We sincerely thank the Editor and the three anonymous Reviewers for their time, constructive comments, and valuable suggestions. It has significantly helped us improve the quality of our manuscript. In this point-by-point response, we have addressed all comments raised. Below, we first provide a summary of the key revisions, followed by detailed responses to each specific comment. All changes made in the revised manuscript are clearly highlighted in red text for the reviewers' convenience.

Reviewer 1:

This manuscript proposes a method to improve the accuracy of digital soil maps by iteratively applying bias correction across soil depths in a spatial context. While this topic is highly important in DSM studies, several points need further clarification to strengthen the quality of the manuscript.

We thank the reviewer for your valuable and constructive comments. We have carefully considered your suggestions and have revised the manuscript accordingly. Our detailed responses follow below.

- **The iterations are demonstrated for points with observations; how does this method apply to pixels or points with no prior observations?**

  Thank you for your suggestion. We clarify that the IRC method is designed to update existing prior soil maps when new soil observations become available, not to generate predictions where no prior map exists. The method requires "prior" in our framework refers to the spatially continuous prior soil property map, not to individual point observations. This prior map provides estimates at every pixel across the study area through spatial interpolation techniques used during its creation. When we collect new soil observations at specific locations, we calculate residuals by comparing these new measurements with the corresponding prior map values at those locations.

  In the revised manuscript, from line 212 to 219, we added the relevant content to explain this concept ‘The IRC method does not require prior and new soil observations to be co-located at the same pixels. Instead, the method requires that a prior estimate exists at locations where new soil observations are available. By learning the relationship between residuals (differences between new observations and prior estimates) and environmental and soil covariates at sampled locations, the trained model can interpolate residual corrections across the study area.’

- **What are the covariates to determine probabilistic maps of soil class and soil properties?**

  Thank you for your suggestion. To determine prior probabilistic maps for soil class and properties, we used environmental covariates including NOAA GOES-R Land Surface Temperature, the USGS National Land Cover Dataset, ESA Sentinel-1 and -2, and USGS gamma radiation and magnetic anomaly data. More details on these covariates are available in the S3 table of this paper (Xu et al., 2025).

  In the revised manuscript, we provided more detailed explanation in section 2.1.1, such as content from line 235 to 258: 'In Figure 2b, the table details features used to train the Random Forest regressor for residual prediction. The feature space consists of environmental covariates that remain fixed across iterations and soil covariates that are updated iteratively:

  (1) Environmental covariates (21 dimensions): These capture spatial variations in soil-forming factors and remain unchanged throughout all iterations. The covariates include remote sensing data (Sentinel-1, Sentinel-2, GOES land surface temperature) and terrain attributes, identical to those used in the prior mapping method (Xu et al., 2025).

  (2) Depth information (1 dimension): The centroid (median value) of the soil depth interval for the target layer (e.g., 10 cm for the 5–15 cm layer), describing the vertical position in the soil profile.

  (3) Representative soil property values (1 dimension): The expected value (weighted mean) of the soil property at each pixel in the modeling layer, representing the current best estimate. This is computed as the weighted sum of top-probable values.

  (4) Top-probable soil property values (1 dimension): The current predictions at each pixel, reflecting both intra-pixel and inter-pixel soil heterogeneity.

  (5) Inter-layer differences (5 dimensions): Differences in top-probable predicted soil property values between the target layer and the other five depth layers. For instance, if modeling Depth 2, the inter-layer differences would be (D2−D1), (D2−D3), (D2−D4), (D2−D5), and (D2−D6). These features capture vertical correlations in the soil profile and aid in estimating spatial patterns.

  (6) Weights (1 dimension): Probabilities associated with each top-probable soil property value. These weights remain fixed throughout iterations.'

- **Need to clearly mention the division of data for model development and the iterative bias correction. To my understanding now, you used data from Wosis and SCD to develop the prior soil maps and used your own samples for**

correction. If so, the next question is how do you make sure consistency between Wosis and SCD data when combining those for model development?

Thank you for your suggestion. When develop prior predictions of soil class, we used soil pedons from the National Soil Information System (NASIS) and part of SCD; when develop prior soil properties, we then used a harmonized soil properties database mainly developed using SSURGO. Development of prior soil properties can be found in more detail from this cited paper: *Xu, C., Huang, J., Hartemink, A. E., & Chaney, N. W. (2025). Pruned hierarchical Random Forest framework for digital soil mapping: Evaluation using NEON soil properties. Geoderma, 459, 117392*. And then, when predicting posterior soil properties, we used SCD (the part that did not used in predicting prior soil class and properties), WoSIS, and soil data that USDA and UCR colleagues collected in California.

We also provided the relevant content in the revised manuscript, from line 443 to 446, For the California case study, prior soil property maps were generated using the pruned hierarchical Random Forest (pHRF) method (Xu et al., 2025). The pHRF-derivef soil maps were developed with soil pedons from the National Soil Information System (NASIS) and part of SCD (the remaining data not used in IRC method).

It is important to make sure soil data from difference sources consistent with each other. In the revised manuscript, we mentioned that the SCD soil data used in the study were collected using standardized laboratory methods. SCD is a subset of the National Cooperative Soil Survey (NCSS) database. And WoSIS within the study area are mainly collected from NCSS database. These soil data were collected using standardized methods, such as the Kellogg Soil Survey Laboratory (KSSL). For each soil property, we also set up plausible ranges of soil property values. If any reported value fell outside its range, we removed it from further procedures. For soil texture, compositional constraints were applied. Additionally, in line 218, we mentioned our use of the Integral Suspension Pressure method (ISP+) for precise particle size analysis of soil and sedimentary materials, following standardized methods, when collecting soil samples in California for our measurements. Please find the relevant content in section 2.2.1 in the revised manuscript.

- **While the models were evaluated using the OOB dataset, I assume that you only develop one model for one soil property at a specific depth. How does this approach guarantee that your model is representative of soil variability across California? Also, how did you divide the data for training and validation?**

  Thank you for your suggestion. We trained one model per soil property across all soil layers in one iteration. We developed another new model in the next iteration. And in each iteration, we only updated residuals for one specific depth. In the iteration, we proceeded update residuals for a different depth. In this way, all depths would be updated.

Our methodology ensures representativeness of soil variability across California through several considerations:

(1) Spatially distributed training and validation data: There are more than 3000 new observations for residual correction. Our soil samples span different areas within CA (Figure 3). The Random Forest regressor can learn relationships between residuals and feature space across different environment, capturing heterogeneous soil patterns across the state.

(2) Convergence across iterations: The iterative training process demonstrated that the learned relationship between residuals and soil covariates remained consistent across different iterations, indicating stable and reliable spatial patterns.

(3) Vertical correlation preservation: Training a single model across all soil layers (in one iteration), rather than separate models for each depth, allows the model to learn and preserve vertical correlations within soil profiles.

Regarding data division, validation was performed using out-of-bag samples; and the remaining samples are used for model training. We provided relevant content in the revised manuscript, from line 452 to line 457, as 'Model training and evaluation were performed using out-of-bag (OOB) sampling, with OOB samples (samples withheld from the training process and not used to fit the models) that shared the same geolocation as training samples removed to prevent data leakage and reduce spatial autocorrelation effects. In each iteration, a new Random Forest model is trained to update residuals for one specific depth interval, and the same set of OOB samples remains excluded throughout to ensure independent validation.'

Line 159: Please specify the standardised depth intervals. And briefly explain the method of this harmonisation process.

Thank you for your suggestion. We cited this reference to our revised manuscript *Arrouays, D., McKenzie, N., Hempel, J., de Forges, A. R., & McBratney, A. B. (Eds.). (2014). GlobalSoilMap: basis of the global spatial soil information system. CRC press*., in line 362 in the revised manuscript (Arrouays et al., 2014).

The harmonization was performed using a spline function to interpolate soil property values from the original horizon depths to these standard intervals. We also added explanation and citation for the interpolation method, starting from line 362, as 'The harmonization was performed using equal-area spline functions to interpolate soil property values from the original horizon depths to these standard intervals (Hartemink et al., 2010). The spline function fits a smooth curve through observed values at their measured depths, then calculates the area under this curve within each standardized depth interval and divides by the interval width to obtain the value.'

Figure 1: I guess the number of data points is different for each property. It might be more informative if you could mention it in the figure.

Thank you for your suggestion. We have revised Figure 3 (previously Figure 1) and its caption to address this point.

In section 2.2.1, from line 370 to line 378, we added the following description of sample size variation: 'Location of soil profiles and their distribution of soil property values are presented in Figure 3. Six soil properties are studied: sand content, silt content, clay content, pH, soil organic matter (log-scaled), and oven-dry bulk density. These samples were not co-located with the training samples used to generate the prior maps (samples at the same locations were already removed). The number of observations varies by soil property: pH has the most samples, followed by oven-dry bulk density and soil organic matter. The sample sizes across properties can also be inferred from the frequency histograms shown in the lower-left corner of each panel in Figure 3. Across all depths combined, each soil property has more than 11000 observations in California. The number of observations generally decreases with depth, with depths below 1 m having notably fewer samples compared to shallower layers.'

To make this information clearer to readers, we have updated the figure caption as follows:'... Note that the total number of soil measurements varies by property and generally decreases with depth beyond the surface layer, with the surface layers and depths below 1 m generally having fewer observations.'

Line 177: what are the interval depths for the Wosis dataset?

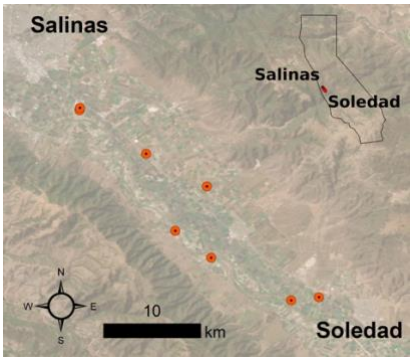
Thank you for the suggestion. The WoSIS dataset comprise vertical soil profiles with measurements tied to soil horizons and their ranges differ by location. We harmonized all profiles to the six GlobalSoilMap standard intervals using spline interpolation. In line 398, we mentioned that the WOSIS reported soil properties for different soil horizons.

Line 186: It might be good to add a reference for the standardised lab method

Thank you for your suggestion. We added this citation in the revised manuscript: Soil Survey Staff: Kellogg Soil Survey Laboratory methods manual, U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln, Nebraska, 2014 (Soil Survey Staff, 2014).

Figure 2: In which part of California are these samples located? An inset map would be helpful for this case. Is there any sample from Wosis and SCD located around these areas? If so, what is the closest distance?

Thank you for the suggestion. In Figure 4 (original Figure 2), we included an inset map in the upper-right corner that shows the location of the Salinas Valley within California. The field measurements were collected between Salinas and Soledad in the Salinas Valley, and this is now clearly indicated in both the figure and the caption.



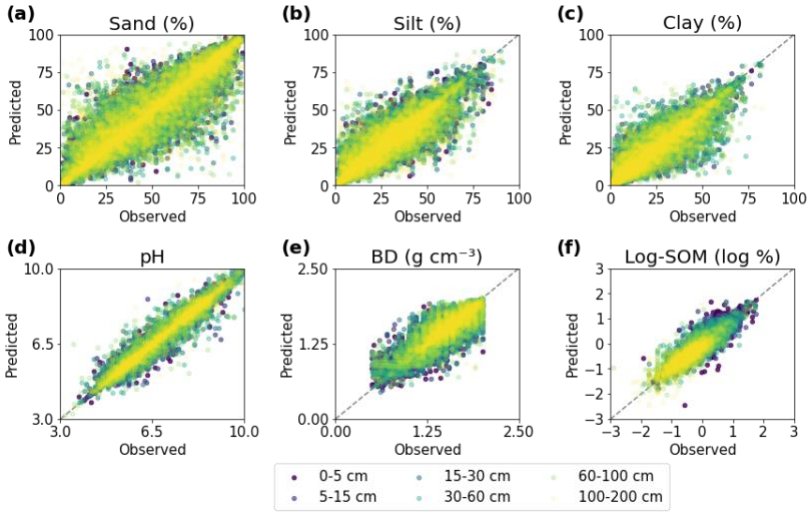
We note that WoSIS and SCD samples do exist near our field sampling locations. One advantage of our field sampling campaign was the ability to access private agricultural lands, which complements the NCSS database (from which WoSIS and SCD are largely derived) where access to private lands is sometimes restricted. However, we did not show inter-sample distances in the revised manuscript, as they are not co-located samples and discussing spatial proximity not fit with the remaining of the paragraph.

Figure 5: This is difficult to see the difference in performance between soil layers. I suggest plotting the evaluation parameter (R2, RMSE, etc) with the depth as the y-axis for each soil property. Also, why is there a negative value on SOM? What does that mean? Units should be provided on each graph of soil properties.

Thank you for the suggestion. We found that adding depth-specific breakdowns for each metric would make the figure overly complex. Instead, we created a table (Table S1) in the Supplementary Information that presents evaluation metrics (R^2, RMSE, bias) for each soil property at each of the six depth intervals. We reference this table in the main text (lines 493) as: "Depth-wise analysis of model performance metrics for each soil layer is provided in the Supplementary Information (Table S1)."

Property	Depth (cm)	RMSE	R^2	ρ
Sand	0–5	6.998	0.903	0.951
	5–15	6.998	0.900	0.949
	15–30	7.475	0.888	0.942
	30–60	9.438	0.829	0.911
	60–100	10.788	0.804	0.898
	100–200	12.594	0.761	0.874
Silt	0–5	5.340	0.861	0.928
	5–15	5.384	0.849	0.922
	15–30	5.754	0.822	0.908
	30–60	6.602	0.768	0.878
	60–100	7.156	0.748	0.867
	100–200	8.488	0.704	0.842
Clay	0–5	3.020	0.947	0.973
	5–15	4.414	0.891	0.944
	15–30	5.111	0.872	0.934
	30–60	5.962	0.793	0.892
	60–100	6.888	0.813	0.903
	100–200	6.947	0.788	0.890
pH	0–5	0.213	0.957	0.979
	5–15	0.260	0.936	0.968
	15–30	0.259	0.943	0.971
	30–60	0.295	0.935	0.967
	60–100	0.302	0.941	0.970
	100–200	0.269	0.956	0.978
BD	0–5	0.166	0.710	0.847
	5–15	0.173	0.673	0.826
	15–30	0.163	0.684	0.832
	30–60	0.170	0.656	0.816
	60–100	0.163	0.691	0.835
	100–200	0.146	0.667	0.821
SOM	0–5	3.335	0.476	0.755
	5–15	2.560	0.514	0.743
	15–30	1.626	0.664	0.826
	30–60	1.522	0.654	0.816
	60–100	1.026	0.706	0.852
	100–200	1.175	0.597	0.794

Regarding SOM, the negative values represent low SOM contents because SOM is plotted on a log scale. To address this comment, we updated Figure 5 as belows in the revised manuscript.



Line 443: What soil properties is this nRMSE value for?

Thank you for your suggestion. This is for all soil properties. We changed this sentence to 'Overall, all soil properties maintain reasonable normalized bias and nRMSE (with nRMSE values consistently less than 0.02 for both prior and posterior predictions)' to avoid ambiguity.

Section 3.2: Which dataset that you use to evaluate in this section? Please clarify

Thank you for your suggestion. We used out-of-bag (OOB) samples for evaluation. We also removed OOB samples that share same geolocation with training samples to avoid data leakage. Please refer to a previous answer for more details.

Figure 6: Is this the average value across the soil depths?

Thank you for your query. Those metrics are averaged across all soil depths. In line 547 in the revised manuscript, we added this modification, as 'Aggregating data from all depths, Figure 6 shows the degree of improvement across different soil properties' to emphasize it.

Section 3.3: What does the reduced PIW in this case mean? To my knowledge, reduced PIW means that the prediction interval is narrower, but that does not guarantee for more accurate prediction.

Thank you for your suggestion. We agree that a reduced prediction interval width (PIW) alone does not guarantee improved accuracy, as it primarily reflects uncertainty range. Section 3.3 illustrates that the iterative residual correction largely reduced uncertainty in the posterior prediction of soil properties, as the title of Section 3.3 shows it is a paragraph discussing uncertainty analysis. Together with previous sections, our results can demonstrate that the iterative correction method improves accuracy and lowers uncertainty, as supported by the reduced bias and RMSE listed in previous paragraphs in the manuscript.

Line 583-585: This approach should also be mentioned earlier in the method section

Thank you for the suggestion. We largely modified section 2 to better illustrate the method. In the revised manuscript, we first described the framework of the IRC method. Then, we provided a worked example, using one selected soil profile to illustrate the evolution across different iteration. Finally, we presented the California case study for its application. Given the substantial revisions throughout this section, we have not copied the updated text here. Please refer to section 2 to see our updates.

Reviewer 2:

This article proposes an iterative correction method to update soil property maps based on the availability of new soil property measurements. The article is broadly well written and represents a unique and useful contribution to digital soil mapping; however, I have significant reservations about the current state of the methods description and recommend MAJOR revision. I honestly did not read past section 2 after spending > 3 hours trying to understand the methods section because a close reading seemed to reveal inconsistent terminology and only made me more confused. Figure 4 was an attempt to graphically explain the model, but it only confused me more.

Also, I initially understood that this method was only applicable to POLARIS or pHRF predictions, but the author's response to the comments from another reviewer stated that this method was applicable to ANY soil map.

I am also deeply concerned about access to the code to reproduce this research. This seems like a method that I would be interested in applying to other soil map products, but the authors only state that the data (not the code?) is available upon request. Why not put this on github?

I'm also confused, does this operate on a cell-by-cell basis or is the distribution adjusted for all grid cells?

I am hopeful that the following comments will be useful to convey areas that I found confusing.

Dear reviewer,

Thank you for your constructive suggestion. We appreciate the time you invested in carefully evaluating the manuscript. Below, we respond to your main concerns and outline how we revised the manuscript accordingly.

Applicability to Any Soil Map: We revised the manuscript to state that the proposed iterative residual correction method is not restricted to POLARIS or pHRF products. Instead, it is applicable to other soil map products that provide a spatially continuous (prior) estimate of soil properties within the study area. And we need such information to calculate residuals (difference between new observation and prior estimate of soil property values), since residual is our target variable. The method's applicability to other soil map is inherent in this design. Other soil dataset that provides a continuous baseline estimate can be used as the "prior map" in this process.

Here we emphasize the requirement for using a spatially continuous soil map, because this method does not require prior and new observations to be collocated at the same pixel. In practice, new and legacy soil samples are rarely at identical locations. More likely, at locations

where new soil observations are available, there are no collocated legacy data present. The producers of the prior soil map normally use nearby information to interpolate estimate of soil property over this pixel. Hence, the main idea is to update existing soil maps when newly collected data becomes available

Cell-by-Cell vs. Global Adjustment: This method performs cell-by-cell updates and produces region-wide distribution adjustment. The process begins by sampling residuals at grid cells where new soil observations are available. A Random Forest regressor is then trained on these cells to learn the relationship between the residuals and the feature space. Once trained, the regressor interpolates this learned relationship across the entire study region.

As a result, the distribution of soil property values is updated individually at each grid cell through predicted residual corrections. At the same time, because the residual correction is applied consistently across all cells within the study region, these local updates collectively alter the region-wide distribution of soil properties.

Code Sharing: We updated our "Code and Data Availability" statement and already made a public GitHub reposotiry to share the code. Here it the link:
https://github.com/emmaxu43/IRC_CA/tree/main

Clarifying Figure: We revised the schematic figure (Figure 1 and 2 now in the revised manuscript) to more clearly distinguish between point-based observations, prior map, components of feature space, predicted residuals, and the interpolated posterior correction. The revised figure 2 used an example to illustrate the iterative correction process. Additional details on how the figure were updated are provided in our point-by-point response to Referee #3. Here, we would like to briefly explain the method:

1. Residual calculation: At locations with new soil observations, residuals are computed as the difference between ground-truth measurements and values from the prior soil map (not restricted to POLARIS or pHRF).

2. Spatial modeling of residuals: A Random Forest regressor is trained using these residuals as the target variable. The feature space includes environmental covariates (e.g., topography, remote sensing data) and soil covariates (e.g., prior soil property values and top-probable values from previous iterations), sampled at the same locations.

3. Map update: The trained model predicts residual corrections for all grid cells in the study area. These predicted residuals are added to the prior map to produce posterior soil property estimates.

4. Iteration: This procedure is repeated until the change in residuals between iterations falls below a user-defined threshold.

Line 14: “Existing soil maps”. Does this mean all soil maps?

Thank you for your suggestion. We agree that the phrase “existing soil maps” is ambiguous. In the revised manuscript, we clarified this in the abstract by replacing it with “existing probabilistic soil maps — spatially continuous soil property maps that provide a prior estimate of soil properties”.

In the context of this study, the prior specifically refers to outputs from the pHRF method. However, the framework is not restricted to pHRF and can be applied to other probabilistic DSM products, provided that a prior prediction surface is available.

In the revised manuscript, we provide the relevant content from line 211 to line 220: ‘In this IRC framework, "prior probabilistic soil property maps" refer to spatially continuous soil property maps that provide an initial (prior) estimate of soil properties with associated uncertainty across the study area. These prior maps provide, for each pixel and depth interval, a distribution of possible soil property values with associated probabilities or weights. The IRC method does not require prior and new soil observations to be co-located at the same pixels. Instead, the method requires that a prior estimate exists at locations where new soil observations are available. By learning the relationship between residuals (differences between new observations and prior estimates) and environmental and soil covariates at sampled locations, the trained model can interpolate residual corrections across the study area.’

Line 27: “prior probabilistic soil property maps”. Does this mean this only applies to POLARIS-like maps? The sentence starting on Line 30. Is this for every pixel in the map?

Thank you for your suggestion. The method is not limited to POLARIS- or pHRF-like products. In this manuscript, pHRF is used as an example to demonstrate the residual correction framework, but the approach itself is more general. Other soil maps can also be used as prior probabilistic soil property maps using this method, please refer to the previous answer.

Regarding line 27: here the “prior probabilistic soil property map” refer to the pHRF-derived soil maps. And the residual correction is applied to every pixel in the map. Residuals are first learned at locations where new soil observations are available, using a Random Forest regressor trained with collocated environmental and soil covariates. The trained model then interpolates residuals across the entire study area, allowing residual corrections to be applied at all pixels within the study region.

Please refer to the previous answer to see modification in the revised manuscript.

Line 73: Latin-hypercube

Thank you for correcting the words. We already made the change in the revised manuscript in line 70.

Line 77: Geostatistical models do not heavily rely on expert knowledge. This is an odd sentence.

Thank you for your suggestion. We replaced 'expert knowledge' with 'Traditional geostatistical models often require presumed parameterization and are constrained by stationarity assumptions' to avoid ambiguity. Please see line 74 in the revised manuscript.

We intended to highlight that traditional geostatistical models (such as Kriging) are often constrained by stationarity assumptions and the requirement to select specific semi-variogram models. These models require subjective decisions regarding parameters (such as the nugget, sill, and range); and assume that spatial autocorrelation follows this form in the study area. Without knowledge in the study area, it is difficult to perform these tasks.

Where soil properties exhibit abrupt transitions or complex non-linearities that violate these assumptions, geostatistical models may fail to capture the true spatial heterogeneity. Applying these models without knowledge about the study area can lead to biased estimations.

Line 118. Improves

Thank you for your correction. We made this correction in the revised manuscript, as 'integrating multiple qualified data sources improves the amount of soil data and reduce prediction uncertainties'.

Line 120. I don't understand this sentence. It does not fit with the rest of the paragraph.

Thank you for your suggestion. In the revised paragraph, we deleted this sentence.

Line 144. Section 2. This section is confusing and I suggest this section is reorganized by first describing the residual correction method, secondly working through a simple example, and thirdly presenting the CA case study and associated data. Step two might not be needed if the authors can more clearly describe the methods, but I would find a clearly worked example helpful for my understanding.

Thank you for your constructive suggestion. This concern is consistent with the comments raised by Referee #3, and we addressed both sets of comments through a substantial revision

of the methods section, including both text and figures. In the revised manuscript, we reorganized Section 2 to:

- first describe the residual correction framework;
- include a worked example using one soil column to demonstrate how the prediction evolves and operates across different iterations;
- then present the California case study and associated datasets as its application.

To avoid an overly lengthy response, we have not included the full revised text here. Please refer to section 2 in the revised manuscript to see our updates. Below, we present the updated figures for illustrating the IRC method.

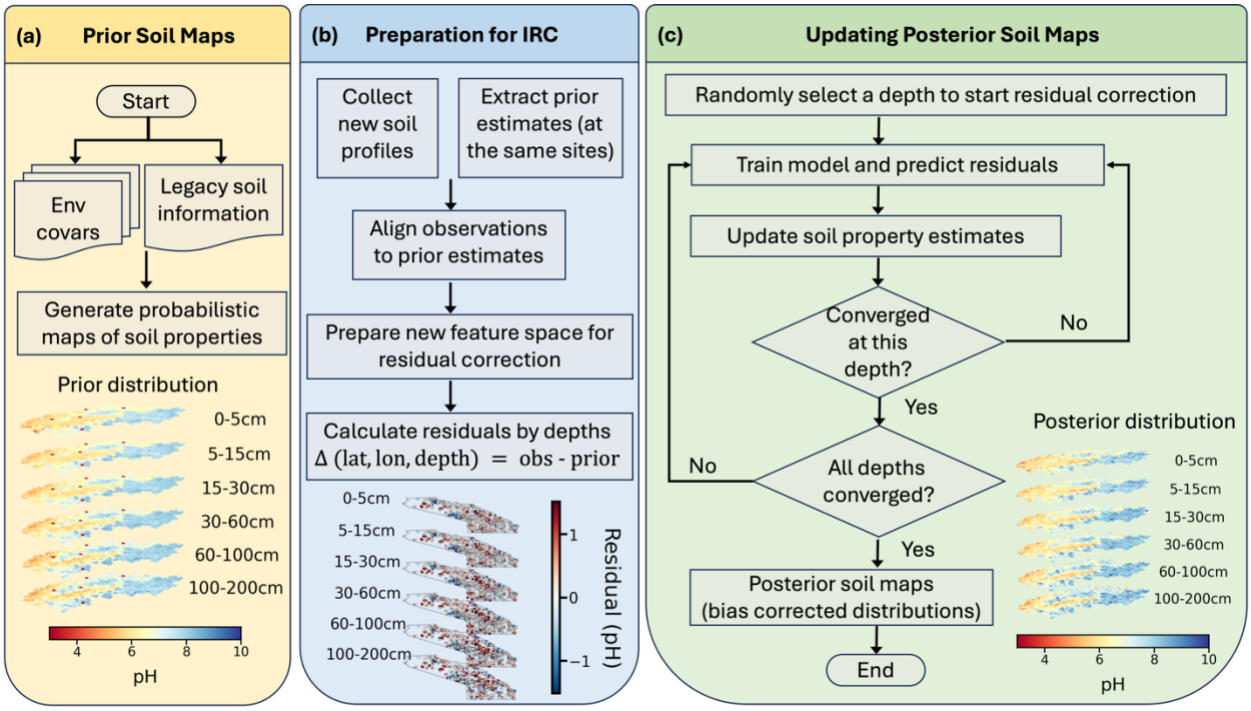


Figure 1: Workflow for updating posterior soil property maps.

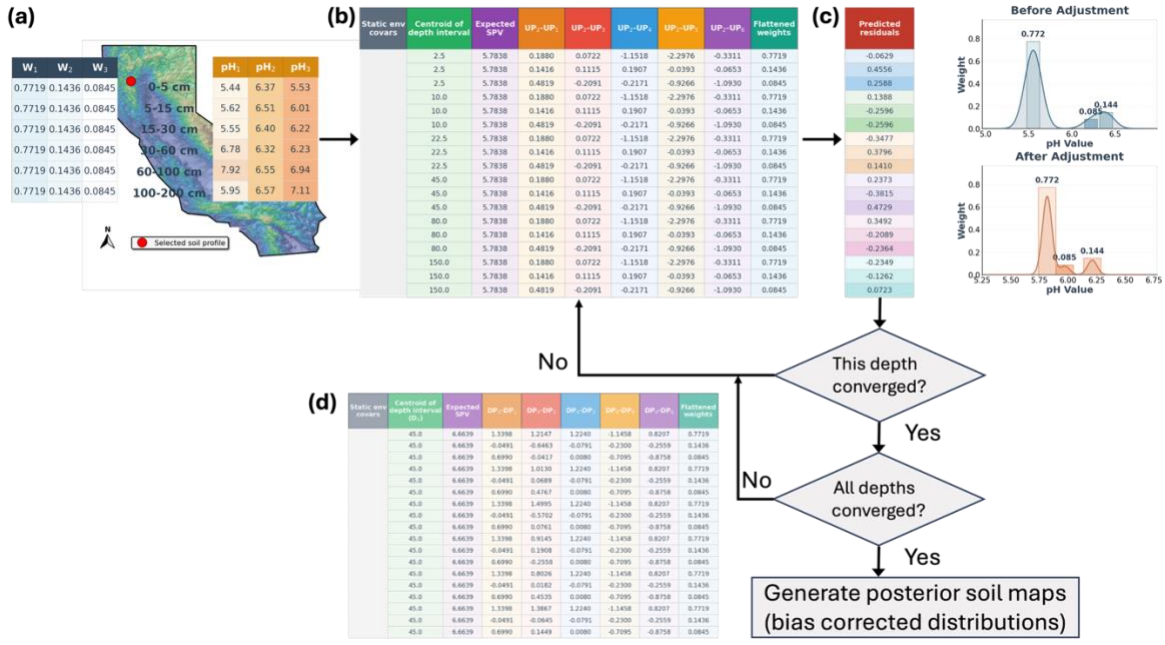


Figure 2: Schematic illustration of the iterative residual correction (IRC) method using a worked example at a randomly selected soil column.

Line 228: Figure 3b/Line 233. What does matching prior soils with new profiles actually mean?

Thank you for your suggestion. "Matching" in this context refers to the process of spatially and vertically aligning the new soil profile data with the prior soil map. For each new ground-truth observation (a soil profile at a known location), we extract the corresponding information from the prior map. This involves:

- Spatial Alignment: Using the geographic coordinates (latitude/longitude) of the new profile to locate the exact corresponding pixel in the prior soil map raster.
- Vertical Alignment: Extracting the prior map's predicted value (e.g., soil organic carbon content) for the specific depth interval that matches the depth of the new observation.

This one-to-one pairing allows us to calculate the residual at that specific point in three-dimensional space (latitude, longitude, and depth). We have incorporated a more detailed and systematic explanation of this procedure in the revised manuscript (Section 2.1.1). To avoid redundancy, we do not repeat the full description here and kindly refer the reviewer to that section for further details.

Line 248: classes should be taxonomic classes.

Thank you for your suggestion. We deleted this sentence in the revised manuscript to address other comments.

Line 253: soil or environmental covariates? Soil covariates are assumed to be other soil attributes.

Thank you for your suggestion. We corrected this terminology to environmental covariates. These features (e.g., topography, remote sensing indices) are environmental covariates. This terminology has been revised consistently throughout Section 2 of the manuscript to avoid ambiguity.

Line 254: Please replace the , with and.

Thank you for your suggestion. In the revised manuscript, we replaced it with 'The covariates include remote sensing data (Sentinel-1, Sentinel-2, GOES land surface temperature) and terrain attributes, identical to those used in the prior mapping method (Xu et al., 2025)' from line 241 to 244.

Line 258: Suggested text: "That prunes less plausible predictions to narrow prediction intervals"

Thank you for your suggestion. We deleted this sentence in the revised manuscript, but in the line 40, 643 and 775, we used narrower PIW to describe the reduced uncertainties.

Line 273: What does soil data mean?

Thank you for your suggestion. In this context, "soil data" refers to soil property values derived from the prior soil map. Specifically, it denotes the prior estimates at a given pixel and depth interval, including the top-probable soil property values used by the model at that stage of the iteration.

We revised the sentence in the manuscript to explicitly distinguish between (i) prior soil map values and (ii) observed soil measurements, to avoid confusion. Please check the updated

section 2, we specifically distinguish prior soil property, soil covariates, and other terms to avoid ambiguity.

Line 277: "Additionally, soil covariates are prepared for residual correction at each depth as feature space for predictive models." This is vague, please explain in more detail.

Thank you for your suggestion. In the revised manuscript, we explained it in section 2.1.1 using a worked example. And Figure 2 also conveys such information. We copy the key information from the revised manuscript to explain how it evolves: 'If either convergence check returns "No" (i.e., D2 has not converged or other layers remain unconverged), the algorithm continues iterating. Here, the soil property values for D2 are updated by adding the predicted residuals to the previous pH values. These updated values are then used to reconstruct the feature space following the same structure described above, updating the representative values, top-probable values, and inter-layer differences. By updating soil covariates layer by layer and iteratively refining the feature space, the next prediction retains prior knowledge while integrating new information about soil heterogeneity and vertical relationships for soil profiles (Wu et al., 2025). A new iteration begins by randomly selecting another unconverged layer, and the process repeats until convergence is achieved across all depth layers.'

Line 282: Are these randomly selected in the algorithm or just for Fig 4?

Thank you for your suggestion. In the algorithm, the starting layer is randomly selected as part of the iterative process. Regardless of which layer is selected to initiate the process, all layers are eventually updated according to the methodology.

Line 284: what variables make up this feature space?

Thank you for your suggestion. The components of feature space include:

- Environmental covariates: such as terrain attributes and remote sensing indices;
- Depth information: median values of the soil horizon interval for the layer under modeling;
- Representative soil property values: expected (mean) values of soil properties at each pixel;
- Top-probable soil property values: multiple most likely predictions from the prior iteration;
- Inter-layer residuals: differences between top-probable predicted values across soil layers;
- Weights: associated with top-probable values, representing their probabilities.

These variables are sampled at the locations of the new soil observations and used to train the Random Forest regressor to predict residuals. This construction avoids direct inclusion of the target residuals (residuals) themselves.

Please refer to section 2.1.1 where we explained this information in more details. Also, we provided visualization of features in the updated Figure 2 (and the figure is provided in a previous answer).

Line 284: How are weights updated? I thought these were weights of soil property-soil taxonomic class weights? Does this mean that the soil taxonomic class predictions are also updated?

Thank you for your suggestion. To clarify, the soil taxonomic class predictions are not updated in this method. The "weights" in the feature space refer to the probabilities associated with the top-probable soil property values and are kept fixed throughout the residual correction process. These weights are used to compute representative values but are not themselves updated by the model. We emphasize this in line 259, as 'These weights remain fixed throughout iterations.'

Line 286: Totally lost by this point. What prior values?

Thank you for your suggestion. "Prior values" refer to the soil property estimates used as the baseline at the start of the iterative residual correction. In the first iteration, these correspond to the values predicted by the pHRF method. In subsequent iterations, the "prior values" are updated to reflect the sum of the previous iteration's residual corrections and the original prior. In line 186, we provided such explanation, as 'Predicted residuals are added to the prior (or previous iteration's) estimates to generate updated soil property values. '

Line 311: What soil observations?

Thank you for your suggestion. In the revised manuscript, "soil observations" refer to the measured soil property values retrieved from the (new) additional georeferenced soil profiles (WoSIS, SCD, and Salinas Valley measurements) that were not used in creating the prior map. The definition of "soil observations" stay consistent throughout the manuscript.

L314: Feature space is refined… Are the environmental covariates also updated/changed?

Thank you for your suggestion. In the revised manuscript, we clarify that 'the feature space for residual modeling is also prepared, consisting of static environmental covariates (which remain fixed throughout iterations) and dynamic soil covariates (which are updated iteratively)' in line 178.

Pleases refer to section 2.1.1 and 2.1.2, where we define the components of feature space include environmental covariates, soil covariates, and weights. The environmental covariates are not updated or changed during the iterative residual correction. Only the soil-related features, such as top-probable values, representative soil property values, and inter-layer residuals, are updated in each iteration. The environmental covariates (e.g., terrain attributes, remote sensing indices) remain fixed and serve as stable predictors for interpolating residuals across the study area. Weights also stay fixed.

L333: soil property distributions?

Thank you for your suggestion. In this context, "soil property distributions" refers to the cell-by-cell distributions of (updated) soil property values across the study region. Convergence is determined based on the median change in soil property values across all pixels: when the median difference between updated residuals and previous residuals falls below a predefined threshold, the iterative updates are considered stabilized. This ensures that further iterations do not largely alter the predictions.

L340: has instead of have.

Thank you for your suggestion. We corrected the word in the revised manuscript.

Fig. 4b. Do these colors mean anything? The rows in 4b seem like they are depth intervals, but these only come from D2? The caption didn't help much.

Thank you for your suggestion. The colors and rows in the original Figure 4b are intended to illustrate the multi-layered structure of the feature space and to differentiate among feature columns. The same color indicates the same type of feature, while different rows represent different values of that feature. For example, if we maintain the top-12 probable soil property values at each pixel, the 12 rows correspond to the 12 values for that pixel and depth.

While the example in Figure 4b uses Depth 2 (D2) for illustration purposes, the algorithm updates one depth layer at a time in each iteration, and all soil layers will all undergo this correction process. During an iteration, residual correction is applied only to the "target layer" (for example, D2), while soil covariates from other layers are used to compute inter-layer differences that capture vertical correlations (this feature is illustrated in a previous answer,

such as D3-D2 or D1-D2). The algorithm then cycles through all depth layers sequentially, ensuring that each layer is updated.

However, please see the updated Figure 2. In the updated figure, we used specific numbers to illustrate changing of feature space.

L343: R3 subscript

Thank you for your suggestion. We already updated subscripts in the revised manuscript.

L387: So these metrics were calculated as the difference between the values used to iterate… I don't understand this.

Thank you for your suggestion. The metrics were computed using out-of-bag (OOB) samples, which are ground-truth observations not used in training the Random Forest regressor. For each OOB sample, we compare the observed soil property value with the expected value of the final prediction (results obtained from the last prediction; aka posterior soil properties) at that location.

To calculate the posterior prediction, at each pixel, we maintain multiple top-probable soil property values from the previous iteration, each with an associated weight. The Random Forest regressor predicts residuals for each of these values in the current iteration. Note the Random Forest regressor is re-trained in each iteration. We then compute the weighted sum of the prior value plus predicted residual to obtain the expected soil property value for that pixel. This expected value is what is compared to the OOB sample to calculate the performance metrics. The reported metrics are computed using all OOB samples and their corresponding expected predictions to summarize the model performance across the study region. In line 387, we explained it as 'Additionally, these metrics are evaluated by comparing the expected values of posterior predictions with co-located soil properties values; not computed on residuals.'

L403. Greek rho instead of p?

Thank you for your suggestion. We used the Greek letter rho to represent the Pearson correlation coefficient, rather than the letter P. We ensured this term is consistently used throughout the revised manuscript.

Reviewer 3:

This paper presents a method to refine predictions of DSM models in a non-parametric manner. I think it may be a significant contribution to DSM and pave the way for promising avenues, such as updating and improving soil maps when newly collected data becomes available. Nonetheless I found the methodology hard to understand (especially how "updated probabilistic maps of soil properties" are obtained and how the "optimization of the feature space" is carried out) and I am concerned it would be hardly reproducible unless the code is made available with a working example. I would also highly recommend the authors to present a minimal example in the paper, on a couple data points.

But I do not think additional experiments or methodological changes are needed: I would therefore recommend to accept the manuscript after some minor revisions.

Comments:

I found the introduction very well written and well supported by extensive references, clearly showing where the current research gaps lie.

Thank you very much for your thoughtful and constructive review, and for recognizing the potential contribution of this work. We agree that the original section 2.2.2 did not sufficiently explain how the posterior soil property maps are obtained. In response, we are pleased to address the points raised:

- **Clarification of Methodology**: We largely revised the method section (section 2) to more clearly explain how the iterative residual correction is implemented and how the updated probabilistic maps are generated. We also gave a worked example using one soil column, in section 2.1.2, to explain this method, especially for explaining how we perform iterative update of feature space.

  We also acknowledge that the term "optimization of feature space" may be misleading, as no explicit optimization objective is used in this process; accordingly, we replaced it with "iterative update of the feature space" throughout the revised manuscript.

- **Code Availability**: We agree that reproducibility is important. We shared the code via this public GitHub repository: https://github.com/emmaxu43/IRC_CA/tree/main

We will also address these points in more detail in our subsequent, itemized responses to the reviewer comments, where we explain precisely how each suggestion has been addressed. We thank the reviewer for recommending acceptance.

Section 2.1: Your method looks very sensitive to discrepancies and biases across different spatial sources. Therefore I would suggest adding a subsection where you explain more clearly how you ensured that the datasets were consistent with each other (include what you replied to Reviewer #1), and clearly indicate how you divided your data into training/test sets.

Thank you for your suggestion. We added a paragraph to explain how we ensure the consistency of different sources of soil data. As stated in the revised manuscript, 'To ensure consistency across different data sources, we applied several quality control steps. First, we checked the physical plausibility of all soil property values by defining a valid range with specific minimum and maximum thresholds for each property… After quality check, the datasets are compatible because the WoSIS records for California are largely derived from the NCSS database, and both the SCD and WoSIS datasets follow standardized laboratory protocols, such as those from the Kellogg Soil Survey Laboratory (Soil, 1996; Soil Survey Staff, 2014)' describes the data blending and quality check for different soil datasets.

Regarding the division of data into training and test sets, we employed out-of-bag (OOB) samples. We explained this in the revised manuscript, as 'Model training and evaluation were performed using out-of-bag (OOB) sampling, with OOB samples (samples withheld from the training process…and the same set of OOB samples remains excluded throughout to ensure independent validation.' We also mentioned the evaluation was performed on values of soil property, instead on residuals in the manuscript.

Section 2.2.2. In my opinion, this section does not explain how posterior soil properties maps are obtained, but simply how some values are updated based on some available ground-truth data. Until I saw the reply you made to Referee#1, I had not understood either how the method applied to pixels with no ground truth observations. You say that you interpolate the residuals by using a model which learns the correlations between the residuals and other environmental covariates: what model is it? Another pHRF? It should be explained in this section.

Thank you for your suggestions. We largely revised the section 2.1 to clarify how the posterior maps are obtained. The posterior soil property maps are obtained through the following steps:

(1) Residual computation: At locations where new soil observations are available, we compute residuals as the difference between the newly observed soil property values and the corresponding values extracted from an existing, spatially continuous prior soil map.
(2) Residual modeling: A Random Forest (RF) regressor is trained to learn the relationship between these residuals and a set of environmental covariates.

(3) Residual interpolation and posterior update: The trained RF model is applied over the entire study area to generate a spatially continuous residual correction map. The posterior soil property map is then obtained by adding this residual map to the prior soil property map. Only values of soil property values are updated. Their weights stay the same.

Regarding the applicability of the method to pixels without ground-truth observations, we acknowledge that our earlier response to Referee #1 may have caused confusion, and we will clarify this point in the revised manuscript. In this study, "pixels with no in situ observations" refers to pixels that do not have co-located soil measurements, but there are soil observations existing nearby. A prior estimate of soil property value can be obtained using spatial interpolation for this specific point or pixel. The proposed residual correction method requires the existence of an initial soil property map (the prior) as a model input.

Importantly, our method does not aim to infer soil properties in regions where no soil observations exist anywhere in the study area. Such a scenario falls outside the scope of this work. Instead, the contribution of our approach lies in updating and improving an existing soil property map when new soil observations become available, by learning and performing residual corrections to improve the estimates of soil properties.

l.273: "By adding these residuals to the prior distributions, the statistical shape of the probability distribution is adjusted (updated property; UP)." Are the residuals really added to prior distributions, or to prior values? From what I understood, residuals are added to prior values.

Thank you for your suggestion. You are correct: residuals are added directly to the prior values at each pixel. At each pixel or sample location, the posterior value is obtained by arithmetic addition of the predicted residual to the prior estimate, while the underlying weights remain unchanged. In line 775, we change the content to 'calculated and added to the prior values of soil property to adjust the statistical shape of the probability distribution pixel-by-pixel.'

At the same time, we clarify that although residuals are added at the value level, this operation may have a direct consequence on the statistical shape of the resulting probability distribution. Shifting individual values, without modifying the weights, can change the range and spread of the posterior distribution. For example, in Figure 7, the posterior soil maps exhibit an expanded range of soil property values compared to the prior maps: lower values become lower and higher values become higher. This leads to an altered distribution of probability mass across value bins and therefore alters the overall distributional shape.

l.277: "Additionally, soil covariates are prepared for residual correction at each depth as feature space for predictive models." What does this preparation involves? And more generally, I find your use of the term "feature space" quite confusing.

Thank you for your suggestions. We agree that we did not clearly distinguish between static environmental covariates and quantities derived from intermediate model predictions, which led to an ambiguous interpretation of what constitutes the feature space and how it evolves during the iterative procedure.

To address this issue, we revised section 2.1.1 and 2.1.2 by defining the feature space at the beginning of the section. In this work, the feature space is defined as the combination of (i) static environmental covariates inherited from the pHRF method, which remain fixed throughout the procedure, and (ii) prediction-derived features (e.g., current soil property estimates, and inter-layer difference of soil property values; as shows in Figure 2); and (iii) weights of soil property values that are updated iteratively after each residual correction step.

We also revised the terminology throughout the section by replacing "optimization of feature space" with "iterative update of feature space", to more accurately reflect the fact that no explicit optimization objective is defined over the feature space itself.

In addition, we suggested the reviewer to read our updated section 2. We large reconstructed this section to address your comments overall.

Section 2.2.2.1 Similarly, I don't understand what you mean by "optimization of feature space".

Thank you for your suggestion. By "optimization of feature space," we were referring to the iterative process of updating and interacting features (static environmental covariates and other updated soil covariates illustrated in the previous answer), as the model progresses.

Rather than using a static set of predictors, the feature space is updated at each iteration to incorporate the most recent updated predictions (UP), allowing the model to capture incremental adjustments and vertical relationships between soil layers that are not reflected in the initial prior map. In addition, we rephrased this in the manuscript as "iterative update of feature space." as mentioned in the previous answer.

In the revised manuscript, in Figure 2, we compared differences between feature in iteration n and iteration n+1 and show in detail what are the components within the feature space (in the feature tables). And the Figure 2 is provided in a previous comment in the response.

l.301 I am very confused by these newly introduced terms: what are representative soil property values and how are they obtained? How do they relate to prior model predictions and/or ground-truth data?

Thank you for your suggestions. The representative soil property value at each pixel corresponds to the expected value (mean) of the soil property distribution, representing the

current best estimate for that location. For a single pixel, there is only one representative value of soil property. In the feature space, for example, if we maintain the top-12 predicted values at a pixel (and 12 weights), all 12 entries along this feature column will have the same representative value. And this value is calculated as the weighted mean of these 12 values.

In line 247, where this term first appears in the method section, we provided definition for it, as 'Representative soil property values (1 dimension): The expected value (weighted mean) of the soil property at each pixel in the modeling layer, representing the current best estimate. This is computed as the weighted sum of top-probable values.'

During each iteration of the Random Forest regression, the model aims to minimize the differences, that is, the difference between the ground-truth observations and the current representative values (predicted residuals plus prior soil property values). Prior predictions are used to compute the representative value at the start of each iteration; for example, in the first iteration, the representative value equals the prior prediction plus the first predicted residual. Subsequent iterations update the representative value by adding newly predicted residuals, progressively refining the soil property estimate at each pixel. From here, you see both prior predictions and ground-truth data are used in the calculation.

l.303 Aren't "top-probable soil property values" prior model predictions?

Thank you for your suggestions. While the top-probable soil property values are related to prior predictions, they are not identical. The prior predictions provide the initial estimate of soil properties at each pixel before any residual correction. During the iterative process, the top-probable values are updated predictions that incorporate both the prior and the residuals predicted by the Random Forest regressor at each iteration.

In other words, prior predictions reflect the previous estimate of soil property values; and top-probable soil property values means adjusted estimate of soil property values (residuals plus priors).

In the revised manuscript, in line 252, we provided the definition of top-probable soil property values, as 'Top-probable soil property values (1 dimension): The current predictions at each pixel (residuals plus previous prediction of soil property values), reflecting both intra-pixel and inter-pixel soil heterogeneity.'

l.306 What is the "target layer" ?

Thank you for your suggestions. In the revised manuscript, we used 'modeling layer' to replace the term "target layer". It refers to the specific standardized soil depth interval (e.g., 0–5 cm, 5–15 cm) that is currently being processed and updated by the model during a given iteration. In our method, residual corrections are applied depth by depth. During each iteration, residuals

are adjusted for only one layer, while information from other layers is used to provide vertical context.

Because our method processes soil properties depth-by-depth, the "target layer" is the one receiving the residual correction, while "other layers" refer to the other vertical intervals used to provide context of vertical profiles. In the revised manuscript, we explicitly emphasize it in line 233, as 'Only one layer is modeled and updated for a given iteration.'

What are the dimensions of this "feature space"? And, from my machine-learning experience, the feature space does not include the target variable, only the explanatory features/variables. It seems also very confusing to me that the residuals are part of this feature space.

Thank you for your suggestions. We use D to denote the dimensions of the feature space. The total dimensions is 30. The feature space is composed of the following components:

$D_{feature} = D_{static\ env\ covar} + D_{depth} + D_{mean\ spv} + D_{top\ spv} + D_{inter\text{-}layer\ difference} + D_{weights}$

$D_{static\ env\ covar} = 21$

These are the environmental covariates used in the pHRF framework.

$D_{depth} = 1$

It is the median values of the soil horizon interval for the soil layer under modeling (current layer).

$D_{mean\ spv} = 1$

It is expected value of soil property value (spv) at each pixel.

Dtop spv = 1

It is top-probable soil property value. On each pixel, we can, for example, have top-12 soil property values. And we can flatten all these values and put them within one feature column.

Dinter-layer difference = 5

With six depth intervals, the residuals between the current layer and the other five layers are included to capture vertical variation.

Dweights = 1

The weights associated with each top-probable soil property value.

Summing these components, the total dimension of the feature space is 30.

The feature space does not directly include the target variable for the current layer (i.e., the residuals being predicted). Instead, it includes features derived from prior information and they are not directly collinear with the current target variables, such as:

- Residuals plus prior soil property values (i.e., top-probable soil property values from previous iterations), computed by adding the predicted residual from the previous iteration to the prior prediction.
- Expected value of top-probable soil property values, computed as the weighted sum of the top-probable values at each pixel.
- Differences in top-probable predicted soil property values between layers, calculated as the difference between top-probable values of the current layer and other layers.

These features are not simply linear transformations of the target variable, and they capture incremental adjustments and vertical relationships. Target variable is residuals.

Figure 4 is full of details that I found a bit confusing, e.g. the "prior distribution" at a random pixel, that is unused, uncommented and does not refer to either A or B. The weights, which are a methodological detail that is not properly explained and thus remains unclear to me. I would highly suggest updating this figure using a more concrete example, with the prior and posterior distributions at 2 real points A and B, at two different depths.

Thank you for your constructive suggestions. We agree that Figure 4 in the original manuscript is dense and may be confusing. In the revised manuscript, we update Figure 1 and 2 to make it more concrete and interpretable. Specifically:

- In Figure 1, we visualized spatial patterns of prior and posterior soil maps over CA. In Figure 2C, we compared "prior" and "posterior" distributions change for the chosen pixel at one depth interval, visualizing how the residual correction updates soil property estimates.
- The weights associated with top-probable values were explicitly annotated and explained, as displayed in Figure 2b and 2d.

For more details, please refer to the updated section 2 and previous comments, since this is a mutual request from the reviewer 2.

Section 2.2.2.3: Using this minimal example, you could show how and at which step this "optimization with constraints" takes place.

Thank you for your suggestion. In the revised manuscript, we updated section 2.1.3 and used an example of sand content to illustrate it, as 'During residual correction, a common issue arises where the addition of residuals to prior soil property values results in values that exceed physical bounds (such as sand content > 100%). To address this, a residual update process with constraints is implemented.

As illustrated in Figure 2c to 2d, after the Random Forest regressor predicts residuals for the layer (D_2), these residuals are added to the previous soil property values to generate updated predictions. Immediately after this addition step, the updated values are examined to check whether they fall within predefined physical bounds (e.g., 0% to 100% for particle size fractions, positive values for bulk density). This constraint check occurs before the convergence evaluation and before the updated values are used to reconstruct the feature space for the next iteration.

If any updated value exceeds the physical bounds, it is adjusted to the nearest valid bound (minimum or maximum). For example, if adding a residual of +15% to a prior sand content of 90% yields 105%, this value is capped at 100%. The "excess" residual (+5% in this case) is then redistributed proportionally (based on their weights) among the other top-probable values at the same pixel, ensuring that the total correction remains consistent with the model's prediction while maintaining physical plausibility. For particle size fractions (sand, silt, clay), an additional compositional constraint ensures that the three fractions sum to 100% at each pixel after residual correction.'

Table 1 : It would be insightful to know how many ground-truth data points were used to refine predictions for each soil property.

Thank you for your suggestion. The number of ground-truth samples varies across different soil depths. Including these detailed sample counts in Table 1 would make it lengthy. To address your comment, we have compiled these layer-specific sample sizes in the Supplementary Material as Table S2.

Depth interval (cm)	Sand	Silt	Clay	pH	SOM	BD
0–5	3,856	3,920	3,083	2,806	2,048	2,254
5–15	3,607	3,621	2,995	3,885	2,770	3,283
15–30	2,044	2,036	1,814	4,247	3,141	3,256
30–60	1,348	1,333	1,305	4,909	3,588	4,217
60–100	1,048	1,047	1,045	4,545	3,306	3,732
100–200	905	896	866	3,717	2,616	2,814
Total	**12,808**	**12,853**	**11,108**	**24,109**	**17,469**	**19,556**

Figure 5 is very insightful and well commented. Similarly, the histograms on Figure 7 that clearly show the fact that the distribution is more spread-out than prior predictions.

We appreciate your positive comments. These maps show that the posterior soil property maps, after residual correction, reveal more spatial variability compared to the prior predictions, capturing soil heterogeneity that is smoothed out in the prior maps.

Citation: https://doi.org/10.5194/egusphere-2025-5107-RC3

Reference

Arrouays, D., McKenzie, N., Hempel, J., Forges, A.R. de, McBratney, A.B., 2014. GlobalSoilMap: Basis of the global spatial soil information system. CRC Press.

Hartemink, A.E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R.A., Minasny, B., Montanarella, L., de Mendonça Santos, M.L., Sanchez, P., Walsh, M., Zhang, G.-L., 2010. GlobalSoilMap.net – A New Digital Soil Map of the World, in: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer Netherlands, Dordrecht, pp. 423–428. https://doi.org/10.1007/978-90-481-8863-5_33

Soil, K., 1996. Survey laboratory methods manual. Soil Survey Investigations Report.

Soil Survey Staff, 2014. Kellogg Soil Survey Laboratory methods manual (No. Soil Survey Investigations Report No. 42). U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln, Nebraska.

Xu, C., Huang, J., Hartemink, A.E., Chaney, N.W., 2025. Pruned hierarchical Random Forest framework for digital soil mapping: Evaluation using NEON soil properties. Geoderma 459, 117392. https://doi.org/10.1016/j.geoderma.2025.117392

Summary of Key Changes in the Revised Manuscript (change list)

We thank the reviewers for their constructive comments, which have substantially improved the clarity and quality of our manuscript. Below, we summarize the major revisions made in response to reviewer feedback.

1. Reorganization of Section 2 (Methods)

We have restructured Section 2 to improve logical flow and readability:

(1) Section 2.1 now presents the general iterative residual correction (IRC) framework, including a step-by-step explanation of the three main components (prior map generation, residual preparation, and iterative correction).

(2) Section 2.1.1 provides a worked example using a randomly selected soil column to illustrate the key steps and components of the IRC method, making the approach more accessible.

(3) Section 2.2 presents the California case study, including detailed descriptions of soil datasets (Section 2.2.1) and implementation details specific to this application (Section 2.2.2).

This reorganization addresses reviewer requests to first describe the framework conceptually, then provide a worked example, and finally present the case study application.

2. Clarification of Method Applicability and Scope

We have clarified that the IRC method is generalizable beyond pHRF and can be applied to other spatially continuous probabilistic soil map:

(1) Added a definition of "prior probabilistic soil property maps" in Section 2, explaining that these refer to any spatially continuous soil maps providing prior estimates with associated uncertainties (not limited to pHRF or POLARIS products).

(2) Emphasized in the abstract and introduction that while this study demonstrates the method using pHRF-derived priors, the IRC framework can correct bias in other existing probabilistic soil maps.

(3) Clarified that the method does not require new observations to be co-located with pixels used to create the prior map, but rather requires a spatially continuous prior map and new soil observations at some locations for model training.

3. Enhanced Figure Quality and Clarity

(1) Figure 1 (Workflow diagram): redesigned this figure to better illustrate the framework

(2) Figure 2 (Worked example): redesigned this figure and only used one soil column as a simple example to represent evolution across different iteration

(3) Figure 4 (formerly Figure 2 - Salinas Valley sampling): inset the location of sampling place in California

(4) Figure 5 (Performance comparison)

a. Added units directly to axis labels for all soil properties (mass %, g/cm^3).
b. Clarified in the caption that SOM is plotted on a $\log_{10}$ scale, explaining why negative values appear (representing SOM < 1%).

4. Addition of Supplementary Information

We have created supplementary materials to provide additional detail without overcrowding the main text:

(1) Table S1: Depth-specific model performance metrics (R^2, RMSE, bias) for each soil property at all six depth intervals.

(2) Sample size information: Added description of sample sizes vary by property and depths in California.

5. Code sharing

We have updated the "Code and Data Availability" statement:

Provided a public GitHub repository containing the complete code for reproducing our analysis: https://github.com/emmaxu43/IRC_CA/tree/main

Point-by-point responses and minor revisions are provided at the beginning of this response document. Please refer to these answers for more details. Thank you for your time and contribution to make this work better.

Chengcheng Xu

On behalf of other co-authors