Improvement of Soil Properties Maps using an Iterative Residual Correction Method

Chengcheng Xu[1], Elia Scudiero[2,3], Ray Anderson[3,2], Nathaniel Chaney[1]

[1] Department of Civil and Environmental Engineering, Duke University, Durham, NC 27705, USA

[2] Department of Environmental Sciences, University of California Riverside, Riverside, CA 92521, USA

[3] United States Department of Agriculture – Agricultural Research Service, George E. Brown Jr. Salinity Laboratory, Agricultural Water Efficiency and Salinity Research Unit, Riverside, CA 92507, USA

Correspondence to: Chengcheng Xu (Chengcheng.xu@duke.edu)

Short Summary

Accurate soil information is vital. This study developed a method to improve existing probabilistic soil maps, spatially continuous maps providing prior estimates, by correcting their probability distributions as new soil data emerges. By iteratively adjusting previous predictions, the method increases both accuracy and certainty of soil maps. Its application in California enhanced predictions for several soil properties. This method can be further used for more soil properties and regions.

Abstract

Accurate mapping of soil properties is vital for many applications, yet existing models for digital soil maps often underestimate their spatial variability or prediction uncertainties,

which introduces risk for applications such as irrigation and drainage management. This study introduces an approach, iterative residual correction (IRC), to update existing probabilistic soil maps when new soil observations become available. We demonstrated its application for enhanced soil mapping performance using a Californian case study. To implement this, we first generate prior probabilistic soil property maps using a pruned hierarchical Random Forest (pHRF) method. These prior estimates are then refined by integrating additional soil profile data and iteratively adjusting residuals of distribution of soil properties (reducing differences between observations and prior predictions) pixel by pixel. For this purpose, we employed Random Forest regressors to gradually adjust the soil property distributions and incrementally correct prior bias. Updated soil maps were evaluated over California and at 1-km resolution to test the methodology, using additional soil observations from the World Soil Information Service, the Soil Characterization Database, the University of California Riverside, and the United States Department of Agriculture Agricultural Research Service. Posterior soil texture predictions achieved an RMSE below 10, a 7% relative reduction in errors (mass fraction of the fine-earth fraction) over priors. RMSE and spatial representation for soil organic matter and bulk density also improved. Furthermore, the method reduced prediction uncertainties (narrower prediction intervals compared to the priors) and enforced physical constraints on soil property bounds. Looking forward, this IRC method offers a scalable pathway to improve existing probabilistic soil maps, providing a strategy for the evolution of digital soil products as new soil observations emerge.

1 Introduction

Soils play an important role in regulating Earth's water, energy, and nutrient cycles (Vereecken et al., 2016). Soil maps guide agricultural practices, ecosystem management, hydraulic modeling, and climate studies, such as crop modeling, flood risk assessment, groundwater management, and climate change (Vereecken et al., 2022). The importance of soil maps has increased with the advent of precision agriculture, including site-specific seeding, irrigation, and fertilization recommendations that intrinsically depend on high-resolution soil properties (Jiang et al., 2011; Li et al., 2019; Mueller et al., 2001; Ortuani et al., 2016). However, the accuracy and reliability of these management actions heavily depend on the quality of soil maps as a critical decision-making input. Traditional soil surveys involve field observations, laboratory analyses, and expert interpretation, but are labor-intensive and expensive (Grunwald et al., 2011; Rossiter et al., 2022; Soil Survey Staff et al., 2023). These limitations have driven the development of digital soil mapping (DSM) techniques. DSM leverages decades of soil data collection and sharing, establishing quantitative models to generate georeferenced soil maps (McBratney et al., 2003).

Digital soil maps are typically derived from existing soil surveys, geostatistical models, machine learning, or hybrid approaches. Soil survey-based soil mapping method, which use low, high, and representative values to describe soil property distributions for each soil component (Soil Survey Staff et al., 2023). The method typically approximates each soil component as a triangular distribution (Chaney et al., 2016; Soil survey staff, 2023), potentially oversimplifying multi-modal distributions of soil properties in some cases

(Haghverdi et al., 2020; Nussbaum et al., 2023). Additionally, estimating soil properties from synthetic sampling within a map unit could create artificial spatial patterns, adding noises into the mapping results (Chaney et al., 2019). Developments such as Latin-hypercube sampling and landscape adaptive covariance functions have improved the representation of spatial patterns of soil properties (Minasny and McBratney, 2006). Yet, soil survey-based approaches remain valuable particularly in areas where soil profile data is limited (Nauman et al., 2024). Geostatistical models often require presumed parameterization and are constrained by stationarity assumptions, which is difficult to apply in areas with insufficient field knowledge (Oliver and Webster, 2014). To address these challenges, non-parametric models, such as Random Forest, trained with hybridized soil data that combine soil surveys with georeferenced soil profiles show potentials in improving soil mapping, particularly for large-scale maps (Chaney et al., 2019; Nauman et al., 2024).

Map of soil properties have been observed with bias compared to field observations in certain areas due to many factors (Hengl et al., 2017; Powers et al., 2011). At the measurement level, sampling methods may favor certain landscape positions or soil conditions, causing a clustered representation (Ramcharan et al., 2018). In areas with coarse sampling density, models trained on unrepresentative data are likely to deviate from actual observations (Sharififar et al., 2019). Commonly used DSM models can show bias. For example, Random Forest classifier favors the majority class (Chen et al., 2004), and Random Forest regressors struggle to capture extreme values (Nauman et al., 2024).

Furthermore, certain areas may not be fully captured by the DSM model and the selected feature space, such as areas with complex glacial pattern, parent material transitions, and alluvial processes (unaddressed problem in SOLUS; SoilGrids 2.0; (Nauman et al., 2024; Poggio et al., 2021)). Model-based solutions include using ensemble models to enhance accuracy compared to a single model (Sylvain et al., 2021). Post-processing methods, such as regression kriging and bias-corrected decision trees, can also be used (Hengl et al., 2004). Yet, kriging-based methods are limited by second-order stationarity assumptions, where the spatial covariance structure is assumed to be invariant across the study area, limiting its efficacy in heterogeneous landscapes. In the presence of abrupt environmental transitions, a stationary "global" variogram can fail to account for local spatial non-stationarity, often resulting in "bull's eye" artifacts around isolated observations (Minasny and McBratney, 2016). Non-parametric models can be used for bias correction that overcome the limitation of making presumed distributions and can adapt to spatially heterogeneous landscapes. However, their data-driven nature makes quantifying how prediction confidence varies spatially important. Without uncertainty quantification, users cannot assess location-specific reliability, limiting their utility in practical decision-making (Schmidinger and Heuvelink, 2023).

DSM products represent soil properties as multi-dimensional matrices showing vertical and horizontal soil variation (Vereecken et al., 2022), with each pixel containing weighted possible values and their prediction uncertainties. These uncertainties can be represented either as continuous values through prediction intervals or as discrete classifications with

111 associated class probabilities (Chaney et al., 2016, 2019; Hengl et al., 2017; Ramcharan et
112 al., 2018). Common quantification approaches include geostatistical techniques like
113 kriging, where the nugget term accounts for measurement errors while kriging variance
114 reflects spatial uncertainty patterns (Chilès and Delfiner, 2012; Takoutsing et al., 2022),
115 and machine learning methods such as Quantile Random Forest (QRF) which generates
116 probability distributions from decision tree outputs using values of soil properties (Poggio
117 et al., 2021; Shi et al., 2024). For discrete classifications, uncertainty derives from soil
118 raster probabilities during soil taxa classification (Chaney et al., 2016; Odgers et al., 2015).
119 Given the data-driven nature of DSM and frequent limitations in soil profile availability,
120 integrating multiple qualified data sources improves the amount of soil data and reduce
121 prediction uncertainties (Nauman et al., 2024), particularly in regions where predictions
122 must rely more heavily on legacy soil data.
123
124 Taken together, these limitations point to a research gap. Existing DSM products contain
125 biases that are difficult to correct with current methods. Kriging-based approaches require
126 stationarity assumptions and careful variogram parameterization. Non-parametric
127 regressors such as QRF can model residuals, but they operate on numerical inputs alone.
128 They cannot directly incorporate categorical soil survey information (soil taxa and map-unit
129 estimates). Also, QRF often runs only as a single-pass simulation. Bayesian updating offers
130 an alternative but still requires specifying prior distributional forms and likelihood
131 parameters. There is a need for a framework that integrates legacy soil survey (map unit-
132 based estimates of soil property, georeferenced soil taxa) and georeferenced pedons to

update posterior distributions and improve soil mapping performance, without using synthetic sampling, distributional assumptions, or site-specific parameterization. To resolve these limitations, we present a hybrid DSM approach combining pruned Hierarchical Random Forest (pHRF) with iterative residual correction (IRC) method (Xu et al., 2025). The pHRF method leverages the National Cooperative Soil Survey (NCSS) soil survey data and georeferenced soil taxa information to generate prior distributions, while additional soil profiles correct biases in prior predictions. This method builds on development in previous research while addressing specific limitations. Sylvain et al. (2021) applied XGBoost (sequential decision trees) and ensemble models to correct deterministic soil property maps, demonstrating reduced bias for many soil properties (Sylvain et al., 2021). Zhang et al. (2010) introduced a bias-correction technique with Random Forest models to mitigate their tendency to regress toward mean values, though not in DSM contexts (Zhang and Lu, 2012).

Building on these foundations, our approach extends these concepts by probabilistically updating posterior distributions at each location through an iterative correction process that continues until convergence across vertical intervals. Vertical soil profiles are respected through layer-by-layer residual correction. Specifically, after correcting each depth layer, its updated soil property values replace the original feature column for that layer when correcting the next layer down. This preserves inter-layer correlations while dynamically updating the feature space at each step. Unlike methods requiring distributional assumptions, our non-parametric framework adapts to different landscapes

and data scenarios. The models implement residual correction by minimizing the differences between priors and new observations to adjust posterior distributions (one iteration corrects only one soil layer), with the entire process continuing until property variations stabilize between different iterations. In the California case study, the stopping criterion is declared when the median change in residuals drops below the 5th percentile of the residual distribution observed across the last three iterations for each depth layer. Accordingly, this study has four main objectives: (1) To develop the IRC framework as a non-parametric method for updating posterior distributions of soil properties by iteratively correcting residuals between previous predictions and georeferenced soil pedons, without synthetic sampling or distributional assumptions; (2) To implement and evaluate IRC using six soil properties: sand, silt, clay, pH, bulk density, and soil organic matter across California, assessing predictive performance using RMSE and R^2 and uncertainty quantification; (3) To assess the added value of iterative correction by comparing model performance relative to both the pHRF prior and a single-pass (non-iterative) residual correction baseline; (4) To study the spatial structure of posterior predictions by comparing semi-variograms and predicted soil profiles from prior and posterior maps. While this study focuses on California, this soil mapping framework can be potentially applied to the Contiguous United States (CONUS) in future work.

2 Methods

This study introduces a hybrid framework for digital soil mapping (DSM) that updates existing probabilistic soil property maps using newly collected soil observations. The

framework combines prior soil property estimates with an iterative residual correction (IRC) method. The IRC method integrates additional georeferenced soil profiles (soil observations not used to train prior soil maps) and employs non-parametric models to adjust the distribution of prior estimates, thereby correcting biases in the prior soil maps.

The following sections first describe the general residual correction framework (Section 2.1). To illustrate the method concretely, we then provide a worked example using one randomly selected soil column to demonstrate how the feature space is constructed and updated across two consecutive iterations (Section 2.1.1). Building on this example, we detail the key components of the IRC method: the iterative update of feature space (Section 2.1.2), the convergence criterion for residual correction (Section 2.1.3), and the process for updating posterior soil properties with physical constraints (Section 2.1.4). Finally, we present the California case study (Section 2.2), describing the soil datasets used (Section 2.2.1) and the implementation details for applying the IRC method over California (Section 2.2.2).

2.1 Iterative Residual Correction Framework for DSM

Residual correction is implemented to address underestimated soil property variation in prior maps (tendency to underestimate high values and overestimate low values, smoothing out soil variation across landscape). The overall workflow of the IRC method consists of three components: (1) prior map generation (Figure 1a), (2) residual preparation (Figure 1b), and (3) iterative correction (Figure 1c).

First, probabilistic prior soil property maps are generated or retrieve probabilistic soil property maps from an existing DSM product as the prior soil maps (Figure 1a). These maps represent the initial estimates of soil properties and their associated uncertainties. Second, a residual preparation step is carried out to enable correction using new soil profile observations (Figure 1b). The preparation involves four key steps: (1) adding additional soil profiles from new field measurements or databases; (2) spatially aligning these profiles with the corresponding pixels in the prior soil maps using geographic coordinates; (3) vertically aligning observations with prior predictions at matching depth intervals; and (4) calculating residuals depth by depth as the difference between observed values and prior predictions. During this stage, 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). Detailed construction of the feature space is described in Section 2.1.1.

Finally, iterative residual correction is performed to update soil property estimates across depths (Figure 1c). During each iteration, the model predicts residuals for one depth layer at a time, with the layer selected randomly. A Random Forest regressor is trained to learn the relationship between residuals and the feature space at sampled locations, then interpolates residual corrections across the study area. Predicted residuals are added to the previous iteration's estimates to generate updated soil property values. After each update, convergence is evaluated for the modeling depth by comparing the median

difference between the current residuals and those from the previous iteration. Once this change falls below a predefined threshold, that depth is considered converged and excluded from subsequent updates. The algorithm then focuses on the remaining "unconverged" depths, until convergence is achieved across all layers. After convergence is verified for all depths, the final corrected residuals are added to the prior estimates to update the posterior distributions of soil properties.

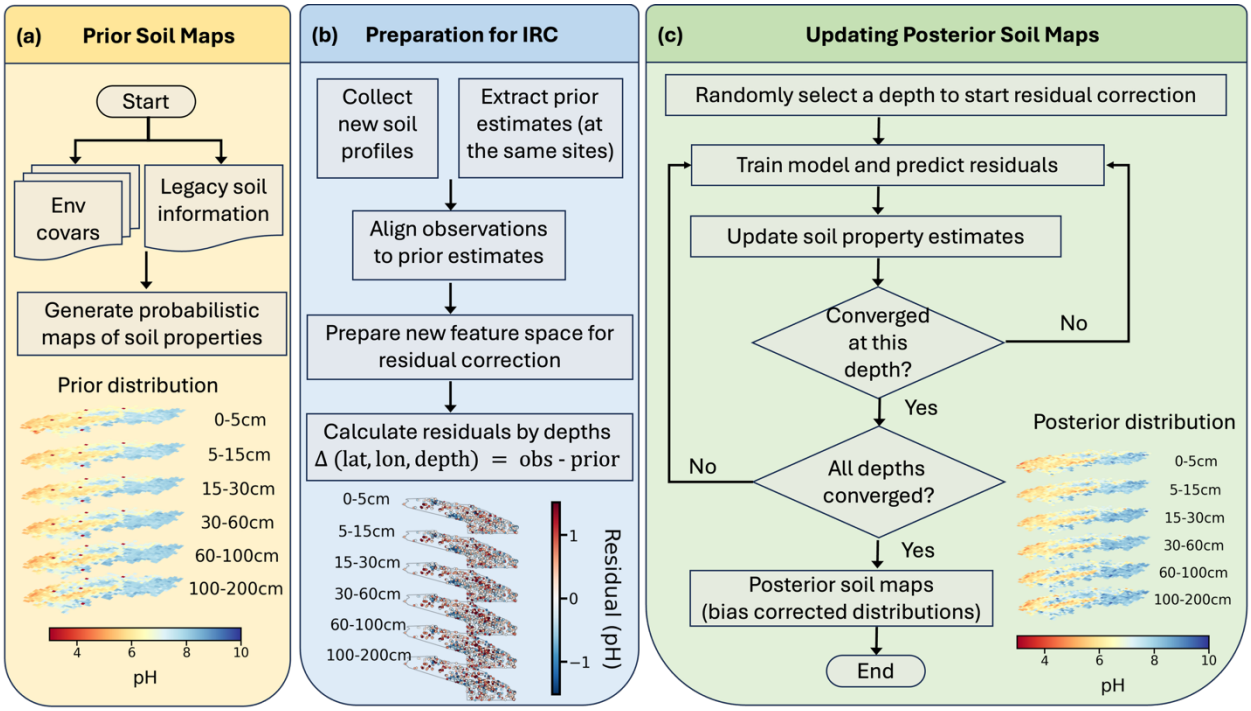


Figure 1: Workflow for updating posterior soil property maps. The process begins with panel (a), the preparation of environmental covariates (env covars) to generate probabilistic maps of soil properties (prior soil maps). As illustrated in panel (b), the preparation for residual correction involves adding additional soil profiles, spatially and vertically aligning prior soil map values with new profile observations, calculating residuals depth by depth, and preparing environmental covariates and soil covariates (new feature

space) for residual correction. Finally, as shown in panel (c), the iterative residual correction step applies bias corrections across different depths, focusing on layers where residuals have not yet stabilized. During each iteration, the model predicts residuals for one depth at a time, randomly selecting a layer. Once residuals for a given depth converge, that layer is excluded from further updates, allowing the model to concentrate on remaining depths until all achieve stability. After verifying convergence across all depths, the algorithm updates the posterior distribution of soil properties and produces the final soil maps (posterior soil property maps).

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.

2.1.1 Worked Example

The iterative residual correction method is further illustrated in Figure 2 using an example with a randomly selected soil column. Figure 2a shows the location of the selected soil column, where additional soil profile observations are available. The right panel displays the top-3 probable pH values (from prior soil maps) at each depth intervals (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, 100–200 cm), while the left panel shows the three weights (probabilities) associated with these pH values. In this simplified example, we use three bins to represent the soil property distribution; however, in actual implementation, more bins are maintained (typically top-12 probable values) to better capture soil variability. For this demonstration, Depth 2 (D_2; 5–15 cm) is randomly selected as the modeling layer to initiate the iterative correction process. Only one layer is modeled and updated for a given iteration. Note that in real model execution, model generally processes over 3,000 soil columns simultaneously in California. Only one column is shown here for clarity.

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 modeling 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.

Let $\{v_i\}_{i=1}^{k}$ denote a set of candidate soil property values with associated normalized weights $\{w_i\}_{i=1}^{k}$, where $\sum_{i=1}^{k} w_i = 1$. Then a representative value $\hat{v}$ is computed as:

$$\hat{v} = \sum_{i=1}^{k} w_i \, v_i$$

(4) Update 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. Top-probable values refer to current estimates of soil property values with top-k ranked highest weights, and k is 12 in this example.

At the current iteration, candidate values are updated as the sum of predicted residuals and the prior estimates of soil property values. Let $\{v_i\}_{i=1}^{k}$ denote a set of candidate values based on their associated weights.

297 Specifically, the k candidates with the highest weights w_i are retained as the
298 "top-probable" values. The weights are normalized such that $\sum_{i=1}^{k} w_i = 1$,
299 defining a discrete approximation to the updated distribution:

300 $$P(V = v_i) = w_i, \quad i = 1, \ldots, k$$

301 where V is the soil property at a given pixel and soil layer. The parameter k
302 controls the level of truncation of the candidate set.
303 (5) Inter-layer differences (5 dimensions): Differences in top-probable predicted soil
304 property values between the modeling layer and the other five depth layers. For
305 instance, if modeling Depth 2, the inter-layer differences would be (D_2−D_1), (D_2−D_3),
306 (D_2−D_4), (D_2−D_5), and (D_2−D_6). These features capture vertical correlations in the soil
307 profile and aid in estimating spatial patterns.
308 (6) Weights (1 dimension): Probabilities associated with each top-probable soil
309 property value. These weights remain fixed throughout iterations.
310
311 In summary, environmental covariates and weights remain static, while depth information,
312 representative values, top-probable values, and inter-layer differences are updated across
313 iterations based on the most recent soil property estimates.
314
315 A Random Forest regressor is then trained using the feature space to predict residuals for
316 the modeling layer (D_2 in this example). The right panel in Figure 2c compares the
317 distribution of pH values before and after residual adjustment in the current iteration. After
318 applying the residual correction, convergence is checked for D_2 by comparing the median

difference between the current and previous residuals. If D_2 has converged (difference below threshold), the algorithm proceeds to check whether all depth layers have converged. If all layers have converged, the iterative process terminates, and the final posterior soil property maps are generated by adding the last predicted residuals to the prior values.

If either convergence check returns "No" (i.e., D_2 has not converged or other layers remain “unconverged”), the algorithm continues iterating. Here, the soil property values for D_2 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.

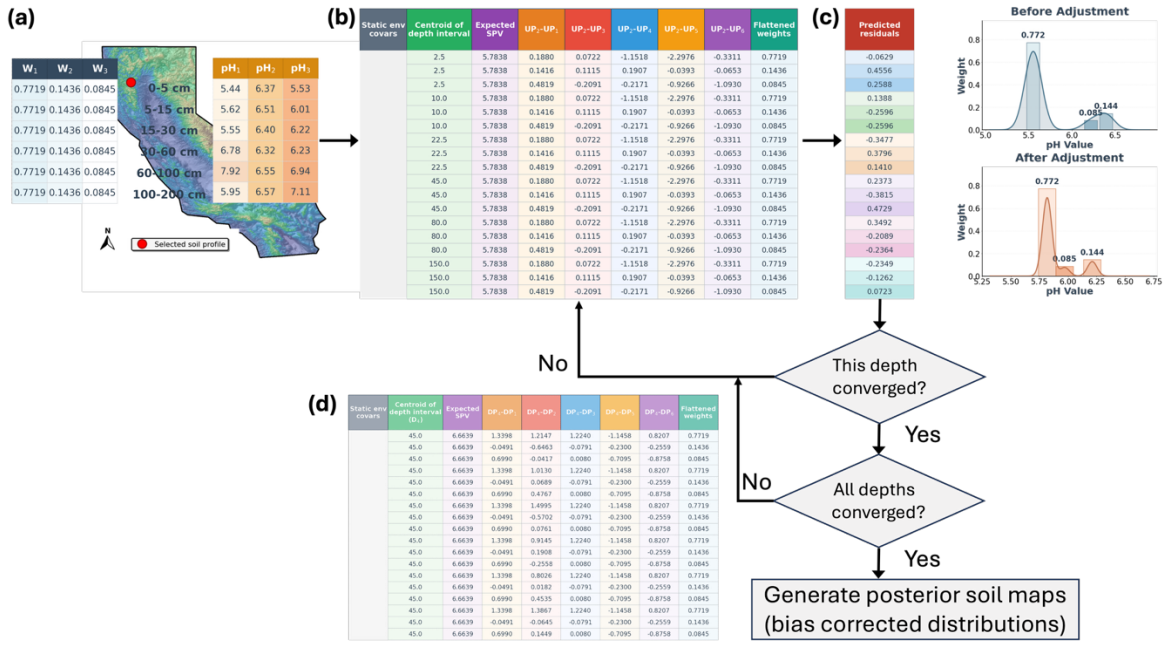


Figure 2: Schematic illustration of the iterative residual correction (IRC) method using a worked example at a randomly selected soil column. (a) Prior distributions and observation location: The map shows the location of the selected soil column within the study area. The right panel displays the top-3 probable pH values at each of the six depth intervals (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, 100–200 cm), while the left panel shows the three weights (w_1, w_2, w_3) associated with these pH values. Depth 2 (D_2; 5–15 cm) is randomly selected for this iteration. (b) Feature space components: The table details the structure of the feature space used to train the Random Forest regressor for residual prediction. The feature space comprises both static and dynamic components. Static components include environmental covariates (satellite imagery, terrain attributes) that remain unchanged throughout iterations, and weights (w_1, w_2, w_3) associated with top-probable values. Dynamic soil covariates that are updated in each iteration include: the centroid of the depth interval (e.g., 10 cm for D_2), the expected (representative) soil

property value computed as the weighted mean, the top-probable soil property values reflecting intra-pixel heterogeneity, and inter-layer differences capturing vertical correlations (e.g., D_2-D_1, D_2-D_3). (c) Residual correction and convergence workflow: A Random Forest model trained on the feature space predicts residuals for the modeling layer D_2. The right panel compares the pH distribution before and after residual adjustment. The flowchart below describes the convergence logic: after predicting and applying residuals to D_2, the algorithm evaluates whether D_2 has converged. If D_2 has converged, the algorithm checks whether all depth layers have achieved convergence. If both checks pass, the final posterior soil property maps are generated by adding the last converged residuals to the prior values. (d) If either check fails, the algorithm updates the soil property values for D_2 by adding predicted residuals, reconstructs the feature space with the updated values, randomly selects another "unconverged" layer, and repeats the process. This iterative cycle continues until convergence is achieved across all six depth layers.

2.1.2 Convergence of Residual Correction

The residual correction process iterates until the residuals stabilize, indicating that further adjustments yield diminishing gains for prediction accuracy. In the algorithm, convergence is achieved by a stopping criterion, which is customizable parameter. It can be a fix constant or different constants for different soil properties. Once this stability is reached, the final converged residuals are added to the prior prediction to generate the posterior soil

properties. To avoid over-correcting bias, only the last converged residuals are added to the prior prediction to generate the final posterior results.

2.1.3 Update with Constraints

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%; fine-earth fraction in mass). 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) in the end of each iteration. For example, if adding a residual of +15% to a prior sand content of 90% yields 105%, this value is capped at 100% (mass fraction). 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

393 that the total correction remains consistent with the model's prediction while maintaining
394 physical plausibility. For particle size fractions (sand, silt, clay), an additional
395 compositional constraint ensures that the three fractions sum to 100% at each pixel after
396 residual correction. While the residual corrections for particle size fractions are modeled
397 independently, any deviation from the unit sum is corrected by redistributing the error
398 proportionally after all iterations are completed. However, it is important to acknowledge
399 the limitations of this approach: since the fractions are not modeled as a joint
400 compositional vector (e.g., within a log-ratio geometry), this independent correction
401 followed by proportional scaling may not fully capture the inherent inter-dependencies and
402 non-linear constraints between soil textures.
403
404 **2.2 California Case Study: Soil Data and Model Implementation**
405 **2.2.1 Soil Data**
406 To demonstrate the IRC method, we apply it to soil property mapping in California. We use
407 georeferenced soil profiles with laboratory measurements of soil properties. We compiled
408 soil profile data from three primary sources: the World Soil Information Service (WoSIS),
409 the National Soil Characterization Database (SCD), and field measurements conducted in
410 California (Batjes et al., 2024; National Cooperative Soil Survey, 2018; Scudiero et al.,
411 2024).
412
413 To ensure consistency across different data sources, we applied several quality control
414 steps. First, we checked the physical plausibility of all soil property values by defining a

415 valid range with specific minimum and maximum thresholds for each property. Any data
416 point falling outside these ranges was considered an error and removed. For soil texture,
417 we required the sum of sand, silt, and clay fractions to equal 100% (mass fraction). If a
418 profile did not meet this compositional constraint, it was excluded. After quality check, the
419 datasets are compatible because the WoSIS records for California are largely derived from
420 the NCSS database, and both the SCD and WoSIS datasets follow standardized laboratory
421 protocols, such as those from the Kellogg Soil Survey Laboratory (Soil, 1996; Soil Survey
422 Staff, 2014). For our own field measurements, we used the Integral Suspension Pressure
423 (ISP+) method to maintain precision for particle size analysis (Corwin and Scudiero, 2020;
424 Scudiero et al., 2024).
425
426 During preprocessing, we harmonized all soil data, which was originally reported at
427 different soil horizons, into six standard depth intervals: 0–5 cm, 5–15 cm, 15–30 cm, 30–
428 60 cm, 60–100 cm, and 100–200 cm (Arrouays et al., 2014). The harmonization was
429 performed using equal-area spline functions to interpolate soil property values from the
430 original horizon depths to these standard intervals (Hartemink et al., 2010, p.201). The
431 spline function fits a smooth curve through observed values at their measured depths,
432 then calculates the area under this curve within each standardized depth interval and
433 divides by the interval width to obtain the value. Location of soil profiles and their
434 distribution of soil property values are presented in Figure 3. Six soil properties are studied:
435 sand content (% mass), silt content (% mass), clay content (% mass), pH, soil organic
436 matter (log-scaled % mass), and oven-dry bulk density ($g\ cm^{-3}$). 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.

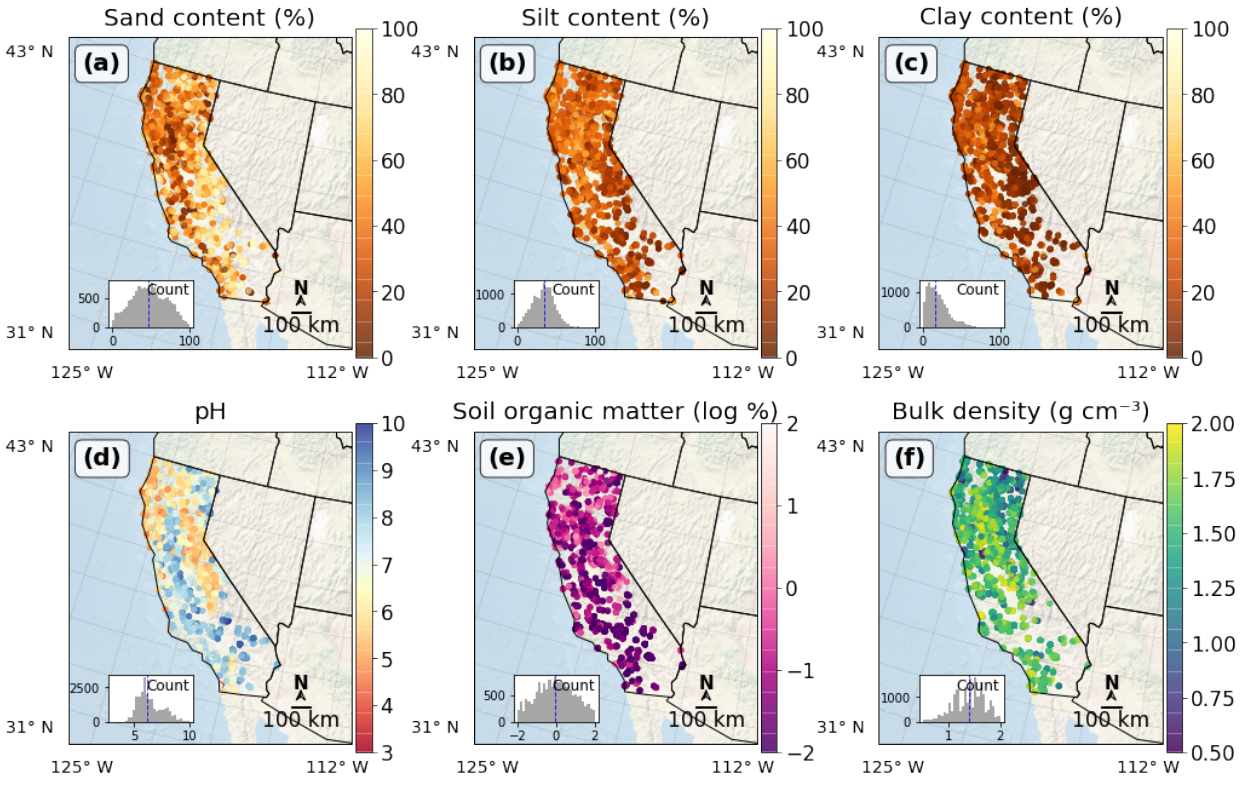


Figure 3: Spatial distribution and statistical characteristics of soil properties observations across California. The figure presents six soil parameters mapped using an Albers Equal Area projection: (a) sand content (% mass), (b) silt content (% mass), (c) clay content (%

mass), (d) pH, (e) soil organic matter (log-scaled % mass), and (f) bulk density (g cm^{-3}). Each subplot displays sample locations as colored points, with field-collected samples shown as triangles to distinguish them from WoSIS (circles) and SCD (squares) samples. Distribution histograms in the lower left corner of each subplot show the frequency distribution of values, with blue dashed lines indicating median values. Distance scale bar and compass rose are provided in the right corner. 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.

2.2.1.1 World Soil Information Service (WoSIS)

The World Soil Information Service (WoSIS), managed by the International Soil Reference and Information Centre (ISRIC), aggregates global soil data from diverse sources, including national soil institutes, research organizations, and collaborative initiatives like the Global Soil Partnership (GSP) and the International Network of Soil Information Institutions (INSII). The database provides soil properties for different soil horizons, georeferenced in decimal degrees, and undergoes quality controls (Batjes et al., 2024). In California, WoSIS typically offers 2,000 to over 5,000 soil observations for the modeling soil property. Samples below 1 m depth are fewer than those from shallower layers.

2.2.1.2 Soil Characterization Database (SCD)

The Soil Characterization Database (SCD) is a subset of the National Cooperative Soil Survey (NCSS) database (National Cooperative Soil Survey, 2018). It records soil properties

for each soil horizon within a soil profile (pedon), including soil texture, bulk density, and water retention. In California, SCD provides between 500 and over 1,000 soil samples per layer for the studied soil property. Each soil profile is georeferenced and includes metadata such as site location, land use, and sampling methods.

2.2.1.3 Ground Truth Soil Sampling and Measurements

Additional soil sampling was conducted to complement georeferenced soil profiles in California for model training and evaluation. These data are reported in (Scudiero et al., 2024) and are briefly discussed here. Multiple fields located between Salinas and Soledad in California's Salinas Valley were selected to collect soil particle size fraction data (Figure 4). These fields, presented as red dots in Figure 4, were chosen because they were accessible, unfarmed during the sampling period, and spread across different parts of the valley.

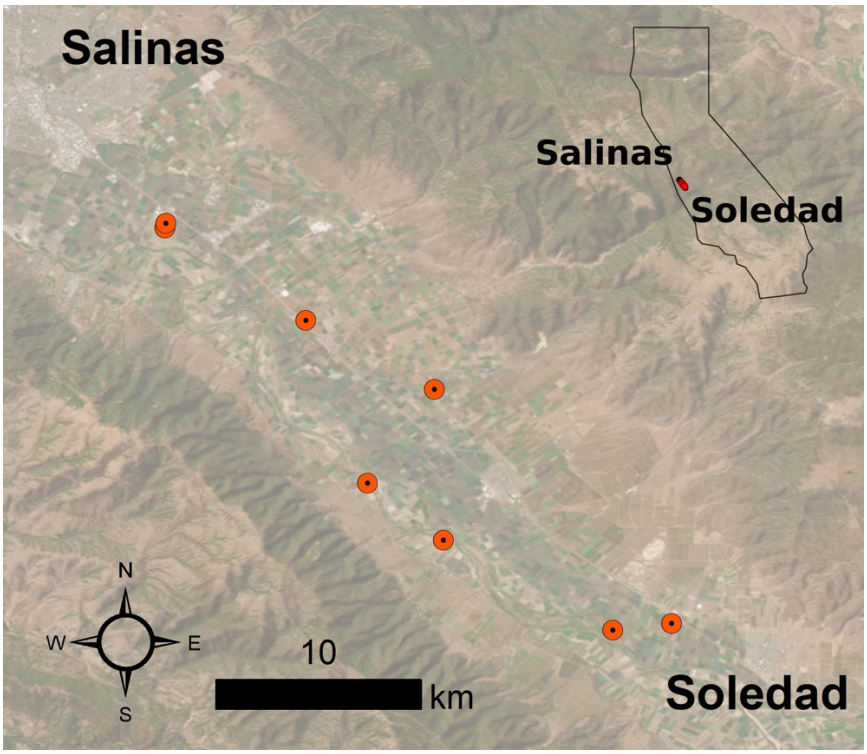


Figure 4: Map of sampling fields in the Salinas Valley in California. Each red dot represents a sampling field between Salinas and Soledad. An inset map (top right) shows the location of the sampling area within California. Scale bar and direction indicator are provided in the left corner. *Basemap: Esri World Imagery. Source: Esri, Maxar, Earthstar Geographics, and the GIS User Community.*

Soil apparent electrical conductivity (ECa) was measured across fields using an electromagnetic induction (EMI) sensor connected to a GPS receiver. Following the ECa-directed soil sampling protocols of Corwin and Scudiero (Corwin and Scudiero, 2020), the most representative soil samples were identified with ESAP software package and the Response Surface Sampling Design algorithm (Lesch et al., 2000; Lesch, 2005). 0-0.8 and

0-1.6 m soil profiles were further analyzed and followed with the expectation that ECa was a regional proxy for the field-scale variability of particle size fraction.

To measure particle size fraction, soil samples were then collected from multiple depths (0–0.1, 0.1–0.4, and 0.4–1.2 m) across fields. After collection, the samples were air-dried, ground, and sieved to remove particles larger than 2 mm; and then measured using the Integral Suspension Pressure method (The improved integral suspension pressure method (ISP+) for precise particle size analysis of soil and sedimentary materials; Wolfgang Durner, Sascha C. Iden) using PARIO™ system (METER Group AG, Munich, Germany).

2.2.2 Model Implementation for the California Case Study

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). After gaining prior estimate of soil properties, the IRC method was then applied using the additional soil observations from WoSIS, SCD, and field measurements, which were not used in generating the prior maps. The convergence threshold for each soil property is set to that the median change in residuals falls below the 5th percentile of the residual distribution observed across the latest three consecutive iterations.

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 all iterations. It should be noted that while removing co-located samples prevents direct data leakage, this OOB evaluation remains subject to the effects of spatial autocorrelation. Because OOB samples may still be located in close proximity to training clusters, the resulting error metrics may reflect the model's data assimilation performance and yield optimistic estimates of model performance.

3 Results

The iterative residual correction (IRC) method is applied to adjust pHRF-derived prior soil properties, including particle size fractions (mass percentage of sand content, silt content, and clay content), pH, oven-dry bulk density (BD; $\mathrm{g\ cm^{-3}}$), and soil organic matter (SOM; log-scaled mass percentage) over California. This correction addresses biases in the prior soil property maps and updates the posterior distributions of these properties. These soil properties are important for land management and serve as essential inputs for pedotransfer functions. The residual correction is performed across California, covering six depth intervals: 0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm, and 100-200 cm.

3.1 Performance Evaluation of Posterior Soil Properties

Table 1 presents the performance metrics for the posterior predictions of six key soil properties: mass percentage of sand content, silt content, and clay content (% mass), pH, oven-dry bulk density (BD; $\mathrm{g\ cm^{-3}}$), and soil organic matter (SOM; mass percentage). The metrics include the root mean square error (RMSE), coefficient of determination (R^2), and correlation coefficient (ρ). For example, sand prediction (% mass) shows an RMSE of 9.322, an R^2 of 0.841, and a correlation coefficient of 0.918. pH prediction shows an RMSE of 0.270, an R^2 of 0.945, and a correlation coefficient of 0.972. These metrics are computed using out-of-bag (OOB) samples from random forest regressors. OOB samples are data points not included in the bootstrap samples used to train each tree in the random forest. Additionally, these metrics are evaluated by comparing the expected values of posterior predictions with co-located soil properties values; not computed on residuals.

Table 1 also shows variations in performance across different soil properties. SOM and bulk density show slightly worse metrics compared to particle size fractions and pH. For instance, SOM predictions (mass percentage) have an RMSE of 1.961, an R^2 of 0.608, and a correlation coefficient of 0.801, and bulk density predictions ($\mathrm{g\ cm^{-3}}$) have an RMSE of 0.164, an R^2 of 0.704, and a correlation coefficient of 0.843. Two main reasons can result in their lower performance. First, these properties are more dynamic in nature compared to particle size fractions and pH. SOM and bulk density can change over time due to factors such as land use practices. The prior predictions are trained using soil survey data that are

older, while the posterior soil profiles used for evaluation may come from a different period. Second, SOM and bulk density are more challenging to model accurately. SOM is influenced by complex biological and soil-forming processes, such as decomposition rates and organic matter inputs. Similarly, bulk density is affected by soil compaction, organic matter content, and soil structure. All of them can vary spatially and temporally. Depth-wise analysis of model performance is provided in the Supplementary Information (Table S1 and S2).

Table 1: Performance metrics (RMSE, R^2, and correlation coefficient ρ) for posterior predictions of soil properties, including sand, silt, clay, pH, oven-dry bulk density (BD), and soil organic matter (SOM). The table summarizes the range (minimum and maximum values) and accuracy metrics for each property averaged across all depth intervals.

Property	Unit	Min	Max	RMSE	R^2	φ
Sand	% mass	0.0	100.0	9.322	0.841	0.918
Silt	% mass	0.0	100.0	6.556	0.788	0.889
Clay	% mass	0.0	100.0	5.891	0.841	0.918
pH	log10([H^+])	3.0	10.0	0.270	0.945	0.972
BD (oven-dry)	$g\ cm^{-3}$	0.5	2.0	0.164	0.704	0.843
SOM	% mass	0.0	100.0	1.961	0.608	0.801

The posterior predictions of soil properties all align with the co-located observations and can capture the general trend of observations (Figure 5). Predictions of pH show the most concentrated clustering to the dashed line, indicating good agreement with observations across all depths. SOM and bulk density show relatively weaker performance compared to other predicted soil properties. And this pattern of reduced accuracy persists throughout all depths.

As Figure 5 shows, the performance of the model tends to decline with increasing soil depth, except for SOM. This decline is primarily due to several reasons. First, the availability of soil data is often greater for shallower layers compared to deeper layers (such as > 1m), which limits the model's ability to learn patterns in deep layers. Second, remote sensing-derived soil covariates can only observe surface properties. Predictions for deeper layers rely on soil horizon information, soil profiles, geology, and parent material-related features. The certainty and quantity of them are less than easily measurable surface covariates. However, SOM shows better performance in deeper layers compared to surface layers. This is likely because surface SOM is highly variable due to factors like residue, land use, and management practices, while deeper SOM tends to be more stable.

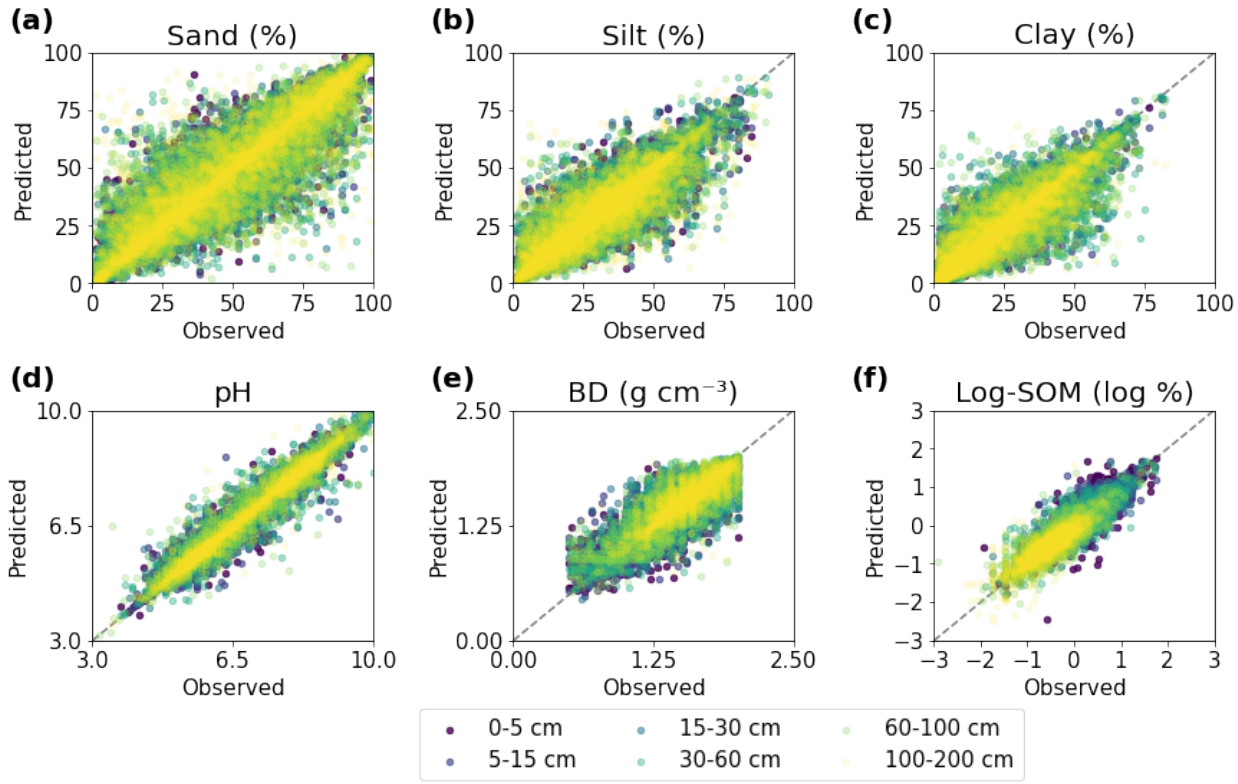


Figure 5: Evaluating posterior predictions with observations for six soil properties: (a) sand (% mass), (b) silt (% mass), (c) clay (% mass), (d) pH, (e) bulk density (BD; $\mathbf{g\ cm^{-3}}$), and (f) log-scaled soil organic matter (SOM; log % mass). The left side shows scatter plots of posterior predictions versus observations across six depth intervals, with each depth represented by a distinct color. The dashed black line represents perfect prediction.

3.2 Comparison of Prior and Posterior Soil Predictions

Prior and posterior predictions of soil properties are compared against co-located observations to assess the added value of residual correction. The radar plots in Figure 6 illustrate the improvements achieved through the residual correction method using three normalized unitless metrics: 1-normalized absolute bias (1-|Bias|), coefficient of

607 determination (R^2), and 1-normalized RMSE by ranges of soil variability (1-nRMSE). These
608 metrics are computed with values of soil properties, instead of on their residuals. Values in
609 Figure 6 closer to the outer edge of each plot indicate better model performance. Overall,
610 all soil properties maintain reasonable normalized bias and nRMSE (with nRMSE values
611 consistently less than 0.2 for both prior and posterior predictions). However, the prior
612 predictions tend to underestimate the variability of soil properties. As a result, the
613 normalized metrics for prior and posterior predictions are similar, while the R^2 values show
614 some differences.
615
616 For all soil properties, posterior predictions consistently outperform prior predictions
617 across all metrics. For particle size fractions, R^2 values show the largest improvements:
618 sand increases from 0.35 to 0.84, silt from 0.19 to 0.79, and clay from 0.25 to 0.84. The
619 nRMSE metric also shows improvements. Sand decreases from 0.19 to 0.09, silt from 0.14
620 to 0.07, and clay from 0.16 to 0.07, showing reductions in prediction errors using the
621 residual correction.
622
623 Aggregating data from all depths, Figure 6 shows the degree of improvement across
624 different soil properties. Prior pH predictions already demonstrate reasonable accuracy,
625 with an R^2 of 0.54 and nRMSE of 0.11. After the residual correction, these metrics improve
626 to 0.94 for R^2 and 0.04 for nRMSE. Bulk density and SOM show the biggest gains. For bulk
627 density, the R^2 increasing from 0.16 to 0.70 and nRMSE reducing from 0.18 to 0.11. Prior

SOM are underfitted with a low R^2 value. With the residual correction, the posterior SOM show a positive R^2 of 0.61. The nRMSE for SOM also improves from 0.07 to 0.04.

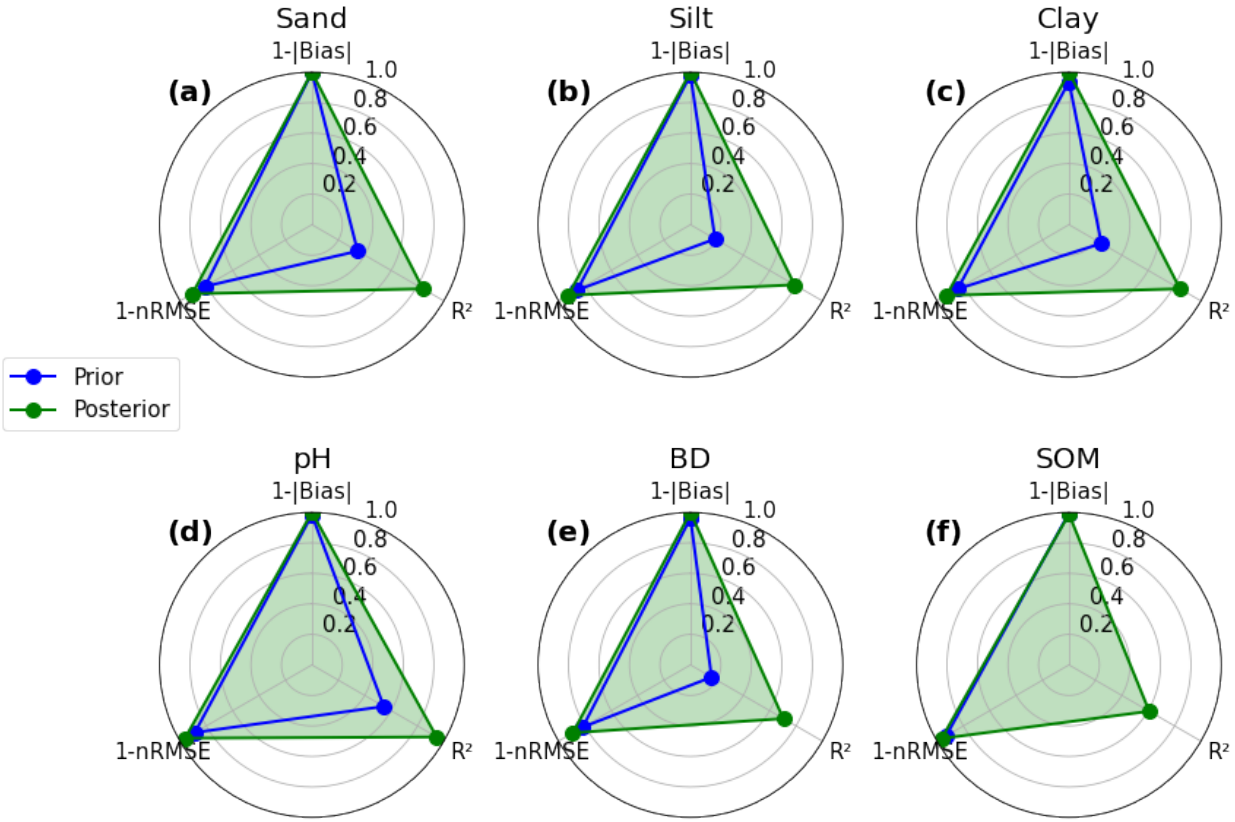


Figure 6: Radar plots comparing the performance metrics of prior and posterior predictions for six soil properties: (a) sand (% mass), (b) silt (% mass), (c) clay (% mass), (d) pH, (e) oven-dry bulk density (BD; g cm^{-3}), and (f) soil organic matter (SOM; log % mass). Each plot presents three metrics: 1-normalized absolute bias (1-|Bias|), coefficient of determination (R^2), and 1-normalized RMSE by ranges of soil variability (1-nRMSE). Prior predictions are shown in blue, and posterior predictions in green. All metrics are scaled from 0 to 1, where values closer to the outer edge of the plot indicate better model performance. The green shaded area highlights the improvement achieved by the posterior predictions over prior estimates.

To evaluate the statistical significance of the IRC method's performance, we compared prediction errors between the prior and posterior estimates using (1000) bootstrap-derived RMSE and 95% confidence intervals (CI) across all soil properties and depths (Figure 7). The posterior estimates consistently reduced RMSE relative to the prior across the entire profile. The non-overlapping confidence intervals in nearly all cases indicate a statistically significant reduction in error, except for SOM from 100 to 200 cm. SOM exhibited wider confidence intervals than other properties, reflecting its skewed distribution, the presence of extreme outliers, and the inherent uncertainty in modeling for such a highly variable property.

In Figure 7, the observed improvements generally diminish with depth, likely due to lower data density in deeper soil layers and the reliance on remotely sensed surface features. Notably, the prior's overall weaker performance stems not only from model architecture but also from the limitations of the harmonized soil property database used for its construction (Chaney et al., 2019). This database smooths out intra-map-unit soil variability and within soil components' variability (Xu et al., 2025). In contrast, the IRC method integrates a large pool of georeferenced soil profiles, allowing for a more detailed recovery of point-based soil variation. While the OOB evaluation can be subject to spatial autocorrelation, we interpret these gains not as evidence of broad spatial extrapolation into unsampled regions, but more as the framework's improved capacity for data assimilation.

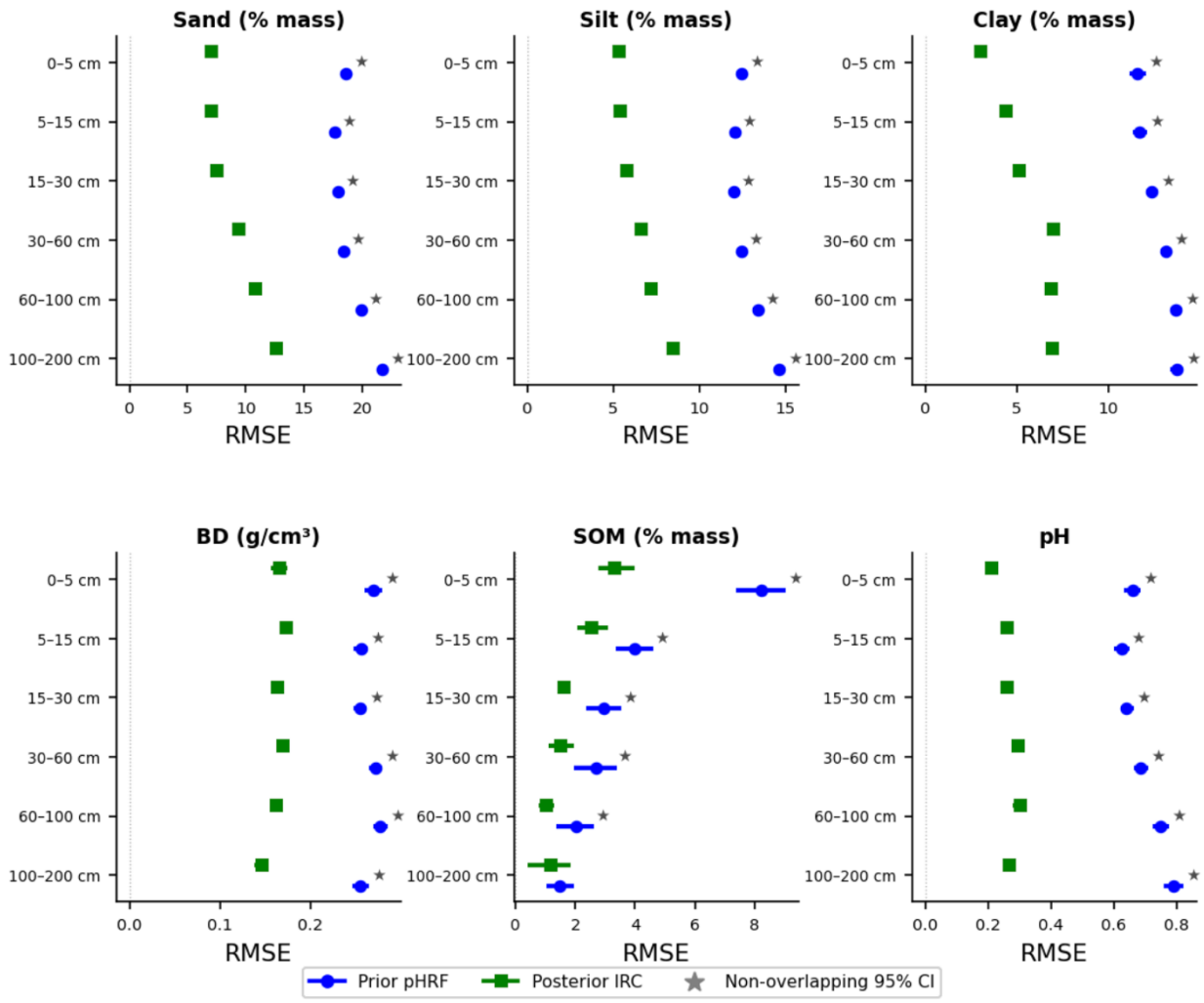


Figure 7: Forest plots of RMSE with bootstrap-derived 95% confidence intervals (CIs) for prior and posterior predictions across soil properties and depth intervals. Each panel corresponds to a soil property (sand, silt, clay, bulk density (BD), soil organic matter (SOM), and pH), and each row represents a depth interval. Blue circles denote mean prior estimates, while green squares denote mean posterior estimates, with horizontal bars indicating 95% confidence intervals based on 1000 bootstrap resamples. Stars (★)

indicate depth–soil property combinations where the confidence intervals of prior and posterior RMSE do not overlap, suggesting statistically significant differences.

Horizontal spatial patterns of the six soil properties are presented in Figure 8. In the Central Valley California, soils are mostly medium textured with about 30% silt and lower sand content compared to surrounding areas. In the Mojave and Colorado Deserts, high sand contents (> 60% mass) with low clay contents are observed. SOM contents are also low in these areas. The histograms show how residual correction adjusts the distribution of soil properties.

For SOM and bulk density, the prior predictions often underestimate the observed variation. Figure 8 shows that the residual correction processes add noticeable spatial variations between prior and posterior soil maps. Prior bulk density values are often clustered around 1.5 $\mathrm{g\ cm^{-3}}$, whereas the posterior histogram presents a broader range, spanning from 1.25 $\mathrm{g\ cm^{-3}}$ to 1.6 $\mathrm{g\ cm^{-3}}$, capturing more heterogeneity of bulk density. Similarly, the residual correction adds soil heterogeneity to SOM. The posterior SOM can delineate water bodies, where SOM content is abruptly lower than the surrounding areas. Additionally, the posterior SOM maps present hill features in the desert areas.

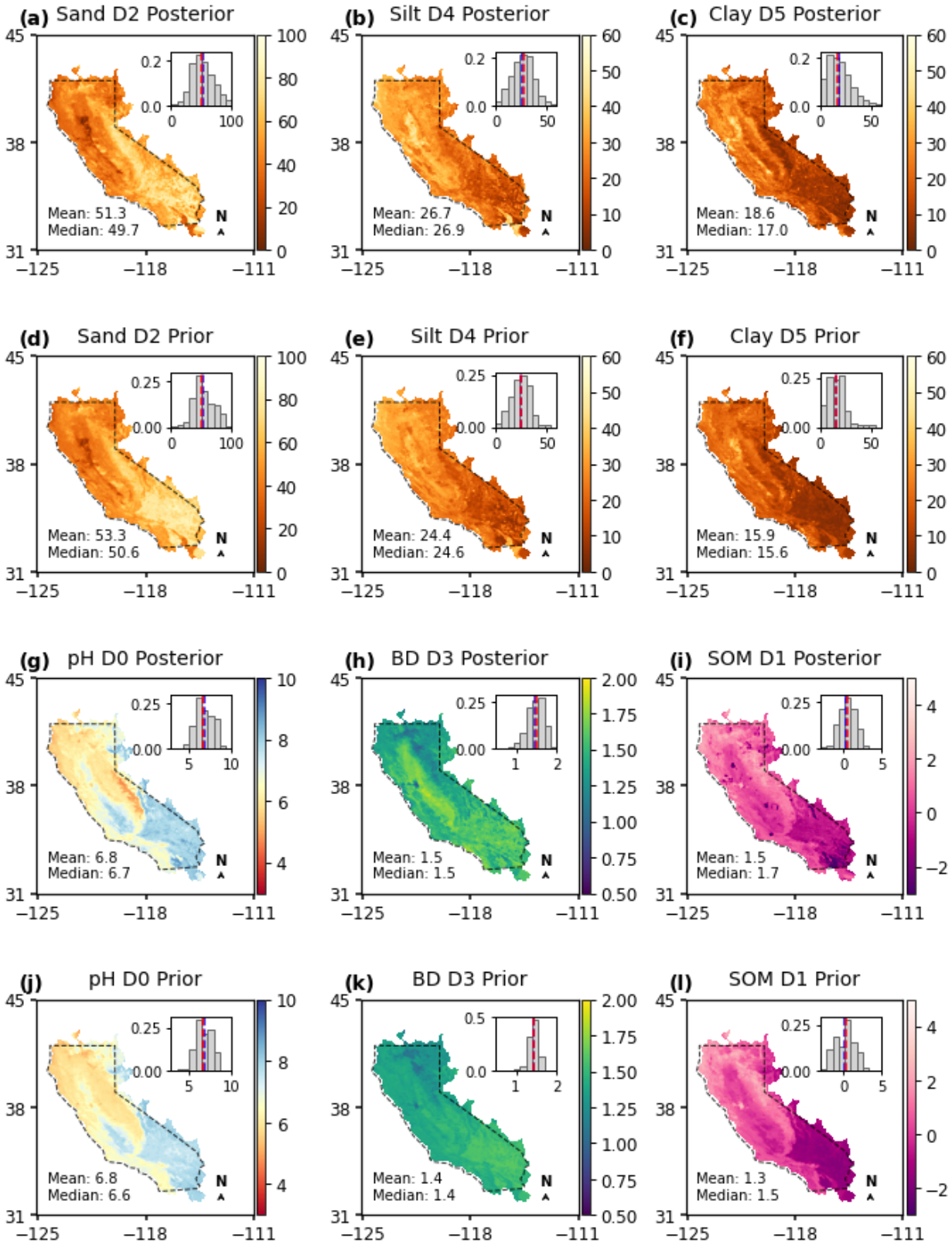


Figure 8: Spatial distribution of six soil properties: sand (% mass), silt (% mass), clay

content (% mass), pH, bulk density (g cm^{-3}), and soil organic matter (log % mass) across California. Maps of prior and posterior soil properties are compared. The corresponding frequency distributions of these soil properties are displayed in the right corner. Dashed polygons represent the continental part of California. In the histograms, the blue and red dashed lines represent the mean and median values, respectively. The maps labeled D0 to D5 correspond to the first vertical layer down to the deepest layer. Note the map and distribution of soil organic matter (SOM) is log-scaled. Mean and median values are computed from the original SOM data.

Figure 9 presents a comparative directional semi-variogram analysis for the six soil properties at the 5–15 cm depth over areas in the Central Valley, California. The plots contrast the observed spatial variance (black solid; OOB samples generated only) with the prior predictions (blue dashed) and the posterior results (green dash-dot), utilizing spherical models to quantify the nugget (C_0), sill (C), and range (a). High nugget values across the properties reflect soil variability, sampling variance, and measurement error that remain irreducible at the current model resolution (1km). Directional semi-variograms compute variance along a specific geographic azimuth (shown by the inset map in each panel), allowing detection of anisotropy. Across all panels, the results indicate that the priors generally fail to capture the extent of spatial variance and correlation length of soil properties. The implementation of the IRC shifts the posterior semi-variograms towards the observed trend, increasing the captured sill and adjusting the effective spatial range to better align with the observed spatial dissimilarity of the domain.

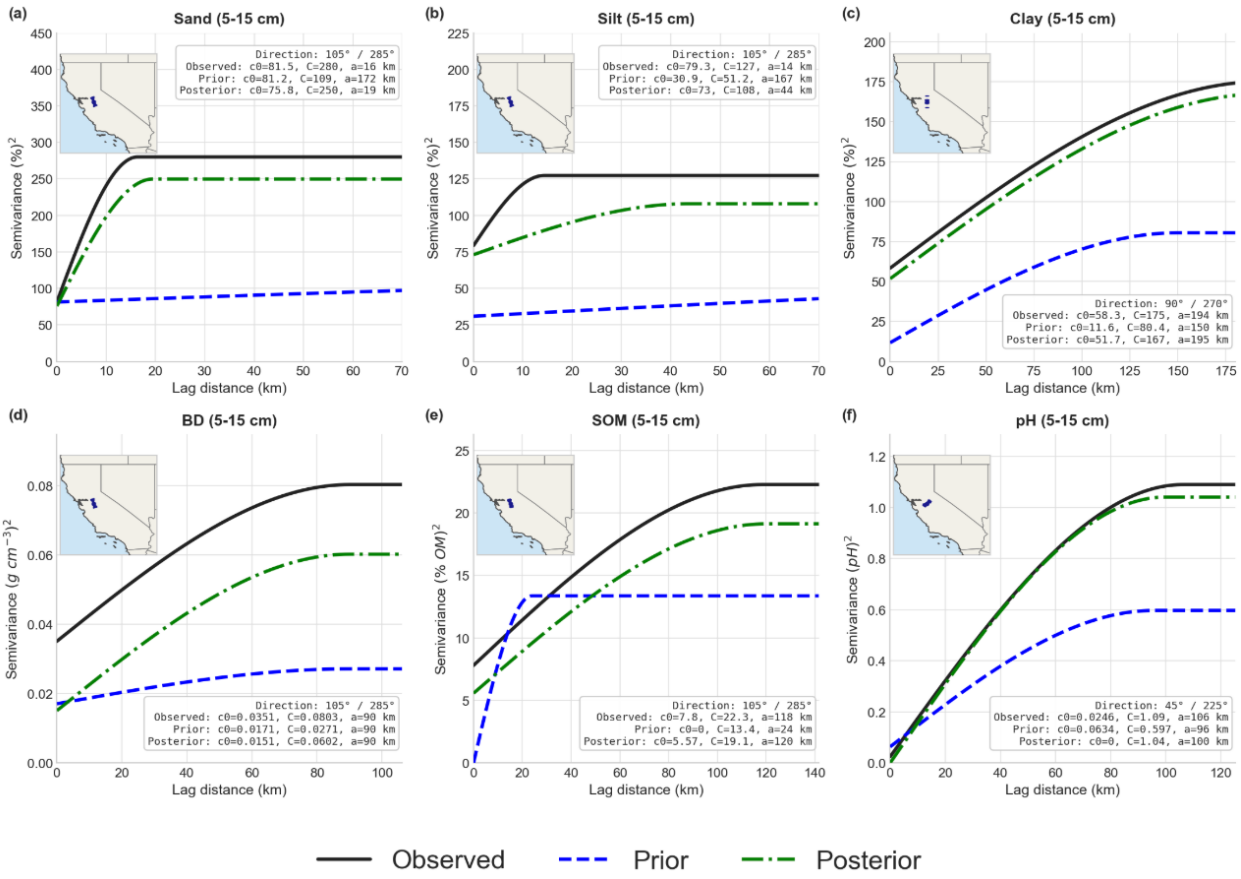


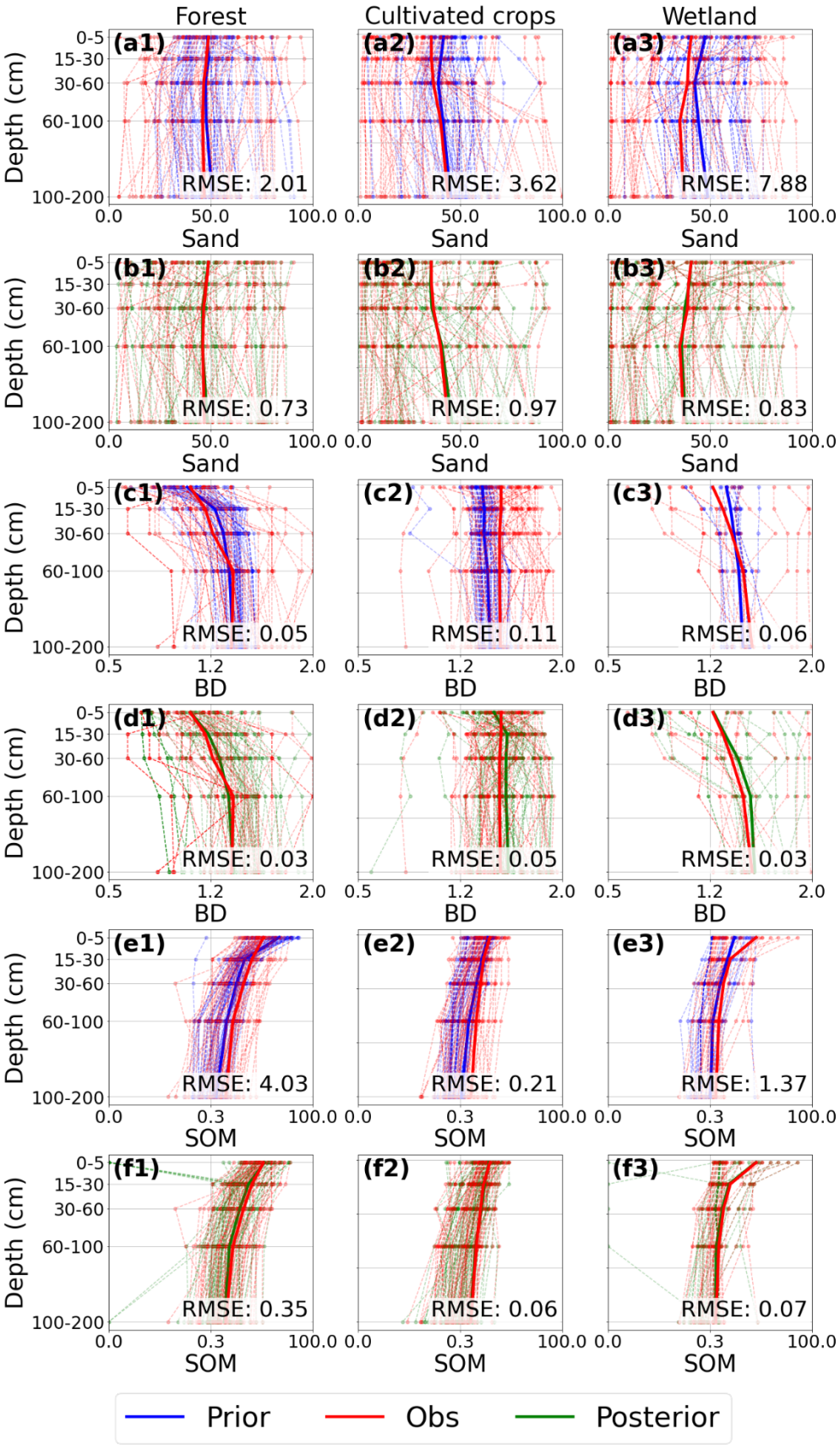
Figure 9: Directional semi-variograms for six soil properties at the 5–15 cm depth, computed within the Central Valley, California subregion and fitted with spherical models. Each panel shows curves for the observed samples (black solid), the prior spatial prediction (blue dashed), and the posterior prediction (green dash-dot). The fitted spherical parameters, nugget (c_0), sill (C), and range (a), are reported in the box. The inset maps indicate the sampling location and the direction along which spatial dissimilarity is measured. For example, a 105°/285° axis corresponds to a NNW–SSE geographic azimuth.

Soil profiles used for evaluating residual correction are grouped according to their corresponding pixel's land use classification from the National Land Cover Database (NLCD). Figure 10 presents selected vertical soil profiles of sand content, oven-dry bulk density, and SOM across three land use categories: forest, cultivated crops, and wetland. The number of samples varies by land use, with forests having the most, cultivated crops approximately half as many, and wetlands the fewest across California. To ensure a balanced visualization, a similar number of profiles are selected from each category. Sand content is chosen due to its broader range of variation (0-100% mass) compared to silt and clay (< 60% range). SOM and bulk density, which show relatively lower performance metrics, are included to assess the model's 'lower-bound performance'. These vertical profiles were not used during model training.

In Figure 10, solid lines represent the mean soil profiles for sand content, oven-dry bulk density, and SOM across forest, cultivated crops, and wetland land use categories. Blue lines, red lines, and green lines indicate prior, observation, and posterior predictions. Comparing the solid lines, the posterior predictions align more closely with the observed data compared to the prior estimates. However, the degree of alignment varies by soil property. For sand content and SOM, the posterior predictions show better agreement with observations, while bulk density predictions exhibit greater discrepancies, particularly in cultivated areas.

For sand content, the residual correction process improves estimates, especially in wetlands, with RMSE decreasing from 7.68 to 0.77 (% mass). Bulk density predictions perform better in forested and wetland areas. In cultivated crops, the posterior predictions show larger discrepancies. This suggests that bulk density is more challenging to predict in agricultural lands, particularly in shallow layers, likely due to agricultural activities. For SOM, the residual correction effectively improves estimates, especially in the surface layers of wetlands.

Dashed lines in Figure 10 represent individual soil profiles. Prior predictions often underestimated the variability in soil properties, struggling to capture extreme values. After the residual correction, the posterior predictions are better able to approximate these extremes. However, the correction process sometimes introduces additional noise. For example, some low SOM values were generated during residual correction, even though such values are not presented in the observed data. It is likely due to that we used the van Bemmelen factor (1.724) to convert the prior soil organic matter to soil organic carbon.


Forest
Cultivated crops
Wetland
Depth (cm)
0-5
15-30
30-60
60-100
100-200
(a1)
RMSE: 2.01
(a2)
RMSE: 3.62
(a3)
RMSE: 7.88
Sand
(b1)
RMSE: 0.73
(b2)
RMSE: 0.97
(b3)
RMSE: 0.83
(c1)
RMSE: 0.05
(c2)
RMSE: 0.11
(c3)
RMSE: 0.06
BD
(d1)
RMSE: 0.03
(d2)
RMSE: 0.05
(d3)
RMSE: 0.03
(e1)
RMSE: 4.03
(e2)
RMSE: 0.21
(e3)
RMSE: 1.37
SOM
(f1)
RMSE: 0.35
(f2)
RMSE: 0.06
(f3)
RMSE: 0.07
Prior
Obs
Posterior

758

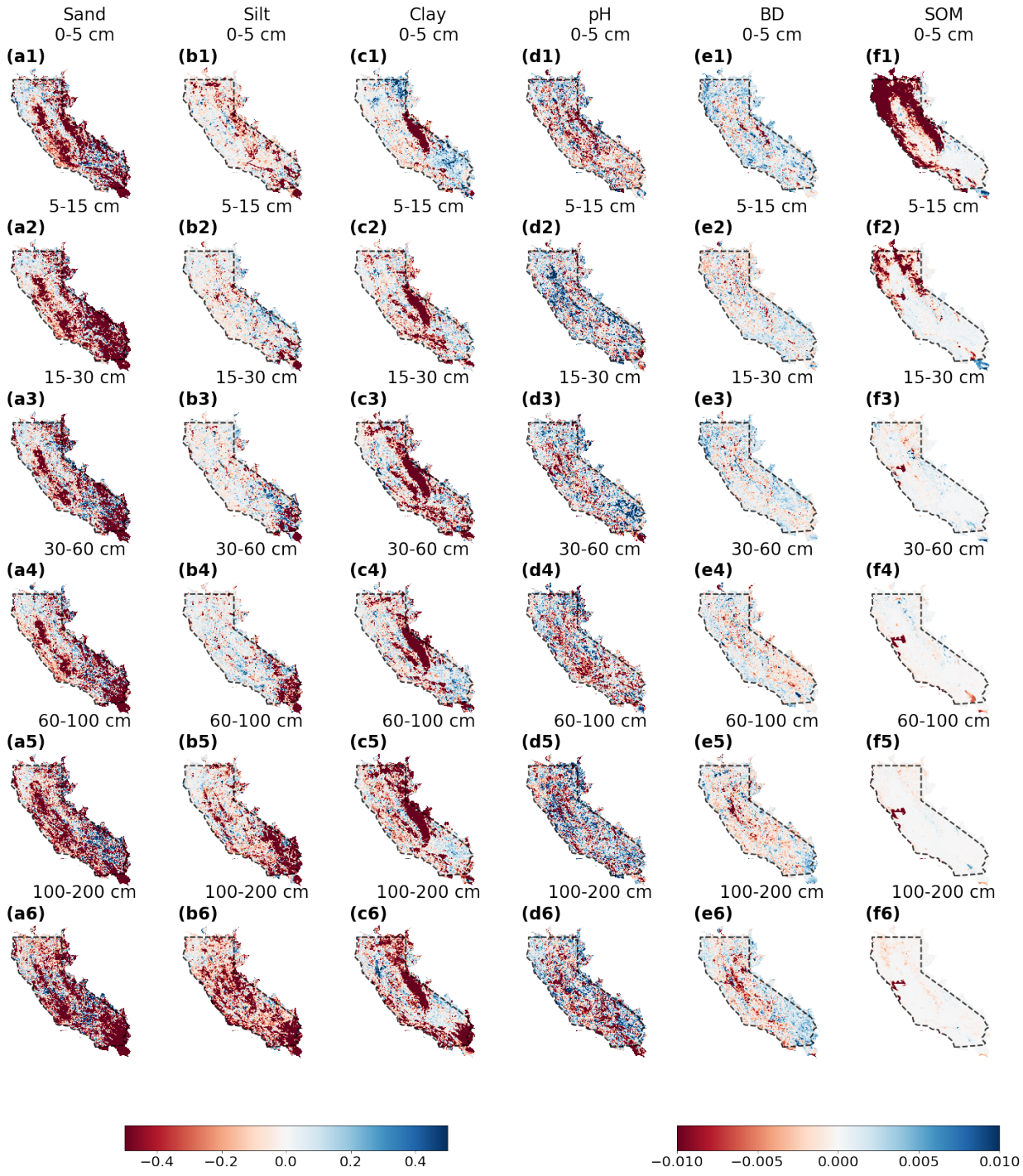
Figure 10: Vertical distribution of soil properties (sand content, oven-dry bulk density BD, and soil organic matter SOM) across three land use categories: forest, cultivated crops, and wetland. Prior estimates (blue), posterior estimates (green), and observations (red) are shown as depth profiles. Dashed lines represent individual measurements, and solid lines show mean values. RMSE is computed elementwise to evaluate model performance across all depths. X-axis and Y-axis represent value ranges of a soil property and vertical depth intervals, respectively.

3.3 Uncertainty Analysis

Figure 11 shows the differences between 5% — 95% posterior and prior prediction interval widths (PIWs) for six soil properties, sand, silt, clay, pH, bulk density, and SOM, from surface to 2-m deep. The differences are calculated by subtracting the prior PIWs from the posteriors. Red areas present a reduction in posterior PIW, indicating the residual correction has reduced uncertainties of soil properties predictions. Blue pixels suggest the opposite. White areas represent regions where the prior and posterior uncertainties are similar.

In Figure 11, most pixels show reduced uncertainty for sand content after residual correction, particularly in agricultural and desert regions. This improvement is attributed to the inclusion of additional soil profile data from these areas. For clay content, the posterior predictions consistently show reduced uncertainty across the Sierra Nevada Mountain ranges. For SOM, the posterior PIWs improved in shallower layers (0-15 cm) over both the

781 Coastal Ranges and the Sierra Nevada Mountains, with the coastal line showing notably
782 narrower PIWs. For pH, the results present a mixed pattern of PIWs after residual
783 correction, with some areas showing reduced uncertainty and others showing the
784 opposite. Similarly, bulk density exhibits a mixed pattern, though deeper layers (60 cm to 2
785 m) generally show reduced uncertainty in the Central Valley, California.



786

787 **Figure 11: Differences of 5% — 95% posterior and prior prediction interval widths**

788 **(PIWs) for soil properties across different depths. Each column represents a specific**

789 **soil property and rows show different depths. Black polygons represent the**

790 **continental part of California. Differences between posterior and prior PIWs are in a**

red-to-blue color scale. Red pixels indicate a decrease in posterior PIW, indicating residual correction reduces uncertainties. Vice versa for blue pixels. White areas indicate similar extent of uncertainties. The left colorbar corresponds to sand, silt, clay with wider ranges of PIW differences. The right colorbar represents other properties with smaller PIW changes.

Figure 12 evaluates the uncertainty quantification by visualizing the coverage of the 90% prediction intervals (PIs) across the range of soil properties. With the x-axis representing the sorted observation deciles and the y-axis showing predicted values, the fan charts present a consistent shift in the posterior (green) band toward the observed trend (black line) compared to the prior (blue) band. This indicates that the IRC method recalibrates the central tendency and shifts the uncertainty distribution to better align with the observed variance of soil properties. In deciles 1 and 10 across all panels, most notably for soil texture and bulk density, the prior bands are difficult to encompass the observed mean, indicating an underestimation of extreme-value variance. This suggests that the prior model and input data (especially the Harmonized soil properties database) suffer from "underestimate soil variability" problem, where extreme soil values are over-smoothed.

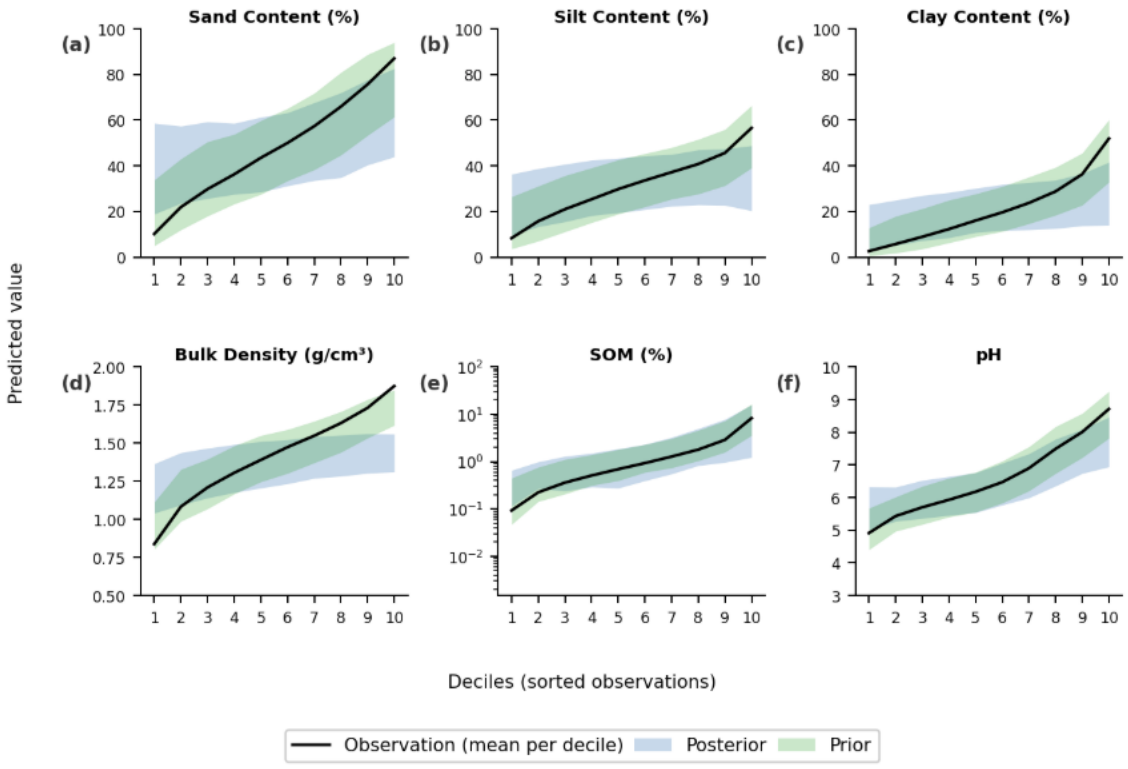


Figure 12: Prediction band fan charts comparing prior and posterior 90% prediction intervals (PI) across ten observation deciles. Observations (black line) are rank-sorted into ten deciles to evaluate model performance across the range of each soil property. The shaded bands represent the 90% prediction interval (5th to 95th percentiles) for the prior (blue) and the posterior (green).

3.4 Compare single-pass and iterative residual correction

Figure 13 evaluates the efficacy of the iterative approach in addressing the "smoothing to the mean" phenomenon for residual correction. SOM from 100 to 200 cm is chosen for several reasons: (1) As shown in Figure 7, this layer was the only one where the 95% confidence intervals of the RMSE for the prior and posterior results overlapped. We contend that this overlap may not imply a lack of model efficacy but can reflect the

inherent properties of deep SOM. SOM values at this depth are typically much smaller than those in surface layers, leading to a narrower absolute range of RMSE that makes “global” improvements appear marginal. (2) SOM distribution is skewed; most samples consist of low values, while only a small fraction constitutes the upper tail. Consequently, mean metrics are heavily weighted by these low-value samples, effectively "washing out" the specific impact of the residual correction on extreme values. To isolate the gain provided by the iterative process and to test if multiple iterations are necessary, we focus exclusively on the tail distribution of SOM from 100 to 200 cm (Figure 13).

Figure 13 examines model performance in the upper tails of the SOM distribution (100-200 cm deep) by comparing the observed values, the single-pass correction, and the final iterative posterior (IRC) results across extreme quantile subset (top 20%, 10%, 5%, and 1% of observations). Figure 13a shows the mean SOM values within each upper-tail subset. Both priors and posteriors underestimate SOM across all thresholds, with the underestimation becoming more pronounced at higher quantiles. Figure 13b quantifies the mean extent of upper-tail underprediction. The iterative posterior consistently reduces this bias relative to the single-pass model, particularly in the most extreme subsets. Figure 13c presents the relative improvement in mean absolute error (MAE) by percentage achieved by the iterative posterior. The improvement increases monotonically with extremity, reaching 25.6% for the top 1% subset. Overall, these results indicate that the iterative correction method improves the model’s ability to represent high-end SOM values compared to single-pass simulation. However, these results should be interpreted with

caution. The number of samples decreases substantially in the extreme tail (only 17 samples for the top 1%), which may limit the statistical robustness of the reported improvements. Future work needs to perform more robust and independent evaluation of tail behavior. Incorporating external datasets from independent measurements would help assess generalizability. In addition, extending the evaluation beyond California to broader geographic and environmental domains will increase the robustness of tail evaluation.

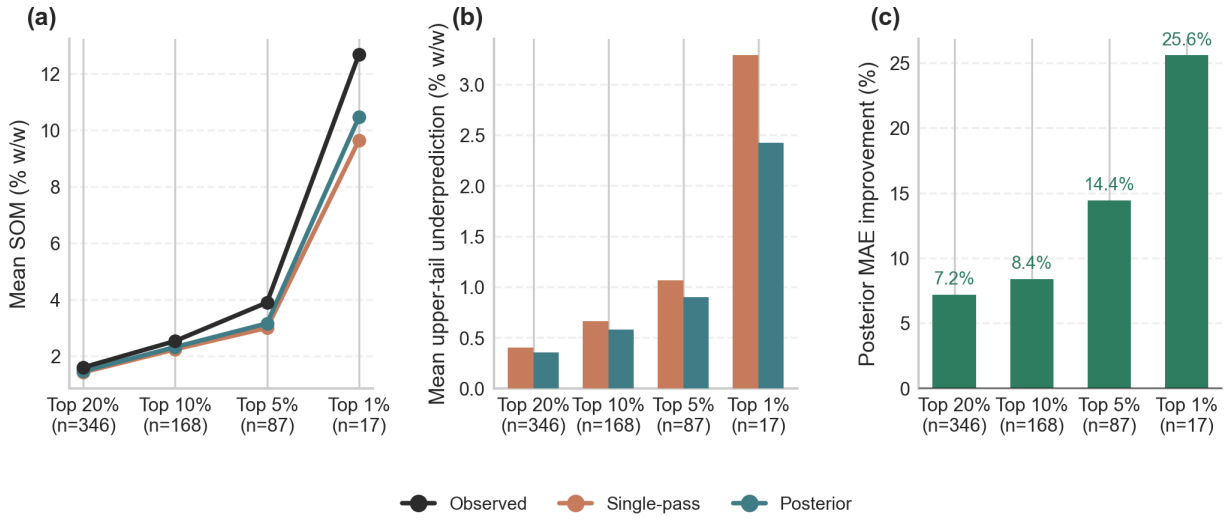


Figure 13: Performance of single-pass and iterative posterior models for the upper tail of SOM (100-200 cm) distribution. (a) Comparison of mean observed, single-pass, and posterior predicted SOM values across 4 groups of upper-tail percentiles. (b) Mean upper-tail underprediction (absolute bias) for single-pass (orange) and posterior (blue) models. (c) Percentage improvement in Mean Absolute Error (MAE) achieved by the posterior model relative to the single-pass model. All metrics are calculated for the 100–200 cm depth interval across the top 20%, 10%, 5%, and 1% of the SOM distribution.

4 Discussion

4.1 Performance of the IRC Method and Its Implications

The California case study achieved four research goals: (1) developed and implemented the IRC method for non-parametric updating of soil properties, (2) demonstrated the IRC method improve soil properties estimate compared to the pHRF-derived priors, (3) an iterative correction that showed added value of over a single-pass residual correction, (4) presented improved vertical and spatial structure by using the IRC method compared to the priors. Performance gains mainly stem from two aspects. First, the innovative model architecture: Unlike single-pass bias correction, the IRC method iteratively updates the feature space layer by layer, preserving vertical structure and enabling stepwise approximation. The convergence criterion prevents overfitting, while physical constraints keep posteriors realistic. The iterative posterior reduces MAE by 25.6% for extreme deep-SOM values relative to a single-pass model, with the margin growing in the tail where single-pass methods struggle. Second, the integration of additional georeferenced soil data: The prior combines georeferenced soil taxa and harmonized survey data (mainly SSURGO-derived); the posterior adds georeferenced soil profiles.

Improved soil mapping has broad implications for the scientific community and land management applications. The IRC posteriors shift sills towards observed values and adjust the effective correlation range, indicating that the method recovers observed spatial structure that was previously smoothed. This is directly relevant to applications that depend on local-scale soil variability, including site-specific irrigation management (Jiang

et al., 2011; Ortuani et al., 2016) and catchment-scale hydraulic modelling (Vereecken et al., 2022). The posterior 90% prediction band shifts towards the observed rank-ordered values, particularly in the extreme deciles where priors were most overconfident. The IRC framework therefore improves uncertainty calibration in addition to reducing prediction errors, a combination that is essential for soil maps to be reliably used in risk-sensitive applications such as flood modelling and agricultural decision support (Chaney et al., 2015; Vereecken et al., 2022). The profile-level analysis also demonstrates that the IRC framework improves the vertical coherence of predictions, which has direct relevance to soil hydrological modelling where the entire depth profile determines water retention, drainage, and root-zone moisture dynamics (Vereecken et al., 2016; Xu et al., 2023).

4.2 Limitations in Soil Profile Data

The effectiveness of residual correction depends on the spatial and vertical distribution of soil profiles used to calculate residuals. In regions with sparse sampling, such as California's desert areas (Figure 1), the limited number of profiles leads to interpolating the entire area using limited observations. If soil heterogeneity is not captured by these limited samples, the residual correction would overlook it. For soil texture, most data collected by staff working on multiple projects under the National Institute of Food and Agriculture (NIFA) and the Sustainable Agricultural Systems (SAS) programs range from the surface to 1.1 meters deep (additional field measurements used in this work). We use spline interpolation to predict soil texture data beyond 1.1-m depths. It assumes vertical continuity in soil properties, which may not reflect abrupt changes in subsurface layers.

Uncertainty also arises from converting some soil organic carbon (SOC) data to soil organic matter (SOM). We used the van Bemmelen factor (1.724) to convert SOC to SOM profiles. This factor does not hold true in scenarios such as organic-rich soils. Adding data quality controls, such as filtering profiles based on metadata (such as soil type, land use), could filter out samples that are not suitable for this conversion. However, this conversion still has uncertainties, since even for mineral soils, this factor still has a certain extent of variation depending on the organic matter composition (lower for soils with more decomposed organic matter), soil types (forest soils or wetland soils with anaerobic decomposition), and environmental influences (such as microbial activity).

4.3 Computational Challenges

The iterative residual correction process on distributions requires computational resources, particularly when applied to large-extent or high-resolution datasets. This process involves adjusting multiple values for each pixel, as each pixel represents a distribution of soil properties. This process can be approached in two ways. The first method involves correcting the residual values for each pixel, adding these residuals to update the posterior values of soil properties, and then converting these updated values to generate a posterior distribution of soil properties. The second method first converts all pixel values into the same histogram bins and then corrects the shape of these histogram bins for each pixel. Thus, the number of values retained per pixel affects computational expense. Based on our experience, using method two, especially for soil texture, requires

100-bin histograms. Using method one with 20 most probable prior property values for residual correction can achieve comparable results while reducing memory usage.

The iterative process of updating features and correcting residuals also plays a role. In our simulations, we observed that subsequent residual corrections generally align with previous ones. To ensure consistency, we require the corrections to converge more than three times across different depths. For example, residual correction for a 1-km soil property map over California takes approximately two hours after preprocessing the input data. However, processing higher-resolution datasets, such as those at a 10-meter scale, can demand significantly more computational resources. This highlights the trade-off between resolution and computational efficiency in DSM projects.

4.4 Temporal and Spatial Constraints

The current method does not account for temporal changes in soil properties, limiting its applicability to dynamic properties like soil organic matter or bulk density. Incorporating temporal covariates (such as seasonal land surface temperature, recent land-use changes) or stratifying soil profiles by collection date could address this. However, such improvements rely on the availability of temporally resolved soil data, which are often limited in quantities and sampling frequency.

Spatial clustering of soil samples poses another challenge. While duplicate profiles were removed during data preprocessing, nearby samples may still share a certain level of

similarity due to spatial autocorrelation. This could lead to overly optimistic evaluation of residual correction performance. Two methods can help address this issue:

(1) Cross-validation with spatial considerations: Implement a cross-validation method for splitting training and validation sets with attention to sample locations. Ensure a minimum distance between training samples and evaluation data.

(2) Independent dataset evaluation: Use independent datasets to evaluate the model. CONUS-wide instrumental network, such as the U.S. Climate Reference Network and the National Ecological Observatory Network, provide independent soil data. However, these datasets have limitations as they were collected with clustering to certain landscapes, potentially introducing bias in the evaluation.

4.5 Similar Studies

Several continental-scale DSM products (or methods) are compared, including the Soil Survey Geographic Database (SSURGO), the Gridded National Soil Survey Geographic Database (gNATSGO), the Probabilistic Layers for the Assessment of Soils (POLARIS), Soil-Landscape Unified Synthesis (SOLUS), and the pruned Hierarchical Random Forest with iterative bias correction (pHRF with IRC) soil properties. SSURGO is a traditional, polygon-based product derived from expert field surveys and remains widely used in agricultural applications (Soil Survey Staff et al., 2023). gNATSGO mainly builds on SSURGO by rasterizing its map units to improve spatial coverage. And its estimation of soil properties still rely on utilizing metadata of legacy soil data (Soil survey staff, 2023). These two still inherit legacy data's limitations, such as scale inconsistency between soil map units and

derived soil maps, inconsistencies with field observations, and report distribution of soil properties with only three values (low end value, representative value, and high end value) (Rossiter et al., 2022; Soil Survey Staff, 2025; Xu et al., 2025).

Development of the following DSM products incorporates quantitative models in their methodology. POLARIS produces probabilistic soil property maps using machine learning and the DSMART algorithm (Chaney et al., 2016, 2019; Odgers et al., 2015), while the uncertainties in the DSMART algorithm can propagate into POLARIS. SOLUS integrates legacy soil data with georeferenced field observations and employs linear adjusted Random Forest to predict soil properties (Nauman et al., 2024). SOLUS hierarchizes soil data with different qualities into its training dataset, giving more attention to georeferenced observations. However, since it also uses resampled soil data derived from polygon-based soil map units, this process may introduce additional uncertainties into the final product. The pHRF with IRC follows a different approach. Unlike most DSM methods that directly predict soil properties from input data, the pHRF with IRC follows a two-step approach: the pHRF first generates prior estimates of soil taxa and property values, which provide broad spatial coverage but can exhibit overconfident or biased predictions in certain regions; the IRC then iteratively corrects these priors by assimilating georeferenced soil profiles, reducing systematic underestimation and improving both predictive accuracy and uncertainty calibration. In future work, the pHRF with IRC method will be applied on large scale and assessed with more soil properties to evaluate its generalizability and robustness.

The comparison also shows distinctions in how different products represent uncertainty. SSURGO provides soil properties at the map unit level, whereas gNATSGO is a gridded product derived from SSURGO. However, their reported low, representative, and high values originate from component-level summaries rather than spatially explicit estimates and are thus unable to capture within-unit spatial variability or continuous gradients. POLARIS provides "full" distributions per pixel but derives them from synthetic sampling and a Harmonized (SSURGO) soil properties database, which can propagate upstream sampling randomness and smoothing artefacts. The IRC framework differs fundamentally in that uncertainty estimates are updated dynamically by assimilating new observations, and the prior and posterior leverage different sources of soil data that are available over the CONUS. Importantly, the innovation extends beyond model architecture: the prior is built from georeferenced soil taxa and a harmonized soil survey database (SSURGO), providing spatially continuous initial estimates rooted in legacy soil knowledge, while the posterior further leverages georeferenced soil profiles (NCSS) to refine those estimates where observation exist. This two-source design shows a practical advantage: in regions without georeferenced soil profiles, the prior soil survey provides an initial estimate; in regions where profiles are available, the IRC framework uses them to learn additional soil variability and correct prior biases. This design directly benefits land-surface model (LSM) parameterization. LSMs require not only accurate estimates of soil properties such as texture, bulk density, and soil organic matter, but also well-calibrated uncertainty bounds to propagate input uncertainty through simulations of soil water retention, carbon cycling,

and energy partitioning (Baroni et al., 2017; Vereecken et al., 2022). By combining the broad spatial coverage of soil surveys with the local accuracy of georeferenced profiles, the IRC framework produces soil maps that are both spatially complete and locally refined.

5 Conclusion

This study demonstrates that iterative residual correction (IRC) is an effective approach to improve existing probabilistic soil property maps when new georeferenced observations become available. By integrating additional georeferenced soil profiles to adjust pixel-wise distributions of soil properties, the IRC framework directly addresses two of the most persistent limitations in digital soil mapping: systematic bias from "regression-to-the-mean" in DSM models, and overconfident or poorly calibrated uncertainty estimates inherited from the prior polygon-based survey inputs.

The California case study confirms that the research objectives are achieved. The IRC method substantially improves predictive accuracy for all six soil properties examined. The gains are not marginal: R^2 for sand, silt, and clay more than doubled relative to priors; pH posterior R^2 reached 0.94; and the most challenging properties (SOM and bulk density) both of which are temporally dynamic and difficult to map, transitioned from near-zero to meaningful R^2 values (0.61 and 0.70, respectively). The implementation of iterative correction provides additional benefit beyond a single-pass model, particularly for capturing extreme values in skewed distributions (25.6% MAE improvement for the top 1%

of SOM from 100 to 200 cm), confirming that multiple passes contribute to model performance improvement.

The method also improves the spatial fidelity of predictions. Semi-variogram analysis in the case of Central Valley shows that posterior predictions better reproduce the observed sill and effective range, indicating that IRC recovers spatial heterogeneity that the prior model suppresses. Land-use-stratified vertical profiles further show that improvements are consistent across forest, cultivated, and wetland ecosystems, with the largest gains in wetland environments, a result consistent with the known inadequacy of polygon-based surveys in ecologically complex or under-sampled landscapes.

For the field of digital soil mapping, the IRC framework offers three contributions beyond incremental accuracy improvement. First, it provides a scalable pathway for map evolution: as new soil data become available from field campaigns or sensor networks, the framework can assimilate them without rebuilding the underlying prior model, making continuous updating computationally tractable. Second, it enforces physical constraints (e.g., non-negativity, particle-size fraction compositional closure) during correction, ensuring that posterior distributions remain physically realizable (a property not guaranteed by unconstrained regression-based bias correction). Third, by improving both the central tendency and the calibration of uncertainty estimates, the framework produces maps that are more reliable as inputs to risk-sensitive downstream applications, including irrigation scheduling, flood risk assessment, and land-surface model parameterization.

Limitations of the current study include the reliance on OOB evaluation, which may overstate performance in regions of high spatial autocorrelation. Additional limitations are the absence of temporally resolved covariates for dynamic soil properties, and the constraint of a California domain. While this domain is geographically diverse, it does not test the framework's generalization to other climatic or lithological regimes. Future work will apply the IRC framework across the CONUS, evaluate it against spatially independent soil datasets, and may explore extensions to soil hydraulic properties and other ecologically relevant soil attributes. Taken together, this work establishes IRC as a promising and practically deployable method of the next generation of digital soil mapping systems.

Data Availability

Data will be made available on request. Code is available on https://github.com/emmaxu43/IRC_CA/tree/main.

Author Contributions

Chengcheng Xu and Nathaniel Chaney designed the study and developed the methodology. Chengcheng Xu wrote the original draft and wrote the codes to produce the methodology and analyses. Nathaniel Chaney supervised the work, provided resources and funding, and helped guide the research direction. Elia Scudiero provided funding, project management, co-supervision. Elia Scudiero and Ray Anderson provided soil

property samples from California that were used as part of the input dataset. Chengcheng Xu, Nathaniel Chaney, Elia Scudiero, and Ray Anderson discussed the results and contributed to revising and editing the manuscript.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research was supported by the Agriculture and Food Research Initiative Competitive Grant no. 2020-69012-31914 from the USDA National Institute of Food and Agriculture. The authors want to thank Dr. Todd Skaggs for his and his teams' support for gathering input data for this work. His and Dr. Ray Anderson's efforts are supported by USDA-ARS, Office of National Programs (projects 2036-61000-019-000-D and 2036-61000-019-006-R). The U.S. Department of Agriculture prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program (not all prohibited bases apply to all programs). Persons with disabilities who require alternative means for communication of program information (braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call

(800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

Financial Support

The study was supported by USDA-NIFA-AFRI-006739 grant for sustainable agricultural systems.

References

Arrouays, D., McKenzie, N., Hempel, J., Forges, A. R. de, and McBratney, A. B.: GlobalSoilMap: Basis of the global spatial soil information system, CRC Press, 496 pp., 2014.

Baroni, G., Zink, M., Kumar, R., Samaniego, L., and Attinger, S.: Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales, Hydrology and Earth System Sciences, 21, 2301–2320, https://doi.org/10.5194/hess-21-2301-2017, 2017.

Batjes, N. H., Calisto, L., and de Sousa, L. M.: Providing quality-assessed and standardised soil data to support global mapping and modelling (WoSIS snapshot 2023), Earth System Science Data, 16, 4735–4765, https://doi.org/10.5194/essd-16-4735-2024, 2024.

Chaney, N. W., Herman, J. D., Reed, P. M., and Wood, E. F.: Flood and drought hydrologic monitoring: The role of model parameter uncertainty, Hydrology and Earth System Sciences, 19, https://doi.org/10.5194/hess-19-3239-2015, 2015.

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., and Odgers, N. P.: POLARIS: A 30-meter probabilistic soil series map of the contiguous United States, Geoderma, https://doi.org/10.1016/j.geoderma.2016.03.025, 2016.

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., McBratney, A. B., Wood, E. F., and Yimam, Y.: POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties Over the Contiguous United States, Water Resources Research, https://doi.org/10.1029/2018WR022797, 2019.

Chen, C., Liaw, A., and Breiman, L.: Using random forest to learn imbalanced data, University of California, Berkeley, 110, 24, 2004.

Chilès, J.-P. and Delfiner, P.: Geostatistics: modeling spatial uncertainty, in: Geostatistics: modeling spatial uncertainty, John Wiley & Sons, Ltd, 147–237, https://doi.org/10.1002/9781118136188.ch3, 2012.

Corwin, D. L. and Scudiero, E.: Field-scale apparent soil electrical conductivity, Soil Science Society of America Journal, 84, 1405–1441, https://doi.org/10.1002/saj2.20153, 2020.

Grunwald, S., Thompson, J. A., and Boettinger, J. L.: Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues, Soil Science Society of America Journal, 75, 1201–1213, https://doi.org/10.2136/SSSAJ2011.0025, 2011.

Haghverdi, A., Najarchi, M., öztürk, H. S., and Durner, W.: Studying unimodal, bimodal, PDI and bimodal-PDI variants of multiple soil water retention models: I. Direct model fit using the extended evaporation and dewpoint methods, Water (Switzerland), 12, https://doi.org/10.3390/w12030900, 2020.

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., and Zhang, G.-L.: GlobalSoilMap.net – A New Digital Soil Map of the World, in: Digital Soil Mapping: Bridging Research, Environmental Application, and Operation, edited by: Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E., and Kienast-Brown, S., Springer Netherlands, Dordrecht, 423–428, https://doi.org/10.1007/978-90-481-8863-5_33, 2010.

Hengl, T., Heuvelink, G. B., and Stein, A.: A generic framework for spatial prediction of soil variables based on regression-kriging, Geoderma, 120, 75–93, 2004.

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS ONE, https://doi.org/10.1371/journal.pone.0169748, 2017.

Jiang, Q., Fu, Q., and Wang, Z.: Delineating site-specific irrigation management zones, Irrigation and Drainage, 60, 464–472, https://doi.org/10.1002/ird.588, 2011.

Lesch, S., Rhoades, J., and Corwin, D.: ESAP-95 version 2.01 R: User manual and tutorial guide, Research Rpt, 146, 17, 2000.

Lesch, S. M.: Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties, Computers and Electronics in Agriculture, 46, 153–179, https://doi.org/10.1016/j.compag.2004.11.004, 2005.

Li, N., Zhao, X., Wang, J., Sefton, M., and Triantafilis, J.: Digital soil mapping based site-specific nutrient management in a sugarcane field in Burdekin, Geoderma, 340, 38–48, https://doi.org/10.1016/j.geoderma.2018.12.033, 2019.

McBratney, A. B., Mendonça Santos, M. L., and Minasny, B.: On digital soil mapping, Geoderma, 117, 3–52, https://doi.org/10.1016/S0016-7061(03)00223-4, 2003.

Minasny, B. and McBratney, A. B.: A conditioned Latin hypercube method for sampling in the presence of ancillary information, Computers & geosciences, 32, 1378–1388, 2006.

Minasny, B. and McBratney, Alex. B.: Digital soil mapping: A brief history and some lessons, Geoderma, 264, 301–311, https://doi.org/10.1016/j.geoderma.2015.07.017, 2016.

Mueller, T. G., Pierce, F. J., Schabenberger, O., and Warncke, D. D.: Map Quality for Site-Specific Fertility Management, Soil Science Society of America Journal, 65, 1547–1558, https://doi.org/10.2136/sssaj2001.6551547x, 2001.

National Cooperative Soil Survey: NCSS Soil Characterization Database (Lab Data Mart), 2018.

Nauman, T. W., Kienast-Brown, S., Roecker, S. M., Brungard, C., White, D., Philippe, J., and Thompson, J. A.: Soil landscapes of the United States (SOLUS): Developing predictive soil property maps of the conterminous United States using hybrid training sets, Soil Science Society of America Journal, 88, 2046–2065, https://doi.org/10.1002/saj2.20769, 2024.

Nussbaum, M., Zimmermann, S., Walthert, L., and Baltensweiler, A.: Benefits of hierarchical predictions for digital soil mapping—An approach to map bimodal soil pH, Geoderma, 437, 116579, https://doi.org/10.1016/j.geoderma.2023.116579, 2023.

Odgers, N. P., McBratney, A. B., and Minasny, B.: Digital soil property mapping and uncertainty estimation using soil class probability rasters, Geoderma, 237, https://doi.org/10.1016/j.geoderma.2014.09.009, 2015.

Oliver, M. A. and Webster, R.: A tutorial guide to geostatistics: Computing and modelling variograms and kriging, CATENA, 113, 56–69, https://doi.org/10.1016/j.catena.2013.09.006, 2014.

Ortuani, B., Chiaradia, E. A., Priori, S., L'Abate, G., Canone, D., Comunian, A., Giudici, M., Mele, M., and Facchi, A.: Mapping Soil Water Capacity Through EMI Survey to Delineate Site-Specific Management Units Within an Irrigated Field, Soil Science, 181, 252, https://doi.org/10.1097/SS.0000000000000159, 2016.

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/SOIL-7-217-2021, 2021.

Powers, J. S., Corre, M. D., Twine, T. E., and Veldkamp, E.: Geographic bias of field observations of soil carbon stocks with tropical land-use changes precludes spatial extrapolation, Proceedings of the National Academy of Sciences, 108, 6318–6322, https://doi.org/10.1073/pnas.1016774108, 2011.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., and Thompson, J.: Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution, Soil Science Society of America Journal, 82, 186–201, https://doi.org/10.2136/sssaj2017.04.0122, 2018.

Rossiter, D. G., Poggio, L., Beaudette, D., and Libohova, Z.: How well does digital soil mapping represent soil geography? An investigation from the USA, SOIL, 8, 559–586, https://doi.org/10.5194/soil-8-559-2022, 2022.

Schmidinger, J. and Heuvelink, G. B. M.: Validation of uncertainty predictions in digital soil mapping, Geoderma, 437, 116585, https://doi.org/10.1016/j.geoderma.2023.116585, 2023.

Scudiero, E., Corwin, D. L., Markley, P. T., Pourreza, A., Rounsaville, T., Bughici, T., and Skaggs, T. H.: A system for concurrent on-the-go soil apparent electrical conductivity and gamma-ray sensing in micro-irrigated orchards, Soil and Tillage Research, 235, 105899, 2024.

Sharififar, A., Sarmadian, F., Malone, B. P., and Minasny, B.: Addressing the issue of digital mapping of soil classes with imbalanced class observations, Geoderma, 350, 84–92, https://doi.org/10.1016/j.geoderma.2019.05.016, 2019.

Shi, G., Sun, W., Shangguan, W., Wei, Z., Yuan, H., Zhang, Y., Liang, H., Li, L., Sun, X., Li, D., Huang, F., Li, Q., and Dai, Y.: A China dataset of soil properties for land surface modeling (version 2), https://doi.org/10.5194/essd-2024-299, 29 August 2024.

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

Soil Survey Staff: Kellogg Soil Survey Laboratory methods manual, U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln, Nebraska, 2014.

Soil survey staff: Gridded National Soil Survey Geographic (gNATSGO) Database for the Conterminous United States, 2023.

Soil Survey Staff: Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States, 2025.

Soil Survey Staff, Natural Resources Conservation Service, and United States Department of Agriculture: Soil Survey Geographic (SSURGO) Database for the CONUS, 2023.

Sylvain, J.-D., Anctil, F., and Thiffault, É.: Using bias correction and ensemble modelling for predictive mapping and related uncertainty: A case study in digital soil mapping, Geoderma, 403, 115153, https://doi.org/10.1016/j.geoderma.2021.115153, 2021.

Takoutsing, B., Heuvelink, G. B. M., Stoorvogel, J. J., Shepherd, K. D., and Aynekulu, E.: Accounting for analytical and proximal soil sensing errors in digital soil mapping, European Journal of Soil Science, 73, e13226, https://doi.org/10.1111/ejss.13226, 2022.

Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M. H., Amelung, W., Aitkenhead, M., Allison, S. D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., Ghezzehei, T., Hallett, P., Hendricks Franssen, H. J., Heppell, J., Horn, R., Huisman, J. A., Jacques, D., Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., McBratney, A., Minasny, B., Montzka, C., Nowak, W., Pachepsky, Y., Padarian, J., Romano, N., Roth, K., Rothfuss, Y., Rowe, E. C., Schwen, A., Šimůnek, J., Tiktak, A., Van Dam, J., van der Zee, S. E. A. T. M., Vogel, H. J., Vrugt, J. A., Wöhling, T., and Young, I. M.: Modeling Soil Processes: Review, Key Challenges, and New Perspectives, Vadose Zone Journal, 15, vzj2015.09.0131, https://doi.org/10.2136/vzj2015.09.0131, 2016.

Vereecken, H., Amelung, W., Bauke, S. L., Bogena, H., Brüggemann, N., Montzka, C., Vanderborght, J., Bechtold, M., Blöschl, G., Carminati, A., Javaux, M., Konings, A. G., Kusche, J., Neuweiler, I., Or, D., Steele-Dunne, S., Verhoef, A., Young, M., and Zhang, Y.: Soil hydrology in the Earth system, Nat Rev Earth Environ, 3, 573–587, https://doi.org/10.1038/s43017-022-00324-6, 2022.

Wu, Y., Huang, Y., Chen, Z., Yao, Z., Fu, Y., Liu, K., Luo, X., and Wang, D.: Iterative Feature Space Optimization through Incremental Adaptive Evaluation, https://doi.org/10.48550/arXiv.2501.14889, 24 January 2025.

Xu, C., Torres-Rojas, L., Vergopolan, N., and Chaney, N. W.: The Benefits of Using State-Of-The-Art Digital Soil Properties Maps to Improve the Modeling of Soil Moisture in Land Surface Models, Water Resources Research, 59, e2022WR032336, https://doi.org/10.1029/2022WR032336, 2023.

Xu, C., Huang, J., Hartemink, A. E., and Chaney, N. W.: 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, 2025.

Zhang, G. and Lu, Y.: Bias-corrected random forests in regression, Journal of Applied Statistics, 39, 151–160, https://doi.org/10.1080/02664763.2011.578621, 2012.