one of the earlier (cited) studies? Especially the K value.

Thank you for this comment. We added a sentence detailing the reference for each parameter and an annexe with the full equations used (appendix B). The process is also detailed in the supplementary material.

1. “In our RUSLE model (Appendix 2), we set K as the soil strength factor based on texture and organic carbon values (Kouli et al., 2009), and C was set with values from Morgan (2005) and

Swarnkar et al. (2018). Slope length and steepness factor LS were based on Desmet and Govers' method (1996), while the R factor from Morgan et al. (1984) was used. Finally, the conservation factor P , which could not be observed, has been set to 1 artificially, as suggested by Mehri et al. (2024). ”

L227 "the different index" -> "the different indices"

We modified this language, miswriting.

*L233 "We performed a standardisation of the predicted values of the texture on 100 % with `TT.normalise.sum` function (Moeys et al., 2024) and a additive-log ratio transformation (Aitchison, 1986) with the `alr` function (Tsagris et al., 2025)." This is not clear. Was the normalization following the MIR inference/wet lab measurements? And then were the `alr` variables used in the mapping, followed by back-transformation (as is done in *SoilGrids v2.0*)?*

Sorry if our explanation were not clear. You understood well, there was a small interference on the texture prediction based on the FTIR model. We did add an extra column to Table 2 to detail the measurements. These values were normalised on 100% before being transformed with the `alr`. The results of the best `alr` model were then transformed back to each texture variables. We added few sentences to clarify the whole process. We also specified that only continuous data were scaled when scaling the covariates. 1. “resulting in 45 models and 50 maps in total (the three texture variables only include two `alr` models).”

1. “We performed a standardisation of the predicted textures values from the Cubist model, with `TT.normalise.sum`”.
1. “Once the models were performed, the additive-log ratio was reversed into the three textures with the `\texttt{alrInv}` function, before being evaluated.”
 1. “We also scaled the continuous covariates [...]”.

L235 "close to a normal distribution Liu et al. (2022)" -> "close to a normal distribution (Liu et al. ,2022)"

We modified this part of the text.

L236 "2021" refers to what?

These data was referring to two references (Poggio et al., 2021; Varon-Ramirez et al., 2022). We added them.

L258 "relative "simple"" -> "relatively simple"

The modification was made.

\S2.4.3 and throughout the paper: what is meant by "soil depth"? Is it the solum (zone of pedogenesis) or to bedrock/completely unweathered parent material? This might be better termed "thickness" but "depth" is indeed commonly used. L365 "shallow and deep profiles" implies only the solum, is this correct?

You highlighted a fundamental notion in soil science. We rely on the IUSS WRB 2022 definition of the *solum* concentrated in the pedogenesis process layers. With “soil depth,” we imply all horizons of the pedogenesis process (*solum*) down to the top of the mineral layer/weathered consolidated

rock (C horizon), as detailed on p. 225 of IUSS WRB 2022. To clarify this point, we added a sentence in the article.

l. “By soil depth, we defined the height of the active horizons in the pedogenesis process (IUSS WRB 2022), the so-called *solum*. The measurement is taken from the top of the *solum* downward to the mineral layer (C horizon) or consolidated rock (R horizon).”.

L295 the correct reference for PICP is Eq. (2) of Malone (2011) not 2017. The formula is not found in Malone 2017. Malone, B. P., McBratney, A. B., and Minasny, B.: Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes, Geoderma, 160, 614–626, <https://doi.org/10.1016/j.geoderma.2010.11.013>, 2011. This equation and the others need definitions of the symbols, although some are standard. For PICP what is "v"? I learn from Malone 2011 it is "the number of observations in the validation [better, evaluation] dataset". What is "PL"? L, U as lower, upper limits can be inferred. Finally, the description "we used the prediction interval coverage probability to evaluate the corresponding prediction within an interval" is not clear. The Malone 2011 description is, to me, clearer: "the PICP is the probability that all observed values fit within their prediction limits".

We would like to thank referee 1 for pointing out the original publication of the PICP in the DSM application. We added the reference to Malone et al. (2011) and their formula, while also retaining the original publication of the PICP from Shreshta et al. (2006).

We did change the sentence explaining the PICP as described below.

l. “Finally, for the QRF models, we used the prediction interval coverage probability (PICP), which correspond to the probability that all observed values fall within their prediction intervals, here set at 90 % “

More generally, we added as you suggested a description of all equation formula excepted for the RPIQ.

L309 It's interesting that silt is so poorly predicted, yet most of these soils are on the silty edge of the texture triangle. And, clay and sand are in Category A and B. Can you explain why the poor result for silt, even though there is a lot of it and with a good range in these samples? This is mentioned on L387.

Thank you for this reflection on the results and their interpretation. Explaining the high silt content in the samples is difficult without a more detailed geomorphological and granulometric investigation. However, some leads can already be found. First, the relatively low sand content is due to numerous erosional agents in the region (water action, wind action, weathering, slope processes), but the relatively low effect of wind action compared to the southern and western parts, where Arenosol can be found (Buringh 1957). The low clay content is due to the Simele plain's higher elevation relative to the Tigris (a 30–50 m difference). These values are consistent with what Abdulrahman et al. (2020) observed in their paper in the Simele Plain.

Predicting silt is difficult for mid-infrared detectors. Hobeley and Prater (2019) highlighted this issue of silt prediction by stating, “This resulted in the best prediction of the silt fraction, which is traditionally poorly predicted by vis–NIR”. These poor results are also reported by Viscarra Rossel and Webster (2012), Janik et al. (2016), and Ng et al. (2022). In the case of Lacerda et al. (2016), the results of the silt prediction are even regarded as so low that they choose to only calculate the silt based on the equation [100 – clay% - sand%]. This approach was not used because it does not allow

an estimate of the silt error and does not account for the possibility that the predicted texture amount is lower or higher than 100% as in a full texture prediction. However, it seems that vis-NIR yields better results than MIR (Ng et al., 2022), which may also explain our poor predictive performance. Another lead is discussed by Hobeley and Prater (2019), stating that RF would perform better than Cubist or other models for silt prediction. We do not fully agree on this statement, as if indeed it yields better results, the silt stays in their publication, the “most badly predicted” variable with an RPD of **3.86** (Table 2). We here preferred consistency with a unified model (Cubist) for all parameters.

As for the chemical or physical reason for such low prediction of silt, it can reside in several reasons that would have to be explored:

- The higher water concentration in clay helps the spectral abortion.
- The quartz mineral is highly present in sands, which have a clear signature for spectral detection (Lacerda, 2016).
- The physical structure of the silt molecule (Janik et al. 2016).

We added the following sentence to detail these potential effects.

1. “Regarding the silt prediction, several reasons could explain its relatively low reliability. First, the use of MIR instead of vis-NIR, and the use of the Cubist model instead of RF, both of which yield better results (Hobeley and Prater, 2016). Second, the nature of the silt molecule, which is more challenging to detect than others (Janik, 2016; Lacerda, 2016).”

References:

- Hobeley, E. U., & Prater, I. (2019). Estimating soil texture from vis-NIR spectra. *European Journal of Soil Science*, 70(1), 83–95. <https://doi.org/10.1111/ejss.12733>
- Janik, L. J., Soriano-Disla, J. M., Forrester, S. T., & McLaughlin, M. J. (2016). Effects of soil composition and preparation on the prediction of particle size distribution using mid-infrared spectroscopy and partial least-squares regression. *Soil Research*, 54(8), 889–904. <https://doi.org/10.1071/SR16011>
- Lacerda, M. P. C., Demattê, J. A. M., Sato, M. V., Fongaro, C. T., Gallo, B. C., & Souza, A. B. (2016). Tropical Texture Determination by Proximal Sensing Using a Regional Spectral Library and Its Relationship with Soil Classification. *Remote Sensing*, 8(9), 701. <https://doi.org/10.3390/rs8090701>
- Ng, W., Minasny, B., Jeon, S. H., & McBratney, A. (2022). Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. *Soil Security*, 6, 100043. <https://doi.org/10.1016/j.soisec.2022.100043>
- Rossel, R. A. V., & Webster, R. (2012). Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. *European Journal of Soil Science*, 63(6), 848–860. <https://doi.org/10.1111/j.1365-2389.2012.01495.x>

L346 "river bakns" -> "river banks". Spell-check.

Thank you for spotting this error, the change was made.

\S4.3 Another interesting comparison with SG2 would be the prediction ranges. SG2 likely smooths more than this study, see Table 7 where the Q1-Q3 range is always much narrower. This can be brought out in the text -- the interesting discussion is about global vs. local models. The SG2 maps are much more uniform than the maps from this study.

You raised an interesting point. Due to its coarser size and also the wider scale of measurement observed the *SoilGrid.2.0* product is likely to produce narrower. We did provide an additional table of the metrics from *SoilGrid 2.0*. (Appendix F) which we added to this discussion.

1. “The standard deviation of the *SoilGrids 2.0* product is smaller, except for the pH, than for our prediction models (Table 7). This shows a narrower distribution of values, likely due to the wide range of input data used for the *SoilGrids 2.0*. The diversity of soil types and input data at the global scale makes the *SoilGrids 2.0* model respond relatively homogeneously at the regional scale.”
1. “We also compared the evaluation metrics of our predicted values with those obtained from *SoilGrids 2.0* (Appendix). Overall, our models outperformed *SoilGrids 2.0* across all evaluation metrics for pH, N, Corg, sand, silt, and clay at all depth intervals. The only exception was the sand R² score at the 60–100 cm depth interval, which was slightly higher for the *SoilGrids 2.0* model.”

L387 "should be interpreted with caution—consistent"... with what?

Indeed, an extra word was added for no reason. We removed “—consistent”

L392 "Abdulrahman et al. 2020 " -> "Abdulrahman et al. (2020)". L390 maybe make it explicit here that this is not a DSM product, rather an expert updating from field work and manual interpretation of remote sensing products (correct?).

Yes, the work of Abdulrahman et al. (2020) consisted of a punctual soil profile sampling study. We used their observations as a comparison to our soil profiles results and to also state that soil profiles (*solum*, see \S2.4.3) with a depth greater than 100 cm are concerning few soil profile (less than the half) and even some have to be treated with caution (B2ss and B2hk). calcrete layer can be very similar with C horizons. We rephrase the sentence of 1.392

1. “[...] although other local profile observations (Abdulrahman et al. 2020) tend to show that profiles below 100 cm are uncommon”

L401 the Hazelton & Murphy guidelines are for conventional mapping, not DSM. They are expressed in terms of map scale and cm² of printed map. Here the product is digital at 25 m resolution. How is the density here converted to match these guidelines? The argument about cLHS is much more relevant for DSM using machine learning from covariates.

You are entirely right. This comparison with Hazelton & Murphy mapping convention is relevant for DSM products. We removed the sentence mentioning it.

L434 formatting problem with the URL <https://mathias-bellat.shinyapps.io/Northern-Kurdistan-map>, which goes over a line break so gives a 404 error if not manually adjusted

Thank you for noticing this error we modified the text and implementation in LaTEX.

L443 "shallower resolution " -> "higher resolution"? And what is that resolution? It's nowhere stated. L435 says 1:200 000 scale, which implies polygons with a minimum legible delineation

(MLD) of 160 ha (0.4 cm² on the map). But L390 "The updated soil classification map (Figure 6) must be interpreted with care, specially at micro-scale (<1:50,000)..." implying a smaller MLD.

You are right, the observations of Buringh (1960) from the previous soil map (1957) are made at the national scale and no information were given for its sampling or whatever. The only comment on the map is made p. 291 (Buringh 1960) "The topographical map of Iraq, quarter of an inch to 1 mile; The provisional geological map of Iraq, 1:2,000,000, 1937; The soil conservation map of Iraq, 1:2,000,000 (Gibbs, 1953); The forest map of North-eastern Iraq, 1:850,000; Some aerial photo-mosaics; Some soil and land classification maps of project areas.". From the reading of the book it is likely that few profiles < 20 where done the north-western Kurdistan region. The only information is the scale of the map which was at 1:1 000 000.

Our own resolution indeed was not stated but based on the number of observations 1.399 (53.5 per 1,000 km²) which regarding Hazelton and Murphy (2016) table 1.2 would be at a reconnaissance 2 : 250 000 scale. However, the recent development of G.I.S and remote sensing has made the relevance of map scale less import. In the supplementary files the complete vector files are available making possible mapping at any scale.

However, to reduce the ambiguity we change the sentence.

1. "[...] must be interpreted with care, especially for local agricultural or construction planning."

References:

- Buringh, P. (1957). Exploratory Soil Map of Iraq (Division of soils and agricultural chemistry, Directorate General of Agricultural Research and Projects,) [Soil map]. Ministry of Agriculture. <https://esdac.jrc.ec.europa.eu/content/exploratory-soil-map-iraq-map-1>
- Buringh, P. (1960). Soils And Soil Conditions In Iraq. Minstry of Agriculture (Iraq). <http://archive.org/details/buringh-p.-soils-and-soil-conditions-in-iraq.-republic-of-iraq>
- Hazelton, P. A., & Murphy, B. W. (2016). Interpreting soil test results: What do all the numbers mean? (Third edition). CSIRO Publishing.

Figure 6 suggest that this is a polygon map.

Sorry, we did change the map into a raster figure instead of a vector figure. The figure was also slightly updated to include Türkiye soils. We also changed the others large scale maps and figure 2 which were vectorised maps.

Figure 3 the inset showing the region is not needed, that has already been shown in Figure 1 and can be found from the coordinates on the main map.

Indeed, the additional small map on the Figure 3 was repetitive regarding the presence of Figure 1, we did remove it.

Figure 8 bicolor key y-axis partially obscured

We apologise, reviewer #1, but we did not understand this comment. In our pre-print version, the y-axis was not obscured. However, we did change the legend position.

Citation: <https://doi.org/10.5194/essd-2025-418-RC1>