Accepted Manuscript
Version of record at: https://doi.org/10.1002/ecs2.4170
Received: 3 December 2021 | Revised: 30 March 2022 | Accepted: 7 April 2022
DOI: 10.1002/ecs2.4170

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

ARTICLE

Macrosystems Ecology

Spatial variability and uncertainty of soil nitrogen across the conterminous United States at different depths

Elizabeth M. Smith[1] | **Rodrigo Vargas**[1] | **Mario Guevara**[1] | **Tonantzin Tarin**[1] | **Richard V. Pouyat**[1,2]

[1]Department of Plant and Soil Sciences, University of Delaware, Newark, Delaware, USA

[2]USDA Forest Service, Northern Research Station, NRS-08, Newark, Delaware, USA

Correspondence
Rodrigo Vargas
Email: rvargas@udel.edu

Present address
Mario Guevara, Centro de Geociencias, Universidad Nacional Autónoma de México, Juriquilla, Queretaro, Mexico

Tonantzin Tarin, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

Funding information
NASA, Grant/Award Number: 80NSSC21K1715; National Science Foundation, Grant/Award Numbers: 1724843, 2103836; National Science Foundation Graduate Research Fellowship; University of Delaware

Handling Editor: Meredith K. Steele

Abstract

Soil nitrogen (N) is an important driver of plant productivity and ecosystem functioning; consequently, it is critical to understand its spatial variability from local-to-global scales. Here, we provide a quantitative assessment of the three-dimensional spatial distribution of soil N across the United States (CONUS) using a digital soil mapping approach. We used a random forest-regression kriging algorithm to predict soil N concentrations and associated uncertainty across six soil depths (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm) at 5-km spatial grids. Across CONUS, there is a strong spatial dependence of soil N, where soil N concentrations decrease but uncertainty increases with soil depth. Soil N was higher in Pacific Northwest, Northeast, and Great Lakes National Ecological Observatory Network (NEON) ecoclimatic domains. Model uncertainty was higher in Atlantic Neotropical, Southern Rockies/Colorado Plateau, and Southeast NEON domains. We also compared our soil N predictions with satellite-derived gross primary production and forest biomass from the National Biomass and Carbon Dataset. Finally, we used uncertainty information to propose optimized locations for designing future soil surveys and found that the Atlantic Neotropical, Pacific Northwest, Pacific Southwest, and Appalachian/Cumberland Plateau NEON domains may require larger survey efforts. We highlight the need to increase knowledge of biophysical factors regulating soil processes at deeper depths to better characterize the three-dimensional space of soils. Our results provide a national benchmark regarding the spatial variability and uncertainty of soil N and reveal areas in need of a better representation.

KEYWORDS
machine learning, MODIS, nutrients, representativeness, soil modeling

INTRODUCTION

Soil nutrients are influenced by several biophysical factors, resulting in a heterogeneous distribution across landscapes and within the soil profile (Bartholomew & Clark, 1965). More specifically, soil nitrogen (N) has characteristically high spatial variability because of extrinsic (e.g., land use and land management) and intrinsic factors (e.g., soil type, parent material, and climate) (Gao et al., 2019). Although spatial variability of soil properties is difficult to observe directly, it influences the potential uses of soil resources, making it imperative to quantify (Campbell, 1979). Recent studies have used various statistical modeling techniques to map the spatial distribution of soil properties or characteristics across different scales (Adhikari et al., 2014; Malone et al., 2009; Ramcharan et al., 2018). These have been excellent efforts, but there is a need to include information from deeper depths, include uncertainty estimates to better interpret the results, and use this information to identify gaps in the spatial information (Stell et al., 2021). We advocate to continue improving soil monitoring and mapping efforts to better represent the lateral and vertical spatial distribution of soil properties including soil N.

Understanding the spatial distribution of soil properties (e.g., soil N) and their relationship with depth is important because variability extends across latitude, longitude, and depth (Poggio & Gimona, 2014). However, most measurement efforts are driven by shallow soil surveys or shallow installation of sensors that often focus on the topsoil (0–30 cm) and that are limited in range due to cost and accessibility (Burton et al., 2020). Looking deeper into the soil and measuring soil properties at deeper depths (e.g., >30 cm depth) is important to test expected functional relationships, such as the effects of soil moisture, temperature, or precipitation on distribution, that have been usually established using shallow measurements (LeBauer & Treseder, 2008; Vargas et al., 2010).

Upscaling in situ soil measurements is a research area of ongoing development (Wadoux & McBratney, 2021). Traditionally, it has been done using polygon-based approaches where equal values are given to defined categories, such as soil types (Cheng et al., 2019), or by parameterizing process-based models to predict soil N across space and time (Grunwald, 2009). Spatial modeling could offer insight into trends, how changes in climate, environmental contaminants, and land management can alter soil properties by utilizing measurements from monitoring networks (Kimsey et al., 2020). However, these efforts have usually missed facets such as depth, uncertainty estimates, and information about ecological or functional relationships (Villarreal & Vargas, 2021).

Digital soil mapping (DSM) is a computer-assisted approach to upscale and produce maps of soil attributes (Zeraatpisheh et al., 2017). An added value of this approach is the growing capability of new approaches to produce uncertainty estimates along with predictions (Guevara et al., 2020). For example, model prediction error provides information about the difference between observed data and model outputs (Malone et al., 2011; Shrestha & Solomatine, 2006). Previous studies have predicted soil C or N at fine spatial scales (Chaney et al., 2016; Morellos et al., 2016; Ramcharan et al., 2018), but they did not provide spatially explicit uncertainty estimates. Other regional studies found that uncertainty is lower in areas with more site-specific data than areas with less data (Kidd et al., 2015; Stell et al., 2021). That said, it has been highlighted the need to account for uncertainty estimates about soil N-related variables (Heuvelink, 2018) to improve the interpretations of soil variability patterns. Reporting of uncertainty estimates is integral for transparency, assessing model outputs, and providing insights for monitoring network design, but is a practice that has not been consistently performed in local-to-global estimates. Specifically, we argue that information of uncertainty provides insights for improvement of monitoring network designs as has been explored in previous studies (Hargrove et al., 2003; Villarreal et al., 2018; Yang et al., 2008).

While spatial predictions (e.g., upscaling) of soil properties are value-added products from local or national inventories, it is important to connect these predictions to ecological processes and identify functional relationships (LeBauer & Treseder, 2008). For example, it is needed to test how these value-added products relate with ecological fluxes and pools such as gross primary productivity (GPP) or biomass. Stand-level studies have found canopy structure is optimized for N and water supply allocation (Peltoniemi et al., 2012). However, complexities such as environmental conditions, site fertility, and plant traits complicate interpretations of such studies at different spatial scales (Tian et al., 2021). Ecosystem level studies found that forests with high GPP exhibited high net ecosystem production (NEP) only in nutrient-rich forests, while nutrient-poor forests had higher ecosystem respiration rates and consequently lower NEP (Fernández-Martínez et al., 2014). However, at large spatial scales the distribution of soil nutrients has been found to affect the distribution and productivity of plants in different ecosystems (García-Palacios et al., 2012; Mou et al., 1995; Robertson et al., 1988). We postulate that these expected relationships should be tested across different spatial scales and use information derived from different soil depths (not just surface information, e.g., <10 cm depth) to evaluate the generality of previous observations and the applicability of value-added products.

The overarching goal of the present work is to provide quantitative information of the spatial distribution

Accepted Manuscript
Version of record at: https://doi.org/10.1002/ecs2.4170

(at 5-km spatial grids) of total soil N (percent by weight of organic and inorganic N in an oven-dried sample) from the International Soil Carbon Network (ISCN; Nave et al., 2016) and associated uncertainties across the conterminous United States (CONUS). We developed a machine learning framework based on DSM to predict soil N and associated uncertainties in a three-dimensional space (i.e., latitude, longitude, and six soil depths). We postulate that uncertainties in soil N predictions will increase with soil depth as less information (i.e., soil N data from surveys) is available for training model estimates. Furthermore, we assess the relationship between soil N at each predicted depth (x, y, and z) with GPP information from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) and biomass information from the National Biomass and Carbon Dataset (NBCD) for the year 2000 to identify potential functional relationships across CONUS. We expect to see positive relationships between soil N with GPP and biomass at surface depths (0–30 cm), but higher uncertainty of soil N predictions could influence these expected relationships. Lastly, we use information from uncertainty estimates to identify locations to optimize future potential soil surveys with different sampling sizes (from 500 to 6000 samples) across CONUS. The novelty of this study includes a value-added product (i.e., a raster dataset) describing the spatial variability of soil N and model prediction uncertainty, as well as spatial information about GPP, and biomass. This study provides information useful for modeling benchmarks, characterization of how the spatial distribution of soil N relates to other ecological processes, and insights for future soil surveys across CONUS.

METHODS

This study is focused within CONUS and is comprised of three main tasks: (Task 1) developing a DSM framework to predict the spatial variability and uncertainty estimates of soil N; (Task 2) exploring spatial correlations between soil N with GPP and biomass; and (Task 3) identifying optimized locations for future soil surveys based on information of uncertainty derived from Task 1. We used the 17 National Ecological Observatory Network (NEON) ecoclimatic domains within the CONUS to analyze the data and summarize our results.

Our framework is based on data compiled by the ISCN (Nave et al., 2016) (years 1944–2014), which is a database with 19,292 observations for soil N (percent by weight in an oven-dried sample) at different soil horizon-based depths (Figure 1). We also compiled 21 environmental

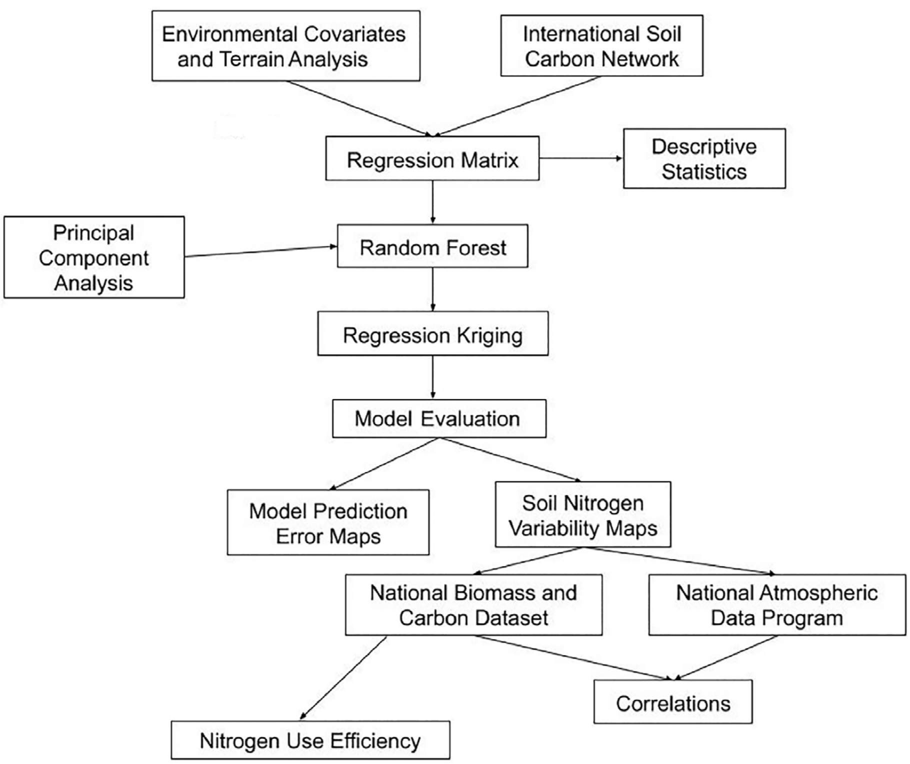

FIGURE 1 Flow diagram of the methodological steps that we used to predict soil N spatial variability, associated uncertainty, and quantify correlations with total N deposition and biomass.

covariates that are important for predicting the spatial variability of soil N (Scull et al., 2003). The covariates are enhanced vegetation index (EVI), geological ages, mean soil moisture, mean accumulated precipitation, mean annual temperature, 5-km digital elevation model (DEM), evapotranspiration, analytical hillshading, slope, aspect, cross-sectional curvature, longitudinal curvature, convergence index, closed depression, flow accumulation, topographic wetness index, slope length and steepness (LS) factor, channel network base level, vertical distance to channel network, valley depth, and relative slope position. Our analyses are summarized using the 17 ecoclimatic domains defined by NEON across the CONUS. We used this network to summarize our findings because each NEON domain represents a region with similar vegetation, climate, and ecosystem functions across the United States (Keller et al., 2008). We discuss details of our framework in the following sections.

Environmental covariates and ancillary data

We harmonized the database of the 21 environmental covariates into a 5 × 5 km regular grid. Climatic covariates such as mean annual temperature and mean annual accumulated precipitation were obtained from climate surfaces for global land areas (Hijmans et al., 2005). Covariates derived from remote sensing such as a DEM, geological ages, and EVI were obtained from SoilGrids1km (Hengl et al., 2014). Additionally, we used long-term (1991–2013) gridded averages of evapotranspiration (Mu et al., 2011; 10.1016/j.rse.2011.02.019) and soil moisture (Guevara & Vargas, 2019; 10.4211/hs.b8f6eae9d89241cf8b5904033460af61). We derive primary and secondary topographic attributes (analytical hillshading, slope, aspect, cross-sectional curvature, longitudinal curvature, convergence index, closed depressions, flow accumulation, topographic wetness index, LS factor, channel network base level, vertical distance to channel network, valley depth, and relative slope position) from DEMs following a standardized approach (Conrad et al., 2015). Information from topographic attributes allowed us to explore the spatial distribution of soil N and generate a general depth-wise pattern of soil N variability using open-source tools for quantitative pedology (Beaudette et al., 2013).

We compiled ancillary information to test for spatial relationships between soil N and GPP or biomass. Daily GPP values were obtained from NASA's MODIS Version 6 GPP product MOD17A2H (Running et al., 2004), and we calculated the long-term annual average (2000–2012) for GPP. Total aboveground, live, dry biomass was obtained from the NBCD 2000 (Kellndorfer et al., 2013). Data from the NBCD were originally reported in metric tons per cell and summarized within 66 USDA Forest Service Forest Inventory and Analysis map zones. We used the *aggregate* and *mosaic* functions in the R raster package (Hijmans et al., 2017) to create a cohesive layer to compare across all soil N depths.

Modeling framework and data analysis

A graphical framework for all the analyses conducted using the ISCN observations with the 21 environmental covariates is presented in Figure 2. We used random forests and regression kriging to derive spatial predictions by applying basic geostatistics to model residuals (Hengl et al., 2004). We used the *fit.gstatModel* function of the R package GSIF (Hengl et al., 2017) to fit a three-dimensional (latitude, longitude, and depth) empirical framework using equal area splines (Bishop et al., 1999; Malone et al., 2009). This approach allowed us to estimate the values of soil N at each one of the six selected depths (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm). The selected depths are standardized depths suggested by the GlobalSoilMap.net consortium (Sanchez et al., 2009). Using recursive feature elimination, a variable reduction technique (Kuhn, 2008), we determined relevant environmental covariates across the full soil profile and at each depth.

In addition to predicting soil N at six depths, we accessed the model prediction error (described as uncertainty) of these predictions using random forests following a previous approach (Meinshausen, 2006). Briefly, in regression kriging the prediction error represents the distance between new point locations and the center of the feature space (Hengl & Toomanian, 2006). For this study, we interpret model prediction error to quantify uncertainty (Malone et al., 2011; Shrestha & Solomatine, 2006). Sources of uncertainty can include errors in the training data, structural error, and randomness of the system modeled (Jansen, 1998).

Cross-validation is a technique for evaluating the accuracy of predictive models (Chen et al., 2021). We used k-fold validation with $k = 10$ to evaluate our model because a larger k results in less bias of the technique (Kuhn & Johnson, 2013). To evaluate the performance of our framework, we used 10-fold cross-validation to derive evaluation metrics such as the explained variance (coefficient of determination), mean absolute error (MAE), and the root mean square error (RMSE). Finally, we identified areas of over and under predicting of soil N by calculating residuals of the model output with respect to the in situ values from the ISCN for each depth.

We tested the spatial dependency of the predicted soil N per depth in each NEON domain using the *autofitVariogram*

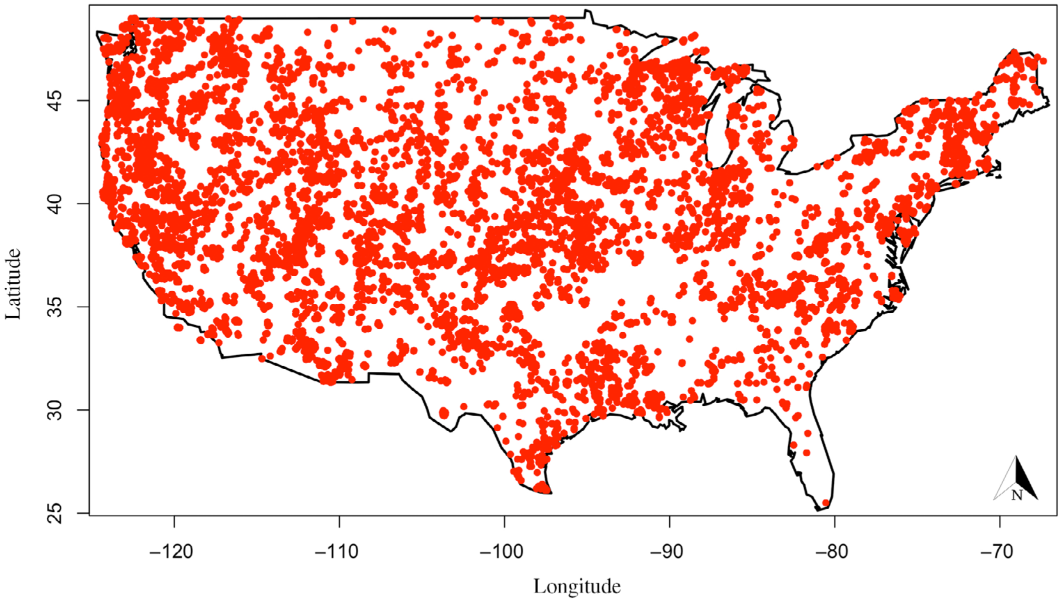

FIGURE 2 International Soil Carbon Network sample points within the conterminous United States. These points were used as training data for the model.

function in R package automap (Hiemstra et al., 2009). Descriptive statistics such as mean, SD of predicted soil N and model prediction error, and the ratio of mean predicted values to model prediction error were calculated for comparison between soil depths. In addition, we explored linear relationships between soil N and GPP or biomass across 5-km grids per depth in each NEON domain (Task 2).

Finally, we used information from our uncertainty estimates of soil N to optimize potential sample locations (Task 3, i.e., inform potential soil surveys). Our premise was that information about uncertainty can be used to optimize future sample locations with the goal of reducing uncertainty in future modeling predictions (Chu et al., 2014; Malone et al., 2019; Stell et al., 2021). A conditioned Latin hypercube sampling analysis (cLHS) was executed using the *clhs* R package (Roudier, 2011) to determine any change in distribution of ancillary data with increased sample points. Previous studies have used a similar approach for identifying locations to optimize environmental monitoring efforts (McKay et al., 1979; Stell et al., 2021; Villarreal et al., 2019). We performed the cLHS analysis to identify locations with sample sizes of 500, 1000, 3000, and 6000 points. The rationale behind these sample sizes is that the Rapid Carbon Assessment included about 6000 randomly selected locations (Soil Survey Staff and Loecke, 2016), the National Forest Inventory included about 3000 locations for soil samples (Domke et al., 2017), and the last two simply represent examples of substantially smaller sample sizes. All analyses were performed using R Studio (RStudio Team, 2021).

RESULTS

The ISCN dataset

The spatial distribution of the available ISCN dataset was disproportionately spread across depths in CONUS (Figure 2) with majority samples located in depth 0–5 cm. The statistical distribution of the dataset was right skewed (skewness: 5.03; kurtosis: 34.23), and most of the soil N values were between 0% and 0.15% by weight in an oven-dried sample. Available measurements of soil N from the ISCN decreased with soil depth: 0–5 cm (n = 13,945), 5–15 cm (n = 1850), 15–30 cm (n = 1778), 30–60 cm (n = 1247), 60–100 cm (n = 311), and 100–200 cm (n = 161). Similarly, the mean soil N from the ISCN dataset decreased with soil depth: 0–5 cm (0.35%), 5–15 cm (0.21%), 15–30 cm (0.13%), 30–60 cm (0.12%), 60–100 cm (0.09%), and 100–200 cm (0.06%).

Environmental covariates

We performed a recursive elimination analysis to determine the relevance of environmental covariates at each depth. Results from the full soil profile (0–200 cm) revealed that mean evapotranspiration, EVI, geological ages, mean accumulated precipitation, and mean annual temperature were the top 5 out of 11 predictors chosen. Other relevant predictors across the full profile were mean soil moisture, DEM, topographic wetness index, LS factor, channel network base level, vertical distance to channel, and vertical distance to channel network. The relevance of environmental covariates was the same across each soil depth. Analytical hillshading, aspect, channel network base level, closed depression, and convergence index were the top five predictors chosen across all depth intervals.

Modeling soil N

A 10-fold cross-validation showed that the average explained variance was 40%, the average RMSE was 0.132, and average MAE was 0.282 across the full profile (Appendix S1: Table S1). Explained variance for our predictions ranged from 38.9% (0–5 cm) to 42.2% (60–100 cm), RMSE ranged from 0.132 (30–60 cm) to 0.135 (60–100 cm), and MAE ranged from 0.279 (0–5 cm) to 0.289 (100–200 cm). Finally, all variograms resulted in a nugget sill ratio of <0.25 for each depth, showing strong spatial dependence for soil N across all depths (Appendix S1: Table S1).

In general, our predictions overestimated soil N values from the ISCN at 0–5 and 5–15 cm (Figure 3a,b) while underestimating soil N values in deeper depths (Figure 3). The highest amount of over estimation (6.73%) occurred in depth 0–5 cm along the east and west coasts of CONUS. Underestimation is more prevalent in deeper depths where higher values were present between 30 and 60 cm, resulting in underestimation up to 0.059%–0.154% at this depth.

We compared the mean and SD of the predicted values as well as those of the model prediction error to better gauge any trends in distribution of predicted soil N across depths for CONUS. The mean predicted soil N (0.096%–0.133%) and SD values (0.067%–0.082%) decreased with depth, while the mean (0.026%–0.051%) and SD (0.008%–0.014%) of the prediction error increased with depth (Table 1). We also compared the ratio of soil N predicted values to model prediction error. The highest ratio was found from 0 to 5 cm (4.181) with the lowest ratio from 100 to 200 cm (1.885) (Table 1). These results highlighted that there was more variability in surface depths than in deeper depths as the mean and SD of

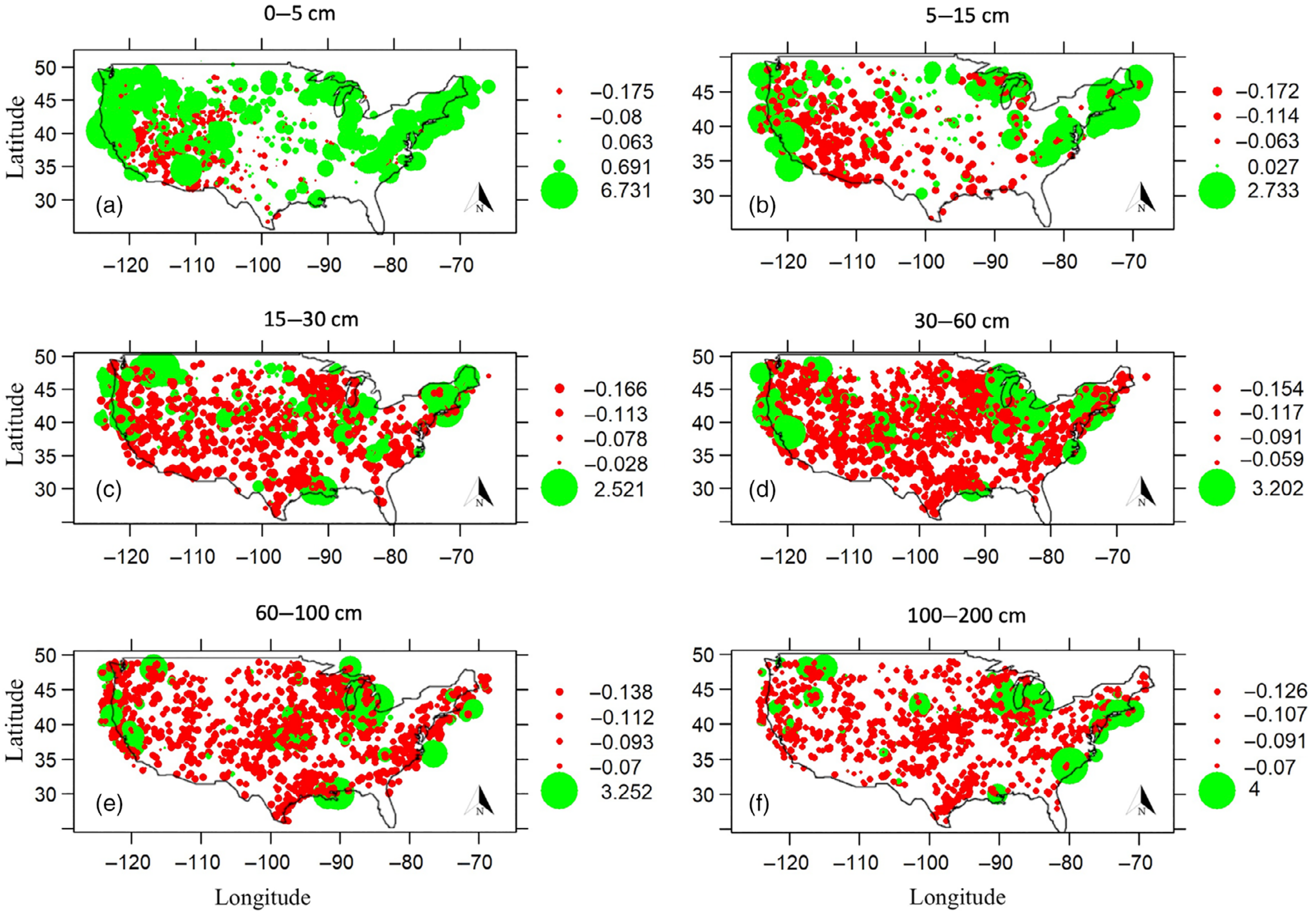

FIGURE 3 Model accuracy maps highlight where the model is over- or underestimating predicted values. The units are presented as percent (%) in an oven-dried soil sample. Underestimation increases with depth. Green dots represent areas of overestimation while red dots represent areas of underestimation. The percent of over- or underestimation corresponds to the size of the dot.

Accepted Manuscript
Version of record at: https://doi.org/10.1002/ecs2.4170

these ratios decrease with depth (Table 1). Finally, the SD of the predicted values (0.067%–0.082%) was lower than the SD of the observed values (0.197%–0.488%) (Table 1; Appendix S1: Table S1).

Spatial variability and associated uncertainty of soil N

We generated 5-km spatial resolution maps to show the variability of soil N at six depths across CONUS (Figure 4). Overall, soil N decreased with depth following the pattern observed for the ISCN dataset (i.e., training dataset) where mean soil N was 0–5 cm (0.35%), 5–15 cm (0.21%), 15–30 cm (0.13%), 30–60 cm (0.12%), 60–100 cm (0.09%), and 100–200 cm (0.06%).

We analyzed in detail the results of soil N from our model for each one of the 17 NEON domains for 0–5 cm depth, as it is the depth with highest amount of soil N (Table 2). The Pacific Northwest had the highest soil N (0.251% ± 0.108) followed by the Northeast (0.235% ± 0.105) (Table 2). By contrast, the Desert Southwest (0.071 ± 0.38) and the Southern Plains (0.089% ± 0.030) had the lowest concentrations.

TABLE 1 Descriptive statistics of percent soil N and model prediction error for each depth across conterminous United States.

Depth (cm)	Soil nitrogen (%)		Prediction error		Soil nitrogen/prediction error	
	Mean	SD	Mean	SD	Mean	SD
0–5	0.133	0.082	0.032	0.008	4.181	9.886
5–15	0.130	0.081	0.029	0.008	4.447	9.627
15–30	0.125	0.077	0.027	0.008	4.631	9.485
30–60	0.155	0.072	0.026	0.008	4.430	9.606
60–100	0.105	0.069	0.031	0.008	3.329	8.352
100–200	0.096	0.067	0.051	0.014	1.885	4.642

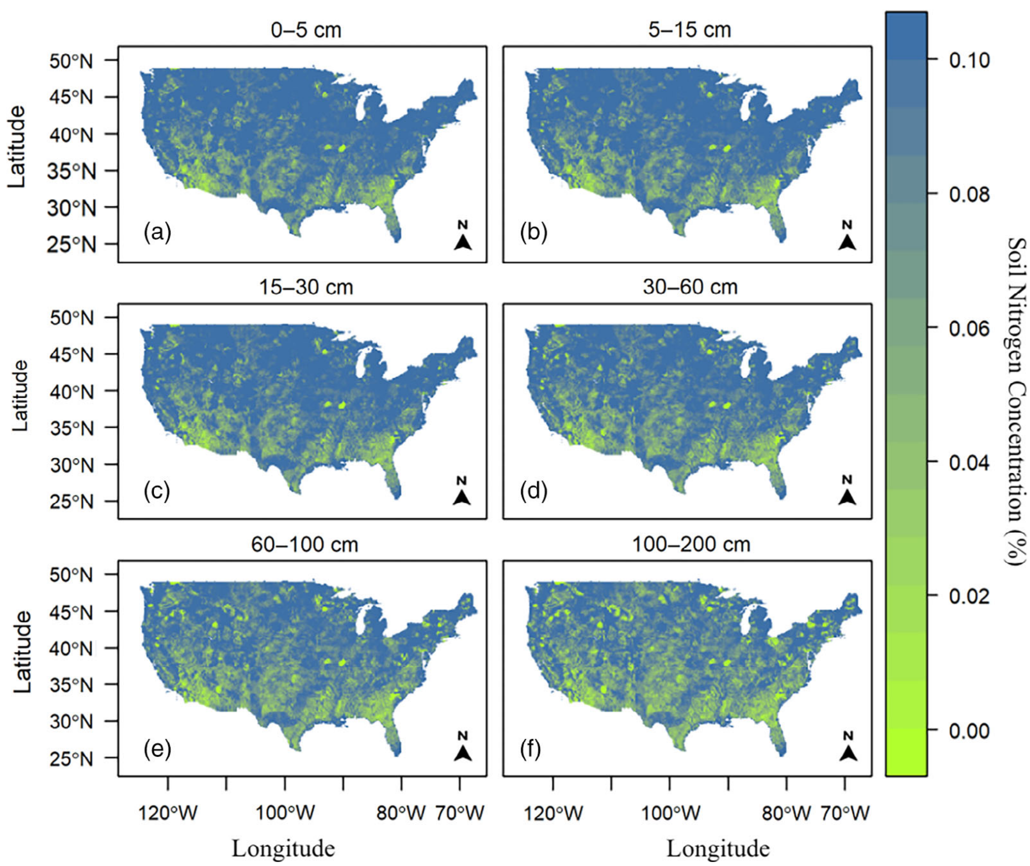

FIGURE 4 Spatial variability of soil nitrogen per depth for baseline year 2016. Soil N concentration (in percent) is higher in surface-level depths (0–60 cm) than in deeper depths (60–200 cm).

TABLE 2 Descriptive statistics of percent soil nitrogen and prediction error, from 0 to 5 cm per National Ecological Observatory Network (NEON) domain across the conterminous United States.

	Soil N (%)		Prediction error		Soil N/prediction error	
NEON domain	Mean	SD	Mean	SD	Mean	SD
Appalachians/Cumberland Plateau	0.150	0.062	0.033	0.006	4.51	10.944
Atlantic Neotropical	0.157	0.073	0.100	0.012	1.564	5.912
Central Plains	0.094	0.028	0.031	0.004	3.006	6.251
Desert Southwest	0.071	0.038	0.032	0.008	2.225	4.709
Great Basin	0.120	0.057	0.027	0.004	4.511	13.573
Great Lakes	0.208	0.107	0.033	0.007	6.389	15.351
Mid-Atlantic	0.124	0.069	0.033	0.005	3.807	12.955
Northeast	0.235	0.105	0.030	0.006	7.897	16.692
Northern Plains	0.136	0.053	0.032	0.006	4.197	8.599
Northern Rockies	0.187	0.073	0.028	0.005	6.615	13.315
Ozarks Complex	0.099	0.040	0.037	0.008	2.685	4.926
Pacific Northwest	0.251	0.108	0.026	0.003	9.511	34.783
Pacific Southwest	0.132	0.074	0.027	0.005	4.815	13.572
Prairie Peninsula	0.135	0.052	0.032	0.005	4.228	10.616
Southeast	0.093	0.010	0.039	0.084	2.416	0.123
Southern Plains	0.089	0.030	0.034	0.007	2.632	4.607
Southern Rockies/Colorado Plateau	0.121	0.061	0.028	0.005	4.337	12.643

Spatial predictions differed from our hypothesis in that model uncertainty did not increase with depth (Figure 5, Table 1). The overall uncertainty across the CONUS was lowest at depths 30–60 cm (0.0261% ± 0.008) but higher at 100–200 cm (0.051% ± 0.014; Table 2).

Relationships across NEON domains and soil depths

At depth 0–5 cm, uncertainty was highest in the Atlantic Neotropical (0.100%). The lowest mean uncertainty was observed in depth 0–5 cm of the Pacific Northwest (0.026%) (Table 2). As expected, uncertainty was highest in the deepest depth (100–200 cm), ranging from 0.032% to 0.051%. It is important to note the consistent high levels of uncertainty across the Atlantic Neotropical domain at all depths. When comparing the soil N and prediction error ratios for NEON regions at 0–5 cm, we can better see the pattern of high variability associated with areas that have higher amounts of soil N. The highest SD of the ratio are seen in areas that had some of the lowest rates of uncertainty (Pacific Northwest, Northeast, and Great Lakes; Table 2).

The Pacific Northwest (169.421 kg/m^2) and the Northeast (101.344 kg/m^2) reported the highest mean biomass, while the lowest mean was reported in the Central Plains (0.828 kg/m^2). A linear regression analysis was executed for all depths to explore soil N and biomass relationships. We found that r^2 decreased with depth 0–5 cm (0.53), 5–15 cm (0.52), 15–30 cm (0.50), 30–60 cm (0.45), 60–100 cm (0.40), and 100–200 cm (0.36), demonstrating a moderate relationship between biomass and soil N with depth (Table 3).

The Great Lakes (0.998 kg C/m^2) and the Atlantic Neotropical (0.649 kg C/m^2) reported the highest mean GPP, while the lowest mean was reported in the Central Plains (0.034 kg C/m^2) (Table 3). Like the biomass analysis, we searched for depth relationships between soil N and GPP using a linear regression approach, but we did not find significant relationships at $\alpha = 0.05$ (Table 4).

Conditioned Latin hypercube sampling analysis

A cLHS analysis was performed with sample sizes of 500, 1000, 3000, and 6000 points (Figure 6, Table 5). As the number of available points increased, the percentage of points within each NEON domain remained about the same. In other words, the percent of needed points within a NEON domain was independent of the overall

Accepted Manuscript
Version of record at: https://doi.org/10.1002/ecs2.4170

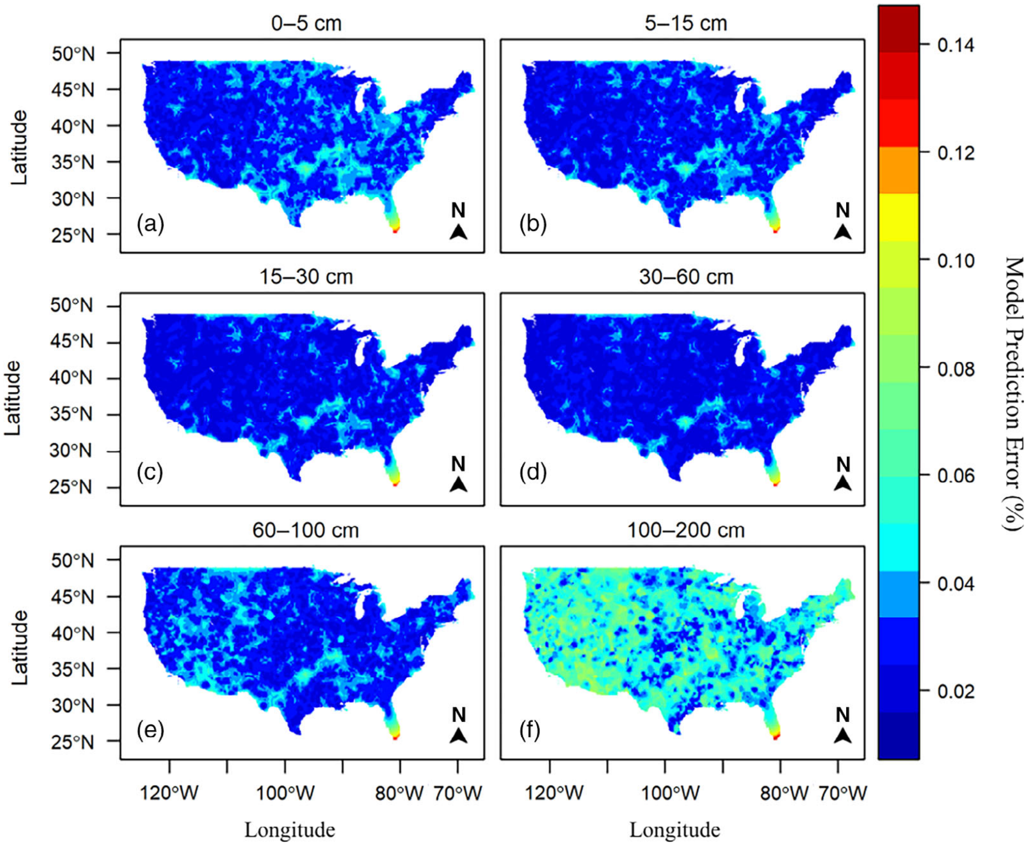

FIGURE 5 Model prediction error across the conterminous United States.

TABLE 3 Descriptive statistics (means) of forest biomass and gross primary productivity (GPP) per National Ecological Observatory Network (NEON) domain across the conterminous United States.

NEON domain	Biomass (kg/m^2)	GPP (kg C/m^2)
Appalachians/Cumberland Plateau	77.37	0.1
Atlantic Neotropical	18.80	0.649
Central Plains	0.83	0.034
Desert Southwest	2.29	0.366
Great Basin	14.85	0.197
Great Lakes	44.22	0.998
Mid-Atlantic	68.93	0.206
Northeast	101.34	0.32
Northern Plains	3.62	0.052
Northern Rockies	52.15	0.051
Ozarks Complex	52.46	0.101
Pacific Northwest	169.42	0.136
Pacific Southwest	58.44	0.259
Prairie Peninsula	13.53	0.121
Southeast	49.75	0.237
Southern Plains	10.22	0.08
Southern Rockies/Colorado Plateau	22.04	0.051

TABLE 4 Linear regression analysis and r^2 values for soil N and forest biomass, and gross primary productivity (GPP) at all depths across the conterminous United States.

Soil depth (cm)	Biomass (kg/m^2)	GPP
0–5	0.53*	0.09
5–15	0.52*	0.1
15–30	0.50*	0.12
30–60	0.45*	0.15
60–100	0.40*	0.2
100–200	n.s.	n.s.

Abbreviation: n.s., not significant.
*Denotes significant p value at $p \leq 0.05$.

sample size of a hypothetical survey (from 500 to 6000 points). The Northern Plains consistently held the highest percentage of points per domain across all sample sizes, while the Atlantic Neotropical held the lowest percentage (Figure 6, Table 5).

DISCUSSION

We developed a framework for determining the three-dimensional spatial variability of soil N and associated

Accepted Manuscript
Version of record at: https://doi.org/10.1002/ecs2.4170

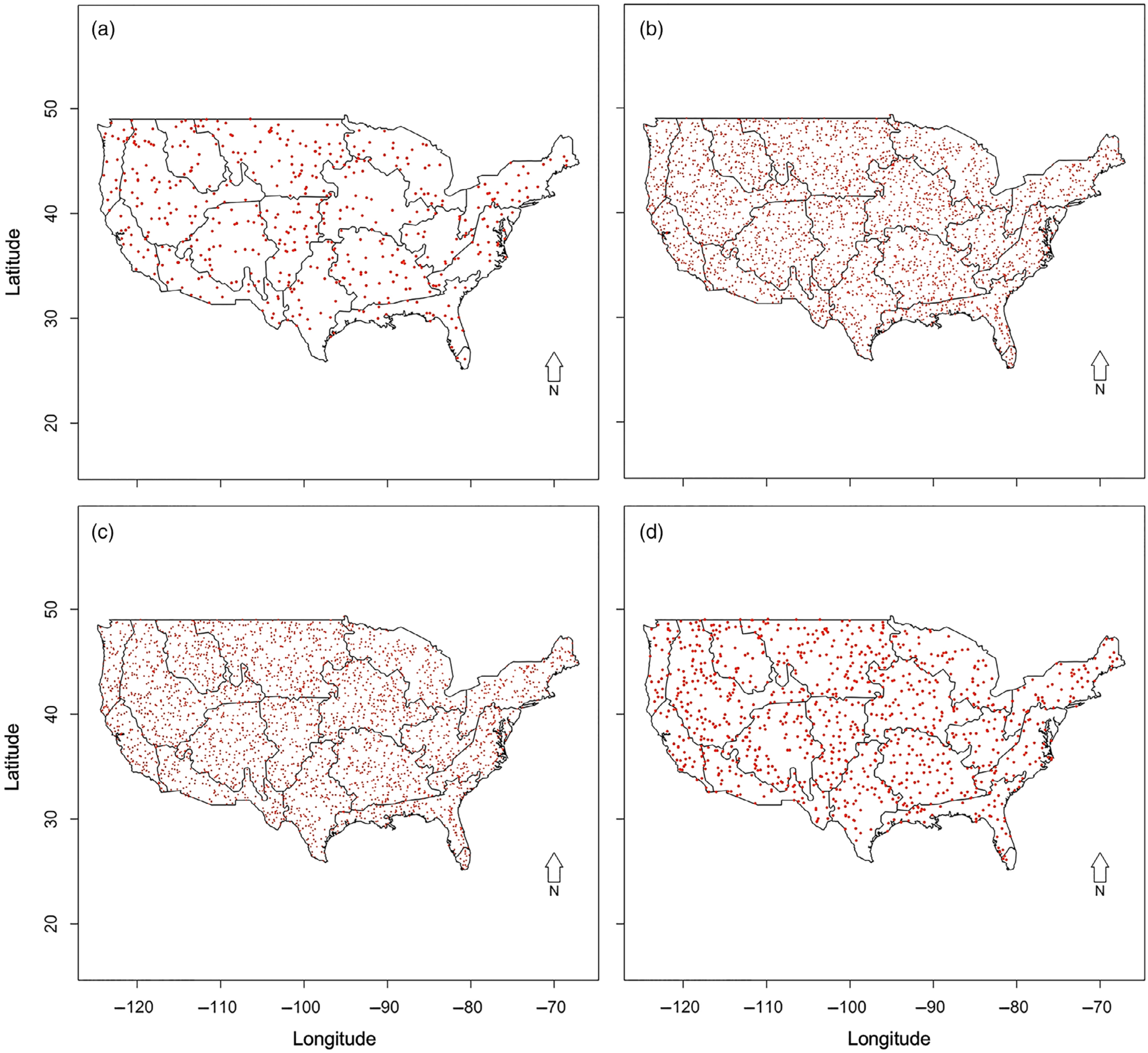

FIGURE 6 Conditioned Latin hypercube sampling analysis (cLHS) for (a) 500 points, (b) 1000 points, (c) 3000 points, and (d) 6000 points.

model uncertainties. These predictions and associated uncertainties were then used as independent variables in a regression analysis with GPP and biomass across CONUS. Additionally, uncertainty values were used to create a multivariate space to determine optimized locations for hypothetical soil surveys with different sample sizes. Our results did not support the first hypothesis with the expectation that uncertainty would consistently increase with soil depth. However, uncertainty was highest from 100 to 200 cm, likely due to less information being available to parameterize the model at deeper depths. We discuss these findings in detail in the following sections.

Our results are congruent with previous studies demonstrating the efficacy of using environmental covariates based on land use and remotely sensed data for landscape scale mapping of soil N (Lamsal et al., 2009). Through the recursive feature elimination analysis, the same five topographic covariates were found important at all depths; however, a similar study found that covariates such as maximum temperature, topographic wetness index, DEM, and slope characterized the overall multivariate environmental space for 0–30 cm depth (Priyadharshini et al., 2021). We highlight that the multivariate relations at different depths are not equal, as deeper depths were more difficult to characterize,

TABLE 5 Conditioned Latin hypercube sampling analysis points results from full profile (0–200 cm) for each National Ecological Observatory Network (NEON) domain to identify potential observation locations.

NEON domain	500 locations	1000 locations	3000 locations	6000 locations
Appalachians/Cumberland Plateau	18 (3.60)	31 (3.10)	110 (3.67)	205 (3.42)
Atlantic Neotropical	1 (0.20)	3 (0.30)	9 (0.30)	18 (0.30)
Central Plains	30 (6.00)	65 (6.50)	167 (5.57)	339 (5.65)
Desert Southwest	34 (6.80)	49 (4.90)	160 (5.33)	306 (5.10)
Great Basin	51 (10.20)	126 (12.60)	322 (10.73)	676 (11.27)
Great Lakes	33 (6.60)	61 (6.10)	193 (6.43)	370 (6.17)
Mid-Atlantic	20 (4.00)	36 (3.60)	142 (4.73)	282 (4.70)
Northeast	36 (7.20)	51 (5.10)	162 (5.40)	315 (5.25)
Northern Plains	58 (11.60)	126 (12.60)	365 (12.17)	740 (12.33)
Northern Rockies	15 (3.00)	40 (4.00)	150 (5.00)	294 (4.90)
Ozarks Complex	31 (6.20)	82 (8.20)	232 (7.73)	465 (7.75)
Pacific Northwest	11 (2.20)	27 (2.70)	91 (3.03)	177 (2.95)
Pacific Southwest	16 (3.20)	34 (3.40)	94 (3.13)	191 (3.18)
Prairie Peninsula	37 (7.40)	76 (7.60)	246 (8.20)	490 (8.17)
Southeast	30 (6.00)	47 (4.70)	135 (4.50)	269 (4.48)
Southern Plains	27 (5.40)	55 (5.50)	169 (5.63)	350 (5.83)
Southern Rockies/Colorado Plateau	48 (9.60)	80 (8.00)	233 (7.77)	470 (7.83)

Note: Values are the number of samples needed within each NEON domain, and associated percentage (in parentheses) for each sampling scenario.

possibly a result of more complex relationships among the covariates. Recent studies concluded similar findings, attributing a decrease in the influence of variables with increasing depth (Kokulan et al., 2018). Thus, it is critical to increase our understanding of the biophysical factors that regulate soil processes at deeper depths to better characterize the three-dimensional space of soils.

Our predicted model explained at least 40% of the variance of soil N at all depths. When comparing the model to the training data, we see that the SD of the predicted values (0.067%–0.082%) was lower than the SD of the observed values (0.197%–0.488%), indicating that variability of the predicted values is less than that of the observed values (Table 1; Appendix S1: Table S1). Consequently, model predictions appear to be more constrained than the actual variability observed in the ISCN dataset (i.e., training data).

We found that the spatial distribution of soil N became more variable with depth and differs across NEON domains (Table 2). Previous studies have quantified soil N on global and regional scales (Batjes, 1996; Wang et al., 2017; Zhou et al., 2020), and our results mirror those of Wang et al. (2017), in which the highest amount of soil N was found in the top 0–5 cm. For example, the Great Lakes, Northeast, and Pacific Northwest had the highest amounts of soil N from 0 to 5 cm across the CONUS. This was expected as studies have shown increased available nitrogen in the Northeast (Butler et al., 2012) and that soils in the Pacific Northwest have characteristics associated with high soil N availability (Littke et al., 2014). We clarify that our estimate of soil N is derived from data collected throughout several decades and compiled by the ISCN; therefore, it represents a baseline of potential soil N across the CONUS. Our spatial predictions are useful for representing general spatial trends, modeling, benchmarking, and for identifying areas where future soil surveys could focus to improve the spatial representation of in situ soil N information. We postulate that increasing information about soil N will better characterize the spatial heterogeneity across CONUS and its relationships with environmental covariates resulting in better model-data agreement and ultimately lower model uncertainty.

Available soil data from the ISCN decreased with depth and arguably affected our capacity to generate stronger soil N and depth relationships. To address this, we used the equal area spline method to standardize all profile data at specific soil depths. However, this method can be sensitive to abrupt changes in the soil profile, which are not necessarily represented by the available data (Hengl et al., 2014), adding another source of uncertainty. Therefore, we highlight the need to continue developing soil covariates that could improve the representation of the variability of soil layers (and soil horizons) that ultimately influence the variability of soil N and other important properties (Hengl & MacMillan, 2019; Yuxin et al., 2021).

Our multiple model runs were consistent, showing a low RMSE across all depths. Mean prediction error was lower at depths 5–15, 15–30, 30–60, and 60–100 cm than at 0–5 cm but with an overall increase in model error with depth from 0–5 to 100–200 cm (Figure 5). This finding was consistent with other studies exploring the spatial variability of soil properties (Adhikari et al., 2014; Wang et al., 2017). A possible explanation is that at deeper depths there is less available soil N data points for training the model and consequently model uncertainty increases. An alternative explanation for this pattern is that the terrain parameters (i.e., independent variables) may not be fully representative of the spatial variability of soil N at deeper depths (Adhikari et al., 2014; Minasny et al., 2006). We clarify that model uncertainty in this study represents the spread (or consistency) among multiple model runs. Consequently, low available training points could result in good model-data agreement (as in our case) but with large model uncertainty. This could represent that model runs are relatively different (in some areas across the CONUS) due to nonlinear interactions between limited available training data and their relationships with environmental covariates (i.e., independent data). We further recognize that one of the largest sources of uncertainty for calculating soil stocks is determining appropriate soil bulk densities along with pedotransfer functions (Guevara et al., 2020). To avoid this additional source of uncertainty, we only report the spatial distribution of soil N as percent across CONUS. We advocate for a better quantification of soil bulk density across available soil data in the ISCN and other soil data repositories (National Academies of Sciences, 2021).

We postulate that reporting both spatial predictions and model prediction uncertainty increases the interpretability and transparency of model outputs (Adhikari et al., 2014). Furthermore, information about the spatial distribution of model uncertainty can be used to improve monitoring efforts. For example, we bring attention to the lack of information across the Atlantic Neotropical, where uncertainty is high in part because of the limited data and the geomorphological conditions of this region that are substantially different than other regions (with more data and arguably better parameterized) across the CONUS.

We found evidence that the expected relationship between soil N, biomass, and GPP is influenced by soil depth. Our results support the expected relationship between soil N and biomass. Previous studies have found that aboveground live dry biomass is positively related to soil characteristics such as soil N (Laurance et al., 1999), that aboveground biomass and its distribution among size classes are related to soil nutrients (Paoli et al., 2008), and that atmospheric nitrogen deposition influences tree growth and survival (Horn et al., 2018). Contrary to our expectations, we did not find a significant relationship between soil N and GPP across CONUS. First, it is known that GPP may be influenced by a wider range of environmental drivers across larger spatial scales (Sun et al., 2018). Second, there could be a mismatch between the information from the ISCN dataset (i.e., data from 1944 to 2014) with satellite-derived GPP representing years 2000–2012, where land use change and disturbances may have influenced potential relationships. Third, it is known that total soil N (i.e., bulk soil N) is a poor proxy for available soil N (Beauchamp et al., 2003; Warren & Whitehead, 1988), and therefore, future analyses should focus on mapping available soil N to test the soil N-GPP relationship at regional scales. Consequently, there is a need to reconcile available information from national soil inventories with models and local-to-regional observations. This effort will allow testing how information from different soil depths influences functional relationships and could be used as benchmarks for sensitivity analyses in process-based models.

Finally, we postulate that uncertainty information provides insights for optimization of sampling designs. Using the calculated model uncertainty, we determine the optimal number of samples needed across each NEON domain. The Northern Plains required the highest number of samples followed by the Great Basin and Southern Rockies/Colorado Plateau. Possible reasons for this could be domain size, geologic history, or terrain and available information within the current ISCN soil database. Furthermore, our results demonstrate that the probability distributions of the optimized locations are independent of the sample size of the hypothetical survey. This information could be used to guide future soil surveys across CONUS, but much more research is needed to test how optimization of in situ information could reduce overall uncertainty in upscaled soil products (Stell et al., 2021).

In this study, we provide gridded estimates of total N as a percent of an oven-dried sample and uncertainty in CONUS, as well as the relationship between soil N, biomass, and GPP. These predictions and relationships are derived from the ISCN and could be used as insights for model benchmarking and ecological applications. We found evidence that climatic and biological soil forming factors are more relevant closer to the surface (0–60 cm) and that topographic features (mainly the analytical hillshading and the slope of terrain) and hydrological related sources of information (evapotranspiration and soil moisture) are more relevant at deeper depths (60–100 cm). We found that the highest concentration of soil N was in the Pacific Northwest, Pacific Southwest, Northern Rockies, Northern Plains, Great Lakes, Prairie Peninsula, Appalachians and Cumberland Plateau, and the Northeast. By contrast, the lowest concentrations of soil N were found in the Desert Southwest, Southeast, Central Plains, and Southern Plains. Overall, soil N tends to decrease with depth, but model prediction uncertainty increases with depth. In general, uncertainty was higher in NEON ecoclimatic regions with less available data and future soil surveys should place special attention in collecting samples across the Atlantic Neotropical, Pacific Northwest, Pacific Southwest, and Appalachian/Cumberland Plateau NEON domains.

ACKNOWLEDGMENTS

Elizabeth M. Smith acknowledges support from the University of Delaware (Graduate Scholar Award) and a National Science Foundation Graduate Research Fellowship. Rodrigo Vargas acknowledges support from the National Science Foundation (Nos. 1724843 and 2103836) and NASA (No. 80NSSC21K1715). We thank Courtney Rempfer for early support organizing databases.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data (Smith et al., 2022) are available from Figshare: https://doi.org/10.6084/m9.figshare.19397156.v1.

ORCID

Elizabeth M. Smith https://orcid.org/0000-0002-1440-7747
Rodrigo Vargas https://orcid.org/0000-0001-6829-5333
Mario Guevara https://orcid.org/0000-0002-9788-9947
Tonantzin Tarin https://orcid.org/0000-0001-5125-579X

REFERENCES

Adhikari, K., A. E. Hartemink, B. Minasny, R. B. Kheir, M. B. Greve, and M. H. Greve. 2014. "Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark." *PLoS One* 9(8): e105519. https://doi.org/10.1371/journal.pone.0105519.

Bartholomew, W. V., and F. E. Clark. 1965, 1965. *Soil Nitrogen*. Madison, WI: American Society of Agronomy. https://dl.sciencesocieties.org/publications/books/tocs/agronomymonogra/soilnitrogen.

Batjes, N. H. 1996. "Total Carbon and Nitrogen in the Soils of the World." *European Journal of Soil Science* 47: 151–63. https://doi.org/10.1111/j.1365-2389.1996.tb01386.x.

Beauchamp, E. G., R. Pararajasingham, and B. D. Kay. 2003. "Relationships of Total Soil Nitrogen to Several Soil Nitrogen Indices." *Communications in Soil Science and Plant Analysis* 34(3–4): 505–18. https://doi.org/10.1081/CSS-120017835.

Beaudette, D. E., P. Roudier, and A. T. O'Geen. 2013. "Algorithms for Quantitative Pedology: A Toolkit for Soil Scientists." *Computers and Geosciences* 52: 258–68. https://doi.org/10.1016/j.cageo.2012.10.020.

Bishop, T. F. A., A. B. McBratney, and G. M. Laslett. 1999. "Modeling Soil Attribute Depth Functions with Equal-Area Quadratic Smoothing Splines." *Geoderma* 91: 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8.

Burton, L., K. Jayachandran, and S. Bhansali. 2020. "Review—The 'Real-Time' Revolution for In Situ Soil Nutrient Sensing." *Journal of the Electrochemical Society* 167(3): 037569. https://doi.org/10.1149/1945-7111/ab6f5d.

Butler, S. M., J. M. Melillo, J. E. Johnson, J. Mohan, P. A. Steudler, H. Lux, E. Burrows, et al. 2012. "Soil Warming Alters Nitrogen Cycling in a New England Forest: Implications for Ecosystem Function and Structure." *Oecologia* 168(3): 819–28. https://doi.org/10.1007/s00442-011-2133-7.

Campbell, J. B. 1979. "Spatial Variability of Soils." *Annals of the Association of American Geographers* 69(4): 544–56.

Chaney, N. W., E. F. Wood, A. B. McBratney, J. W. Hempel, T. W. Nauman, C. W. Brungard, and N. P. Odgers. 2016. "POLARIS: A 30-Meter Probabilistic Soil Series Map of the Contiguous United States." *Geoderma* 274: 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025.

Chen, S., X. Hanyi, X. Dongyun, W. Ji, S. Li, M. Yang, H. Bifeng, et al. 2021. "Evaluating Validation Strategies on the Performance of Soil Property Prediction from Regional to Continental Spectral Data." *Geoderma* 400: 115159. https://doi.org/10.1016/j.geoderma.2021.115159.

Cheng, W., A.-x. Zhu, C.-z. Qin, and F. Qi. 2019. "Updating Conventional Soil Maps by Mining Soil–Environment Relationships from Individual Soil Polygons." *Journal of Integrative Agriculture* 18(2): 265–78. https://doi.org/10.1016/S2095-3119(18)61938-0.

Chu, H.-J., R.-A. Chen, Y.-H. Tseng, and C.-K. Wang. 2014. "Identifying LiDAR Sample Uncertainty on Terrain Features from DEM Simulation." *Geomorphology* 204: 325–33. https://doi.org/10.1016/j.geomorph.2013.08.016.

Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, and J. Böhner. 2015. "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4." *Geoscientific Model Development* 8: 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Domke, G. M., C. H. Perry, B. F. Walters, L. E. Nave, C. W. Woodall, and C. W. Swanston. 2017. "Toward Inventory-Based Estimates of Soil Organic Carbon in Forests of the United States." *Ecological Applications* 27(4): 1223–35. https://doi.org/10.1002/eap.1516.

Fernández-Martínez, M., S. Vicca, I. A. Janssens, J. Sardans, S. Luyssaert, M. Campioli, F. S. Chapin, et al. 2014. "Nutrient Availability as the Key Regulator of Global Forest Carbon Balance." *Nature Climate Change* 4(6): 471–6. https://doi.org/10.1038/nclimate2177.

Gao, X.-s., Y. Xiao, L.-j. Deng, Q.-q. Li, W. Chang-quan, B. Li, O.-p. Deng, and M. Zeng. 2019. "Spatial Variability of Soil Total Nitrogen, Phosphorus and Potassium in Renshou County of Sichuan Basin, China." *Journal of Integrative Agriculture* 18(2): 279–89. https://doi.org/10.1016/S2095-3119(18)62069-6.

García-Palacios, P., F. T. Maestre, R. D. Bardgett, and H. de Kroon. 2012. "Plant Responses to Soil Heterogeneity and Global

Environmental Change." *The Journal of Ecology* 100(6): 1303–14. https://doi.org/10.1111/j.1365-2745.2012.02014.x.

Grunwald, S. 2009. "Multi-Criteria Characterization of Recent Digital Soil Mapping and Modeling Approaches." *Geoderma* 152: 195–207. https://doi.org/10.1016/j.geoderma.2009.06.003.

Guevara, M., C. Arroyo, N. Brunsell, C. O. Cruz, G. Domke, J. Equihua, J. Etchevers, et al. 2020. "Soil Organic Carbon across Mexico and the Conterminous United States (1991–2010)." *Global Biogeochemical Cycles* 34(3): e2019GB006219. https://doi.org/10.1029/2019GB006219.

Guevara, M., and R. Vargas. 2019. "Annual Soil Moisture Predictions Across Conterminous United States Using Remote Sensing and Terrain Analysis Across 1 km Grids (1991-2016)." HydroShare. https://doi.org/10.4211/hs.b8f6eae9d89241cf8b5904033460af61.

Hargrove, W. W., F. M. Hoffman, and B. E. Law. 2003. "New Analysis Reveals Representativeness of the AmeriFlux Network." *Eos* 84(48): 529–35. https://doi.org/10.1029/2003EO480001.

Hengl, T., J. M. de Jesus, R. A. MacMillan, N. H. Batjes, G. B. M. Heuvelink, E. Ribeiro, A. Samuel-Rosa, et al. 2014. "SoilGrids1km—Global Soil Information Based on Automated Mapping." *PLoS One* 9(8): e105992. https://doi.org/10.1371/journal.pone.0105992.

Hengl, T., J. M. de Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, et al. 2017. "SoilGrids250m: Global Gridded Soil Information Based on Machine Learning." *PLoS One* 12: e0169748. https://doi.org/10.1371/journal.pone.0169748.

Hengl, T., G. B. M. Heuvelink, and A. Stein. 2004. "A Generic Framework for Spatial Prediction of Soil Variables Based on Regression-Kriging." *Geoderma* 120: 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018.

Hengl, T., and N. Toomanian. 2006. "Maps Are Not What They Seem: Representing Uncertainty in Soil." In *International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* 805–13. Lisboa, Portugal: Instituto Geographico Portugues IGP.

Hengl, T., and R. A, MacMillan. 2019. *Predictive Soil Mapping with R. OpenGeoHub Foundation*. the Netherlands: Wageningen. ISBN: 978-0-359-30635-0.

Heuvelink, G. B. M. 2018. "Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling." In *Pedometrics*, edited by A. B. McBratney, B. Minasny, and U. Stockmann, 439–61. Progress in Soil Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-63439-5_14.

Hiemstra, P. H., E. J. Pebesma, C. J. W. Twenhofel, and G. B. M. Heuvelink. 2009. "Real-Time Automatic Interpolation of Ambient Gamma Dose Rates from the Dutch Radioactivity Monitoring Network." *Computers & Geosciences* 35(8): 1711–21. https://doi.org/10.1016/j.cageo.2008.10.011.

Hijmans, R. J., J. van Etter, J. Cheng, M. Mattiuzzi, M. Summer, J. A. Greenberg, O. P. Lamigueiro, et al. 2017. "Geographic Data Analysis and Modeling [R Package Raster Version 2.9-5]." R CRAN Project. Comprehensive R Archive Network (CRAN).

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *International Journal of Climatology* 25: 1965–78. https://doi.org/10.1002/joc.1276.

Horn, K. J., R. Quinn Thomas, C. M. Clark, L. H. Pardo, M. E. Fenn, G. B. Lawrence, S. S. Perakis, et al. 2018. "Growth and Survival Relationships of 71 Tree Species with Nitrogen and Sulfur Deposition across the Conterminous U.S." *PLoS One* 13(10): 1–19. https://doi.org/10.1371/journal.pone.0205296.

Jansen, M. J. W. 1998. "Prediction Error through Modelling Concepts and Uncertainty from Basic Data." In *Soil and Water Quality at Different Scales*, edited by P. A. Finke, J. Bouma, and M. R. Hoosbeek, 247–53. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-3021-1_23.

Keller, M., D. S. Schimel, W. W. Hargrove, and F. M. Hoffman. 2008. "A Continental Strategy for the National Ecological Observatory Network." *Frontiers in Ecology and the Environment* 6(5): 282–4. https://doi.org/10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2.

Kellndorfer, J., W. Walker, K. Kirsch, G. Fiske, J. Bishop, L. LaPoint, M. Hoppus, and J. Westfall. 2013. "NACP Aboveground Biomass and Carbon Baseline Data." V. 2 (Nbcd 2000). ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1161.

Kidd, D., M. Webb, B. Malone, B. Minasny, and A. McBratney. 2015. "Eighty-Metre Resolution 3D Soil-Attribute Maps for Tasmania, Australia." *Soil Research* 53(8): 932. https://doi.org/10.1071/SR14268.

Kimsey, M. J., L. E. Laing, S. M. Anderson, J. Bruggink, S. Campbell, D. Diamond, G. M. Domke, et al. 2020. "Soil Mapping, Monitoring, and Assessment." In *Forest and Rangeland Soils of the United States Under Changing Conditions*, edited by R. V. Pouyat, D. S. Page-Dumroese, T. Patel-Weynand, and L. H. Geiser, 169–88. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-45216-2_9.

Kokulan, V., O. Akinremi, A. P. Moulin, and D. Kumaragamage. 2018. "Importance of Terrain Attributes in Relation to the Spatial Distribution of Soil Properties at the Micro Scale: A Case Study." *Canadian Journal of Soil Science* 98(2): 292–305. https://doi.org/10.1139/cjss-2017-0128.

Kuhn, M. 2008. "Building Predictive Models in R Using the Caret Package." *Journal of Statistical Software* 28(5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Kuhn, M., and K. Johnson. 2013. *Applied Predictive Modeling*. New York: Springer. https://doi.org/10.1007/978-1-4614-6849-3.

Lamsal, S., C. M. Bliss, and D. A. Graetz. 2009. "Geospatial Mapping of Soil Nitrate-Nitrogen Distribution under a Mixed-Land Use System." *Pedosphere* 19: 434–45. https://doi.org/10.1016/S1002-0160(09)60136-3.

Laurance, W. F., P. M. Fearnside, S. G. Laurance, P. Delamonica, T. E. Lovejoy, J. M. Rankin-de Merona, J. Q. Chambers, and C. Gascon. 1999. "Relationship between Soils and Amazon Forest Biomass: A Landscape-Scale Study." *Forest Ecology and Management* 118(1): 127–38. https://doi.org/10.1016/S0378-1127(98)00494-0.

LeBauer, D. S., and K. K. Treseder. 2008. "Nitrogen Limitation of Net Primary Productivity in Terrestrial Ecosystems Is Globally Distributed." *Ecology* 89(2): 371–9.

Littke, K. M., R. B. Harrison, D. Zabowski, D. G. Briggs, and D. A. Maguire. 2014. "Effects of Geoclimatic Factors on Soil Water, Nitrogen, and Foliar Properties of Douglas-Fir Plantations in the Pacific Northwest." *Forest Science* 60(6): 1118–30.

Malone, B. P., A. B. McBratney, and B. Minasny. 2011. "Empirical Estimates of Uncertainty for Mapping Continuous Depth Functions of Soil Attributes." *Geoderma* 160: 614–26. https://doi.org/10.1016/j.geoderma.2010.11.013.

Malone, B. P., A. B. McBratney, B. Minasny, and G. M. Laslett. 2009. "Mapping Continuous Depth Functions of Soil Carbon Storage and Available Water Capacity." *Geoderma* 154: 138–52. https://doi.org/10.1016/j.geoderma.2009.10.007.

Malone, B. P., B. Minansy, and C. Brungard. 2019. "Some Methods to Improve the Utility of Conditioned Latin Hypercube Sampling." *PeerJ* 7: e6451. https://doi.org/10.7717/peerj.6451.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21(2): 239–45. https://doi.org/10.2307/1268522.

Meinshausen, N. 2006. "Quantile Regression Forests." *Journal of Machine Learning Research* 7: 983–99.

Minasny, B., A. B. McBratney, M. L. Mendonça-Santos, I. O. A. Odeh, and B. Guyon. 2006. "Prediction and Digital Mapping of Soil Carbon Storage in the Lower Namoi Valley." *Australian Journal of Soil Research* 44: 233–44. https://doi.org/10.1071/SR05136.

Morellos, A., X. E. Pantazi, D. Moshou, T. Alexandridis, R. Whetton, G. Tziotzios, J. Wiebensohn, R. Bill, and A. M. Mouazen. 2016. "Machine Learning Based Prediction of Soil Total Nitrogen, Organic Carbon and Moisture Content by Using VIS-NIR Spectroscopy." *Biosystems Engineering* 152: 104–16. https://doi.org/10.1016/j.biosystemseng.2016.04.018.

Mou, P., R. H. Jones, R. J. Mitchell, and B. Zutter. 1995. "Spatial Distribution of Roots in Sweetgum and Loblolly Pine Monocultures and Relations with Above-Ground Biomass and Soil Nutrients." *Functional Ecology* 9(4): 689–99. https://doi.org/10.2307/2390162.

Mu, Q., M. Zhao, S. W. Running. 2011. "Improvements to a MODIS Global Terrestrial Rvapotranspiration Algorithm." *Remote Sensing of Environment* 115(8): 1781-1800. https://doi.org/10.1016/j.rse.2011.02.019.

National Academies of Sciences, Engineering, and Medicine. 2021. *Exploring a Dynamic Soil Information System: Proceedings of a Workshop*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26170.

Nave, L., K. Johnson, C. van Ingen, D. Agarwal, M. Humphrey, and N. Beekwilder. 2016. "International Soil Carbon Network (ISCN) Database v3-1." https://doi.org/10.17040/ISCN/1305039.

Paoli, G. D., L. M. Curran, and J. W. F. Slik. 2008. "Soil Nutrients Affect Spatial Patterns of Aboveground Biomass and Emergent Tree Density in Southwestern Borneo." *Oecologia* 155(2): 287–99. https://doi.org/10.1007/s00442-007-0906-9.

Peltoniemi, M. S., R. A. Duursma, and B. E. Medlyn. 2012. "Co-Optimal Distribution of Leaf Nitrogen and Hydraulic Conductance in Plant Canopies." *Tree Physiology* 32(5): 510–9. https://doi.org/10.1093/treephys/tps023.

Poggio, L., and A. Gimona. 2014. "National Scale 3D Modelling of Soil Organic Carbon Stocks with Uncertainty Propagation—An Example from Scotland." *Geoderma* 232–234: 284–99. https://doi.org/10.1016/j.geoderma.2014.05.004.

Priyadharshini, R., M. Radha, R. Kumaraperumal, G. Vanitha, and B. Kannan. 2021. "Characterization of Environmental Covariates of Coimbatore District Using Principal Component Analysis." *International Journal of Current Microbiology and Applied Sciences* 10(01): 3114–23. https://doi.org/10.20546/ijcmas.2021.1001.362.

Ramcharan, A., T. Hengl, T. Nauman, C. Brungard, S. Waltman, S. Wills, and J. Thompson. 2018. "Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution." *Soil Science Society of America Journal* 82(1): 186–201. https://doi.org/10.2136/sssaj2017.04.0122.

Robertson, G. P., M. A. Hutson, F. C. Evans, and J. M. Tiedje. 1988. "Spatial Variability in a Successional Plant Community: Patterns of Nitrogen Availability." *Ecology* 69(5): 1517–24. https://doi.org/10.2307/1941649.

Roudier, P. 2011. "Clhs: A R Package for Conditioned Latin Hypercube Sampling."

RStudio Team. 2021. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC. http://www.rstudio.com/.

Running, S. W., R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto. 2004. "A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production." *Bioscience* 54(6): 547. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2.

Sanchez, P. A., S. Ahamed, F. Carré, A. E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, et al. 2009. "Digital Soil Map of the World." *Science* 325: 680–1. https://doi.org/10.1126/science.1175084.

Scull, P., J. Franklin, O. A. Chadwick, and D. McArthur. 2003. "Predictive Soil Mapping: A Review." *Progress in Physical Geography* 27: 171–97. https://doi.org/10.1191/0309133303pp366ra.

Shrestha, D. L., and D. P. Solomatine. 2006. "Machine Learning Approaches for Estimation of Prediction Interval for the Model Output." *Neural Networks: The Official Journal of the International Neural Network Society* 19(2): 225–35. https://doi.org/10.1016/j.neunet.2006.01.012.

Smith, E., R. Vargas, M. Guevara, T. Tarin, and R. Pouyat. 2022. "Dataset: Spatial Variability and Uncertainty of Soil Nitrogen across the Conterminous United States at Different Depths." Figshare. Dataset. https://doi.org/10.6084/m9.figshare.19397156.v1.

Soil Survey Staff, and T. Loecke. 2016. In *Rapid Carbon Assessment: Methodology, Sampling, and Summary*, edited by S. Willis. U.S. Department of Agriculture, Natural Resources Conservation Service. https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052841.pdf

Stell, E., D. Warner, J. Jian, B. Bond-Lamberty, and R. Vargas. 2021. "Spatial Biases of Information Influence Global Estimates of Soil Respiration: How Can We Improve Global Predictions?" *Global Change Biology* 27: 3923–38. https://doi.org/10.1111/gcb.15666.

Sun, Z., X. Wang, H. Yamamoto, H. Tani, G. Zhong, S. Yin, and E. Guo. 2018. "Spatial Pattern of GPP Variations in Terrestrial Ecosystems and Its Drivers: Climatic Factors, CO_2 Concentration and Land-Cover Change, 1982–2015." *Ecological Informatics* 46: 156–65. https://doi.org/10.1016/j.ecoinf.2018.06.006.

Tian, X., F. Minunno, P. Schiestl-Aalto, J. Chi, P. Zhao, M. Peichl, J. Marshall, et al. 2021. "Disaggregating the Effects of Nitrogen Addition on Gross Primary Production in a Boreal Scots Pine Forest." *Agricultural and Forest Meteorology* 301–302: 108337. https://doi.org/10.1016/j.agrformet.2021.108337.

Vargas, R., D. D. Baldocchi, M. F. Allen, M. Bahn, T. A. Black, S. L. Collins, J. C. Yuste, et al. 2010. "Looking Deeper into the Soil: Biophysical Controls and Seasonal Lags of Soil CO_2 Production and Efflux." *Ecological Applications* 20(6): 1569–82.

Villarreal, S., M. Guevara, D. Alcaraz-Segura, N. A. Brunsell, D. Hayes, H. W. Loescher, and R. Vargas. 2018. "Ecosystem Functional Diversity and the Representativeness of Environmental Networks across the Conterminous United States." *Agricultural and Forest Meteorology* 262: 423–33. https://doi.org/10.1016/j.agrformet.2018.07.016.

Villarreal, S., M. Guevara, D. Alcaraz-Segura, and R. Vargas. 2019. "Optimizing an Environmental Observatory Network Design Using Publicly Available Data." *Journal of Geophysical Research: Biogeosciences* 124(7): 1812–26. https://doi.org/10.1029/2018JG004714.

Villarreal, S., and R. Vargas. 2021. "Representativeness of FLUXNET Sites across Latin America." *Journal of Geophysical Research: Biogeosciences*, 126: e2020JG006090. https://doi.org/10.1029/2020JG006090.

Wadoux, A. M. J.-C., and A. B. McBratney. 2021. "Digital Soil Science and Beyond." *Soil Science Society of America Journal* 85: 1313–31. https://doi.org/10.1002/saj2.20296.

Wang, S., Q. Zhuang, Q. Wang, X. Jin, and C. Han. 2017. "Mapping Stocks of Soil Organic Carbon and Soil Total Nitrogen in Liaoning Province of China." *Geoderma* 305(120): 250–63. https://doi.org/10.1016/j.geoderma.2017.05.048.

Warren, G. P., and D. C. Whitehead. 1988. "Available Soil Nitrogen in Relation to Fractions of Soil Nitrogen and Other Soil Properties." *Plant and Soil* 112(2): 155–65. https://doi.org/10.1007/BF02139991.

Yang, F., A. X. Zhu, K. Ichii, M. A. White, H. Hashimoto, and R. R. Nemani. 2008. "Assessing the Representativeness of the AmeriFlux Network Using MODIS and GOES Data." *Journal of Geophysical Research* 113: G04036. https://doi.org/10.1029/2007JG000627.

Yuxin, M., B. Minasny, A. McBratney, L. Poggio, and M. Fajardo. 2021. "Predicting Soil Properties in 3D: Should Depth Be a Covariate?" *Geoderma* 383: 114794. https://doi.org/10.1016/j.geoderma.2020.114794.

Zeraatpisheh, M., S. Ayoubi, A. Jafari, and P. Finke. 2017. "Comparing the Efficiency of Digital and Conventional Soil Mapping to Predict Soil Types in a Semi-Arid Region in Iran." *Geomorphology* 285: 186–204. https://doi.org/10.1016/J.GEOMORPH.2017.02.015.

Zhou, Y., J. Xue, S. Chen, Y. Zhou, Z. Liang, N. Wang, and Z. Shi. 2020. "Fine-Resolution Mapping of Soil Total Nitrogen across China Based on Weighted Model Averaging." *Remote Sensing* 12(1): 1–18. https://doi.org/10.3390/RS12010085.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Smith, Elizabeth M., Rodrigo Vargas, Mario Guevara, Tonantzin Tarin, and Richard V. Pouyat. 2022. "Spatial Variability and Uncertainty of Soil Nitrogen across the Conterminous United States at Different Depths." *Ecosphere* 13(7): e4170. https://doi.org/10.1002/ecs2.4170