Significant Scope for Improvement of Soil Acidity Management in US Croplands

Beck Woollen[1,2*], Cella Schnabel[3,4], Christopher Reinhard[5], Shuang Zhang[6], Tim Jesper Suhrhoff[1,7], Anu Khan[2], Ella Milliken[1,7], Brian Buma[4], Noah Planavsky[1,2,4,7*]

1: Yale Center for Natural Carbon Capture, Yale University; 2: Carbon Removal Standards Initiative; 3: Department of Civil and Environmental Engineering, Cornell University; 4: Environmental Defense Fund; 5: School of Earth & Atmospheric Sciences, Georgia Institute of Technology; 6: Department of Oceanography, Texas A&M University; 7: Department of Earth & Planetary Sciences, Yale University

Corresponding authors: Beck Woollen (beck.woollen@yale.edu); Noah Planavsky (noah.planavsky@yale.edu)

Abstract

Robust soil pH management in agricultural lands is critically important for crop yields and can lead to significant greenhouse gas reductions in agricultural systems. However, costs and logistical constraints can limit the extent to which agricultural soil pH is actively managed. We develop two new estimates of soil pH across the conterminous United States using empirical Bayesian kriging and random forest approaches. We compare these estimated pH values against a new dataset of the agronomically recommended pH at a county level, defined by state-specific soil pH recommendations for the most prominent crop in each county. We estimate between 44–55% of cropland acres in the conterminous United States are below agronomically optimal soil pH. We couple these new estimates with high resolution pH data from six agricultural sites. We find large variability in soil acidity at the field scale, indicating that regional mapping efforts likely underestimate the need for soil deacidification.

1. Introduction

Nutrient availability and crop health are both negatively impacted by acidic soils, which have become increasingly common due to extensive fertilizer application over the last century. Between 1940 and 2015, nitrogen fertilizer consumption in the conterminous United States (CONUS) has increased 33-fold to approximately 11 million metric tons per year (Cao et al., 2018). Consequently, American staple crops have seen steady improvements in yield per acre (Schnitkey et al., 2023; Nielsen, 2020). However, the repeated application of chemical fertilizers has also acidified croplands, negatively affecting yields in multiple regions (e.g., Pennisi, 2023; Ghimire et al., 2017; Guo et al., 2010; Barak et al., 1997). At acidic pH values, multiple macro- and micronutrients become less bio-available to crops, although these dynamics depend on many variables (Hartemink & Barrow, 2023). Enhanced weathering (EW) of agricultural limestone (aglime) is a common approach to mitigate soil acidification (West & McBride, 2005) and thereby improve crop yields (Sirisuntornlak et al., 2021; Ghimire et al., 2017; De Oliveira & Pavan, 1996). There has also been a recent push to use silicate minerals for soil pH management. This process has been observed to drive relatively similar yield increases as standard liming practices (e.g., Beerling et al., 2024).

Using limestone or silicates as soil amendments can drive durable carbon dioxide removal (Kukla et al., 2025). This atmospheric carbon removal can occur years or even decades after rock application (e.g., Kanzaki et al., 2025). In the case of liming, the process is commonly carbon positive initially, before transitioning to carbon negativity (see Raymond et al., 2025 for a discussion). Nonetheless, a recent analysis of the cumulative liming history of the Mississippi River catchment suggested that liming has led to highly efficient carbon removal (i.e., over 85% of the ideal alkalinity flux expected from liming over the past 100 years has been exported from the mouth of the Mississippi River) (Suhrhoff et al., 2026). Further, nitrous oxide, a greenhouse gas roughly 273 times more potent than carbon dioxide over a 100-year time frame, is emitted at higher rates from moderately acidic soils (Intergovernmental Panel on Climate Change, 2023; Chiaravalloti et al., 2023).

Here we present a new analysis of cropland areas across CONUS which, from an agronomic point of view, have below-optimal soil pH. We compile county-by-county agronomic soil pH recommendations for the spatially dominant crop from extension offices. We then map cropland soil pH using empirical Bayesian kriging (EBK) and random forest (RF) methods. This approach matches current pH with targets appropriate for the specific crops grown, contrasting with many studies that use a constant value for optimal soil pH across different crops. By comparing the predicted pH map to spatially explicit soil pH recommendations, we can clearly identify regions of CONUS with high potentials for increased EW deployment. Lastly, we present new high resolution soil pH measurements from six fields across several major agricultural regions in CONUS. We use these new data to highlight that regional scale approaches of estimating area below optimal pH conditions will be underestimates, given the large scale (order of magnitude) variability in pH within single fields (over tens of hectares). Overall, our analysis indicates that, despite a long history of liming in CONUS, there is need for new soil pH management practices and programs.

2. Methods

2.1 Compiling Agronomic pH Recommendations

Cropland composition at the county scale is from the 2022 CroplandCROS dataset (USDA, 2023), which contains 30x30m raster data of crops grown across CONUS. We used a generic counties shapefile to generate cell counts of the crops grown in each of the 3,099 CONUS counties. From these counts, we selected the crop in each county that occupied the most land area. We excluded the "Grassland/Pasture" value from CroplandCROS in raster cell counts, because the breadth of this category leaves considerably more room for ambiguity in pH management practices than harvested crops.

In each state, at least one land-grant university hosts an agronomic extension service that recommends target pH values for crops grown within the state. We compiled soil pH recommendations at the county level, based on the most grown crop identified by cell count in each county. Guidelines used to obtain pH recommendations are detailed in Supplemental Information, Appendix V. Often, target pH will be provided as a range (e.g., University of Georgia recommends a pH of 5.5-6.5 for corn) (Westerfield, 2019). When this was the case, we recorded two different values: an unweighted average of the range (e.g., 6) and an upper value of the range (e.g., 6.5). In cases where a range was not provided, we recorded the stated number for both the average and upper pH recommendation (Figure 1).

2.2 Soil pH Data Collection

We used soil pH data from the National Cooperative Soil Survey (NCSS) to build the EBK and RF maps of soil pH across CONUS (National Cooperative Soil Survey, 2023). The NCSS dataset contains point measurements of soil physical and chemical properties. The main_lab_layer, lab_combine_nasis, and lab_chemical_properties tables were joined by shared attributes to get coordinates, sample depth, observation data, and chemical properties in the same input dataset. To avoid duplicate entries from the same location, duplicate instances of pedon_key were removed. At a given set of coordinates, NCSS data often includes measurements from various depths (e.g., 0-15cm, 15-30cm, and 30-45cm). Recognizing that the most immediate changes to pH from EW happen in the uppermost portion of the soil (Kantzas et al., 2022; Brown et al., 1956), we filtered the NCSS data to include entries where the horizon top (hzn_top) was equal to zero, or the layer sequence (layer_sequence) was equal to 1. The NCSS dataset contains six methods of pH measurements, but data is primarily available for the pH_h2o and pH_cacl2 variables. The pH_h2o variable is defined as a 1:1 soil to water approach, whereas pH_cacl2 is defined as a 1:2 soil to cacl2 ratio. Additional methodological information is available in the database for methodological approaches for specific points (e.g., 4C1a2a vs. 8C1e methods). We used pH_h2o for our analysis, which had a higher n than pH_cacl2 after this filtering was complete (46,417 points vs. 38,709 points). The difference between mean pH_h2o and pH_cacl2 in these data is 0.41 (6.1 – 5.69) (National Cooperative Soil Survey, 2023).

The NCSS dataset contains entries from as far back as the early 20th Century, which we determined are unlikely to represent soil conditions and land use practices on cropland today (e.g., measurements taken before liming and industrial fertilizers were commonplace are unlikely to reflect the influence of these soil amending processes). Measurements predating 1980 were removed. We selected 1980 to balance the advantages of a larger dataset with decreasing likelihood that points represent current conditions as time since measurement increases. Additionally, 1980 represents a practical cutoff point in that commercial fertilizer use by primary nutrient use has been relatively consistent since 1980, compared to a general rise beforehand (Mosheim, 2025). This filtering provided a final input dataset of n = 29,300 soil pH measurements within CONUS, which were used to develop the Empirical Bayesian Kriging (EBK) and Random Forest (RF) models discussed below.

2.3 Empirical Bayesian Kriging

Empirical Bayesian Kriging (EBK) interpolation was conducted via ArcGIS Pro (Figure 2). The EBK was developed using only points on cropland, which we determined using the 2021 National Land Cover Database (Dewitz, 2023). The NLCD contains categories 81 (Pasture/Hay) and 82 (Cultivated Crops). A reclassified NLCD (binary, 81 or 82 = 1, else = 0) raster was used to select points on cropland, yielding an NCSS data subset of 10,722 points. We randomly selected 20% (n

= 2,144) of the data points to be testing data, leaving 80% (n = 8,578) to build the EBK (Figure 2a). The EBK was interpolated at 10 km resolution, in line with the RF model explained below.

2.4 Random Forest

We developed the random forest (RF) model using the full filtered NCSS dataset (n = 29,300), randomly selecting 20% of the data for testing and using the remaining 80% (n = 23,440) to train the model (Figure 3). The RF model uses the NLCD to assign a probability of cropland to each point in the NCSS dataset, which eliminated the need to filter non-cropland points as was done during EBK development. We used 31 predictor variables (Supplemental Information Appendix I). We developed the RF model in Python using the scikit-learn library, optimizing hyperparameters through a grid search. All raster-based covariates that were fed into the RF were re-gridded to a 10 km resolution prior to its development to align with the EBK method.

2.5 Calculating Area Under Recommended pH Levels

The EBK and RF rasters were resampled using nearest neighbor to a 30m resolution, to align with the reclassified 2021 NLCD raster discussed in 2.1 (Figure 2b, 3b). We calculated both the count and percentage of 30m EBK or RF cropland cells below extension-recommended pH. To conduct these counts, cells were reclassified into bins of pH 0.1 and counted using a zonal histogram approach. Initial calculations focused on a precision of pH 0.1, identifying all raster cells at all below recommended pH. This process was repeated to only count cells at least 0.25, 0.5, 0.75, and 1 pH below recommended pH. Distributions of pH for both EBK and RF estimates were recorded, along with the mean and standard deviation difference between recommended and estimated pH. These zonal statistics were also calculated at state and USDA agronomic regional levels.

2.6 Calculating Extent Under Recommended pH Levels

In addition to the descriptive statistics described in section 2.5, we calculated a weighted sum of cropland cells in each county. This weighted sum accounts for both the amount of cropland under recommended pH and the degree to which individual cells are below those values. The statistics discussed in 2.5 provided us with counts of 30x30m cells in each CONUS county at a 0.1 pH precision. By multiplying each of these cell counts by the difference between its pH and recommended pH, we were able to assign greater weight to cells further below recommended pH. In calculating these weighted sums, we only considered the cells in counties for which estimated pH was less than recommended pH, as equal or above-recommended cells may cancel out negative cells within counties. Provided that a negative estimated pH – recommended pH differences indicate potential to raise pH, more negative scores in Figure 7 indicate higher potentials.

$$(1) \quad W_c = \sum_{i:pH_i - \overline{pH_c} < 0} n_{c,i}\left(pH_i - \overline{pH_c}\right)$$

Where:

W_c = Weighted score of EBK or RF cells under recommended pH in county c

c = CONUS county c
$1 \le c \le 3{,}099$
i = index for each pH value at which EBK or RF cells estimate pH
pH_i = the 0.1 resolution pH value at which EBK or RF cells are binned from zonal histogram
$\overline{pH_c}$ = extension-recommended pH value (average) in county c
$n_{c,i}$ = the number of RF or EBK cells in county c, at pH value i

2.7 Evaluating EBK and RF Agreement

Agreement between the EBK and RF models was first evaluated by subtracting the 30x30m rasters from each other. To test whether the EBK and RF models agreed soil pH was overly acidic in each county, we used the statistics calculated in 2.5 to construct a mean $\pm 2\sigma$ range of the mapped minus recommended pH. In cases where the mean EBK $\pm 2\sigma$ range and mean RF $\pm 2\sigma$ range were both below 0, it was concluded both models agreed mean pH was below recommended pH. In cases where both estimates overlapped with 0, it was concluded no discernible difference existed between mapped and recommended pH. In counties where both estimates $\pm 2\sigma$ were greater than 0, we concluded mean estimated pH was above recommended levels. In all other cases, it was concluded the EBK and RF disagreed as to whether mean county pH minus recommended pH overlapped with 0 (Figure 11).

2.8 Error Estimation

We estimated error with the omitted 20% of the NCSS training data for both the EBK and RF models (2,144 and 5,866 points, respectively). Error was not calculated from the original interpolations, but rather their discretized versions with 0.1 pH classes as discussed in 2.5. Although this calculation method does not perfectly reflect the accuracy of the original interpolations because of rounding, it does reflect the accuracy of their modified versions used to calculate area and percentage below extension-recommended (Figure 4a-d).

2.9 EBK and RF Output Comparison to Empirical Data

We compared model output to empirical pH measurements from 6 agricultural fields spanning the Midwest and Southeast, to evaluate differences in model performance at much finer scales than the CONUS level primarily explored in this analysis. For each field, the empirical data was interpolated over the individual parcel using an EBK interpolation in ArcGIS Pro (Figure 12; see Table SI 1 for field-specific data).

3. Results

3.1 Cropland Below Recommended pH Thresholds

The EBK interpolation estimates 81,229,711 hectares of cropland (44.23%) are below the county-specific average pH recommendations of the dominant crop in each CONUS county (Figure 5a-b). When using the upper pH recommendations in each county, 98,315,482 hectares of cropland

(53.53%) are below optimal levels (Figure 5c-d). The RF pH model suggests 84,895,235 hectares of cropland (46.22%) are below average pH recommendations by a pH ≥ 0.1, while 101,769,901 hectares (55.41%) are below upper pH recommendations by the same threshold (Figure 6a-d). When adjusting to a difference of 0.25, 0.5, 0.75, and ≥ 1, the percentages and areas of cropland for both pH estimates decline relatively similarly, as shown in Figure 9a-b.

The EBK and RF each demonstrate a strong directional trend in overly acidic soils. In both cases, counties with high percentages of cropland below recommended pH levels are highly concentrated in the eastern half of CONUS. This pattern is apparent over large agronomic regions (Table SI, 1). Demonstrating this pattern, the EBK and RF respectively suggest 93.08% and 98.86% of cropland to be below average recommended pH in the Appalachia USDA agronomic region (counties in Kentucky, North Carolina, Tennessee, and Virginia). Conversely, the EBK and RF find 3.75% and 2.99% below average pH recommendation in the Mountain USDA agronomic region (Arizona, Colorado, Idaho, Montana, New Mexico, Nevada, Utah, and Wyoming). Data for each USDA agronomic region is available in Supplemental Information, Appendix IV.

3.2 County Hotspot Analysis

Using equation 1, we identified the 50 counties indicated by EBK and RF to have the highest opportunity for pH improvement. These counties are primarily concentrated in the eastern half of CONUS, with particularly distinct hotspots in eastern Arkansas, eastern Texas, and central Florida (Figure 8). The EBK and RF top 50 counties with the highest opportunity to raise soil pH share 34 counties in common. A table of the top 50 EBK and RF counties can be found in Supplemental Information, Appendix 6. Evaluations of all counties (Figure 7a-b) indicates opportunities to raise soil pH are highly concentrated in the eastern half of CONUS.

3.3 Agreement Between EBK and RF

We evaluated agreement between the EBK and RF pH maps (Figure 10). Here, we find the two rasters to be within ±0.4 pH of agreement in 75.8% of 30x30m raster cells. The rasters have an agreement of ±0.1 pH 32.4% of the time. As shown in Figure 8, pronounced differences between the two pH estimates exist in hotspots across CONUS.

To understand the impact of this disagreement on our findings, we calculated the mean difference between estimated and recommended pH and constructed 2σ upper and lower bounds for that mean difference (mean and standard deviations for EBK and RF minus recommended pH shown in Figure SI 6). Comparing these values allowed us to test whether the ±2σ range included zero. In the 1,212 counties where both the RF and EBK ranges were below zero, we concluded with 95% confidence that the two methods agreed mean county cropland pH was below recommended levels. The EBK and RF agreed mean county pH was above recommended levels in 583 counties, agreed there was an insignificant difference between mean county pH and recommended pH (95% confidence interval includes 0) in 567 counties, and disagreed 737 times (Figure 11).

3.4 Field Scale pH Variability

To offer a different perspective on the CONUS-scale analysis above, we also provide new pH maps from six densely sampled field measurements (*n* = 451) from growers enrolled in a pH management program. The 10 km EBK pH estimate fell within the 5–95% pH value distribution in four of the six fields (Figure 12). In all the fields we densely sampled (ranging from 4.37 to 32.41 hectares) we found over an order a magnitude variability in the extent of soil acidity (i.e., more than a pH unit of variability at the field scale).

4. Discussion

We developed two spatially explicit estimates of soil pH across CONUS, each of which was compared against the agronomic extension recommended pH for the most grown crop in every CONUS county. Our finding that approximately half of cropland acres in CONUS are below recommended pH is substantial but aligned with previous research (West & McBride 2005, Shorey 1940). Reaching this conclusion with two separate interpolations strongly indicates an opportunity to raise agricultural soil pH and gain agronomic benefits as well as additional co-benefits (e.g., CO_2 removal via EW). As expected, we observe the highest concentrations of cropland below extension recommended pH levels in the eastern half of CONUS. The RF and EBK pH estimate also suggest several clear hotspots to prioritize pH management, including in Arkansas, Texas, and Florida. Although aglime has been predominantly applied in the eastern half of CONUS, our results suggest past application rates may be insufficient to counterbalance cropland acidification. Each pH estimate also identifies California, Oregon, and Washington as having discernible amounts of cropland below recommended levels (Figures 5-7), indicating potential for increased pH management in the Pacific states.

We find the majority of overly acidic croplands to be within 1 pH unit of optimal levels, indicating substantial potential to improve crop yields through increased EW deployment (Beerling et al., 2024; Goulding, 2016; Pagani & Mallarino, 2012; Gillman et al., 2002). Corn, soybeans, alfalfa, and cotton are some of the most grown crops across CONUS (USDA, 2023); each of which has been shown to have substantial yield improvements (e.g., Sharry et al., 2025; Pagani & Mallarino, 2015; Undersander, 2001). In this respect, our results indicate the issue of overly acidic croplands is both severe and solvable. Moving forward, the statistics presented in this research can be used to diagnose local or regional opportunities to deacidify croplands through EW.

One important consideration for interpreting our two models is their level of error. In both cases, the EBK and RF have errors centered around zero, meaning that neither model systematically over- nor underestimates pH at the CONUS scale (EBK $\bar{x}$ -0.04, σ 0.75; RF $\bar{x}$ 0.003, σ = 0.802) (Figure 4a-d). Given the error is relatively similar for the EBK and RF models, we hypothesize that interpolation method may not be the primary driver of error. Likely influential factors include the relatively low density of input points and the relatively large resolution at which we interpolated (10 km), which was chosen for compatibility with the data products used to generate our RF model. Foremost, no estimates of regional or farmer liming history go into the RF framework. We predict that individual farmer choices on pH management are the largest driver of offsets between predicted and observed pH values. Consistent with this idea, the fields in which we conducted detailed pH mapping were farmers were enrolled in a USDA sponsored pH management program (part of the Climate Smart Commodities program), in general had higher average pH values than regional estimates. These represented growers that have been actively managing their pH and showed strong interest in pH management practices.

Additionally, the timespan of pH measurement may have decreased the accuracy of our RF model in cases where chronological misalignments exist between the date pH measurements were recorded and the dates when other covariate inputs were recorded. Finally, our selection of pH_h2o, rather than pH_cacl2 from the NCSS dataset likely influences our findings. However, provided that the mean of pH_h2o measurements is more alkaline than pH_cacl2, this decision likely decreased the extent to which we estimate croplands to be overly acidic. Errors do change how our findings should be interpreted. Percentages and counts of hectares under recommended pH levels should be taken as general indicators for a given state or region, rather than precise evaluations at a more granular level. The 8,578 training points used for the EBK model spread across 3,099 counties produces an average of 2.77 points per county; the RF average is higher, at 7.56 points per county. We suggest that neither of these values approaches the sufficient sampling density to produce accurate pH diagnostics at the county level or finer.

Beyond error, there are also clear limitations in capturing localized heterogeneity of soil pH at the chosen resolution of 10 km. As previously discussed, soil pH can be quite variable, as it can be spatiotemporally influenced by most every biogeochemical lever within the agronomic system. As presented in Figure 12a-b, there is very high heterogeneity on a hectare-to-hectare basis, with all 6 fields presenting a pH fluctuation greater than 1 unit. Notably, the smallest field in the set (4.37 acres) contains the greatest range in pH (Figure 12b, Table SI 1). Additionally, there appears to be very little clear effect of reducing field-wide heterogeneity by increasing the number of cores per sample, suggesting that local variability, rather than signal-to-noise, is the dominant driver in the value of a given measurement. This is evident in the high overlap in values when comparing field variability in Figure 12b, wherein outliers are almost always overlapping with mean and confidence interval bands of the other fields, except in the most extreme cases. Nonetheless, there are clear regional trends in soil pH that are likely to emerge from the large-scale dataset, as seen in the clear mean decline in pH from the Midwest to East Coast (Figure 12b). This points to our models, and likely many other similar-resolution pH estimates, as best suited for generalized, large-scale analyses. Granular local measurements, compared to these regional trends, however, will likely be useful to inform how local land management practices shape local pH values. We recommend future comparisons of field-scale measurements with larger interpolations standardize pH measurement method where possible.

5. Conclusion

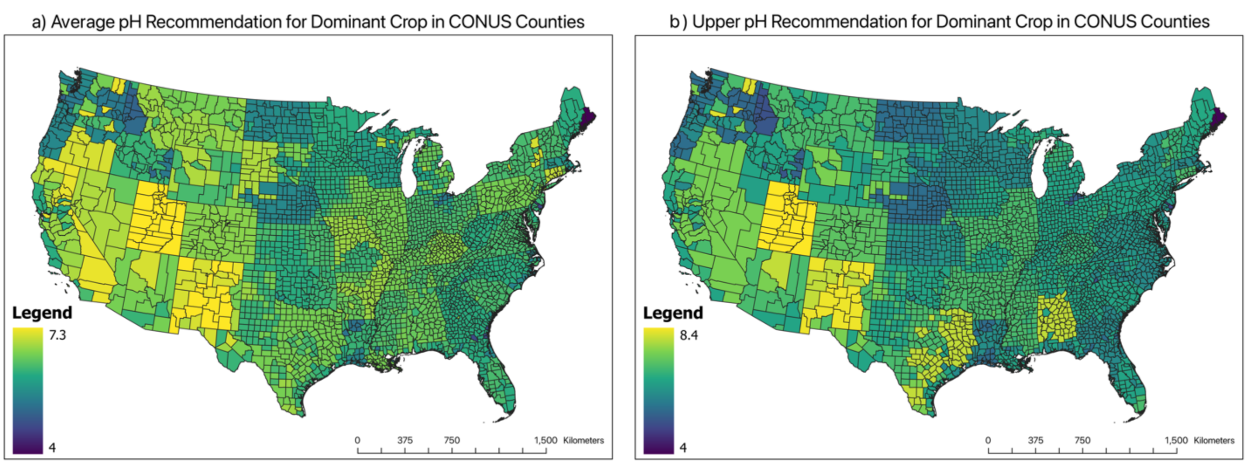
In this study, we compiled soil pH recommendations for the most grown crop in every county across CONUS. This exercise produced a heterogenous set of recommendations, that vary by crop, state, and soil type. We then estimated the pH of croplands across CONUS, using both EBK and RF approaches. Although the EBK and RF are prone to differences at a county-by-county level, their instances of agreement outnumber disagreements and identify specific geographic opportunities to raise cropland soil pH. The two pH estimates are well aligned at larger spatial scales and reinforce previous research on concentrations of acidic soils in the eastern half of CONUS. At the national scale, the EBK and RF are in strong agreement on the amount of cropland below optimal levels—differing just 2% from each other. Our findings corroborate previous research that swaths of agricultural land are too acidic, directly decreasing crop yields. This analysis highlights the potential to deliver economic and environmental benefits through the deployment of EW and additional aglime across CONUS.

6. Author Contributions and Acknowledgements

CRediT statement: NJP, CTR and BW conceptualized the study. BW, CS, TJS, SZ, and EM were involved in data curation. BW, CS, and EM completed formal analysis and investigation for the study. BW, CS, NJP, and BB contributed to methodology development. NJP, BB, and AK were involved in project administration and supervision. BW drafted initial text. All authors were involved in review and editing.

The authors thank Noemma Olagaray, Sophie Spiegel, and Marya Matlin-Wainer for their contributions to agronomic extension data collection in the early stages of the project. Dr. Lucia Dora Simonelli offered valuable help with visualizations. NJP, BW, and TJS acknowledge funding from the YCNCC. NJP acknowledges funding from the Grantham Foundation and Environmental Defense Fund. TJS acknowledges funding from the Swiss National Science Foundation (Grant P500PN_210790). CTR acknowledges funding from the Grantham Foundation. NJP and CTR were co-founders of the CDR company Lithos Carbon but have no remaining financial ties to the company.

7. Figures



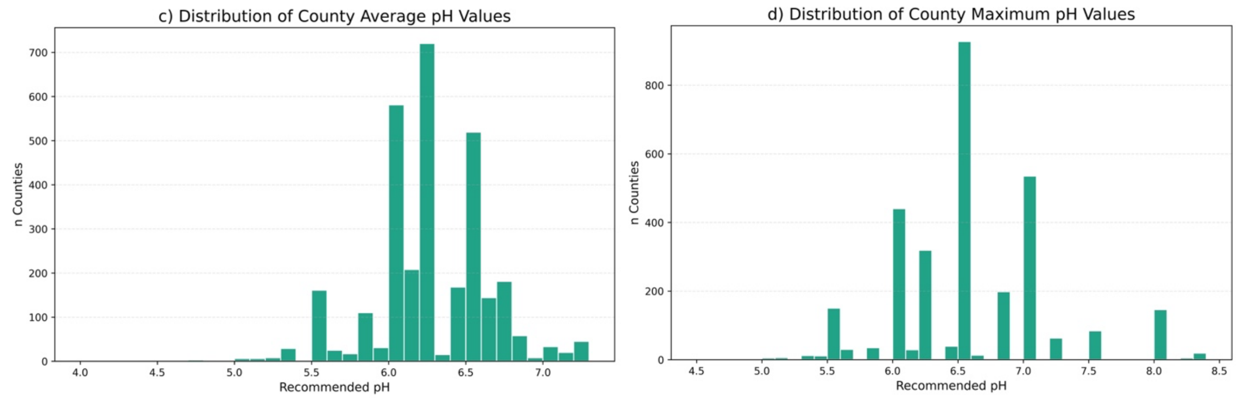


Figure 1(a): Average pH recommendation for the most spatially abundant crop in each CONUS county (e.g., if corn, with a range 5.5-6.5, value = 6); *(b)*: Upper pH recommendation for the most spatially abundant crop in each CONUS county (e.g., if corn, with a range of 5.5-6.5, value = 6.5); *(c)*: Histogram of county-specific average pH recommendations; *(d)*: Histogram of county-specific upper pH recommendations.

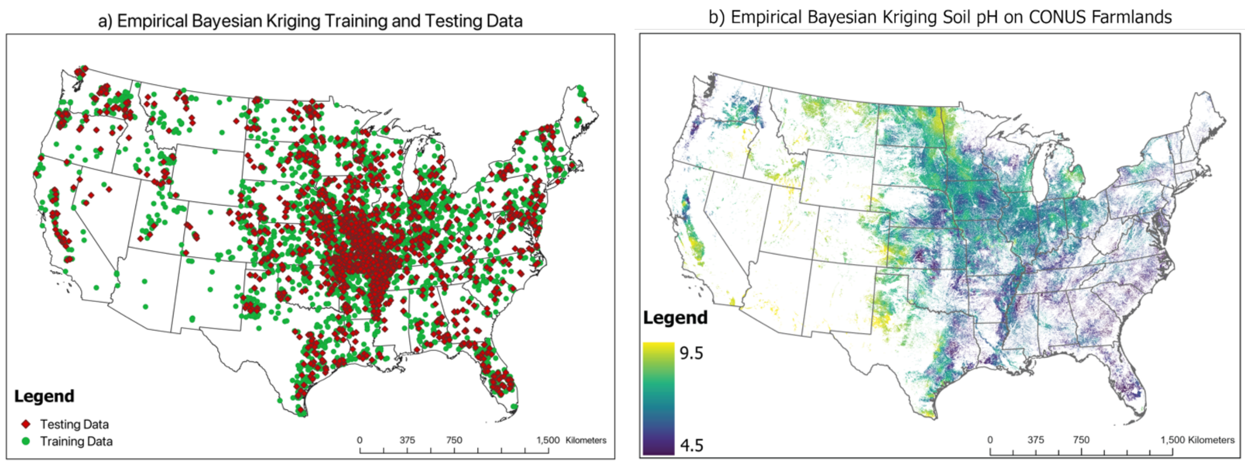


Figure 2(a): EBK NCSS training and testing points (80% and 20%, respectively) that are filtered to be on cropland, from 1980 or later, and from the uppermost portion of the sampled site; *(b)*: EBK interpolation of pH with 10 km resolution, resampled and clipped to cropland from the 2021 NLCD (30x30m resolution).

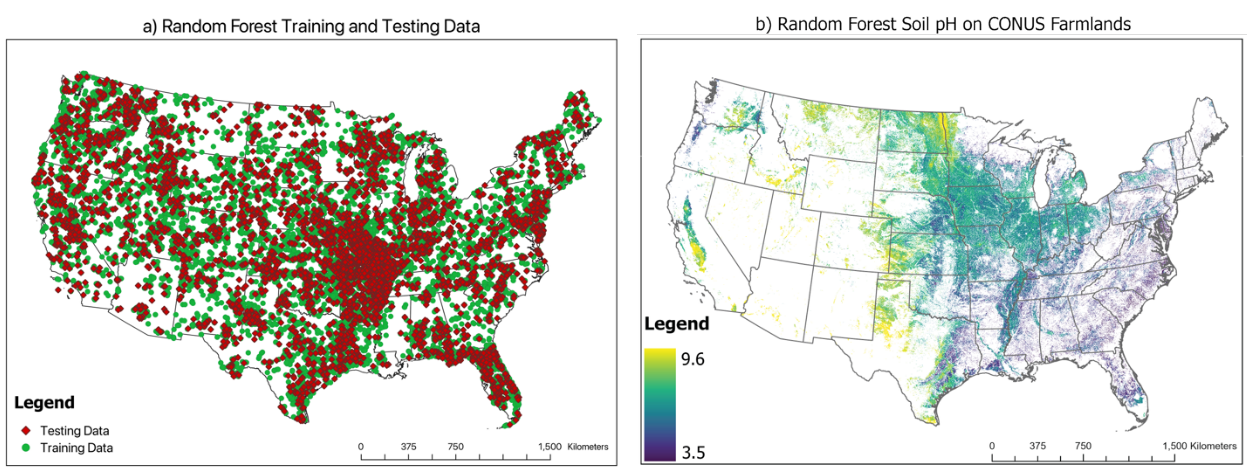


Figure 3(a): RF NCSS training and testing points (80% and 20%, respectively) that are filtered from 1980 or later and from the uppermost portion of the sampled site; *(b)*: RF prediction of pH with 10 km resolution, resampled and clipped to cropland from the 2021 NLCD (30x30m resolution).

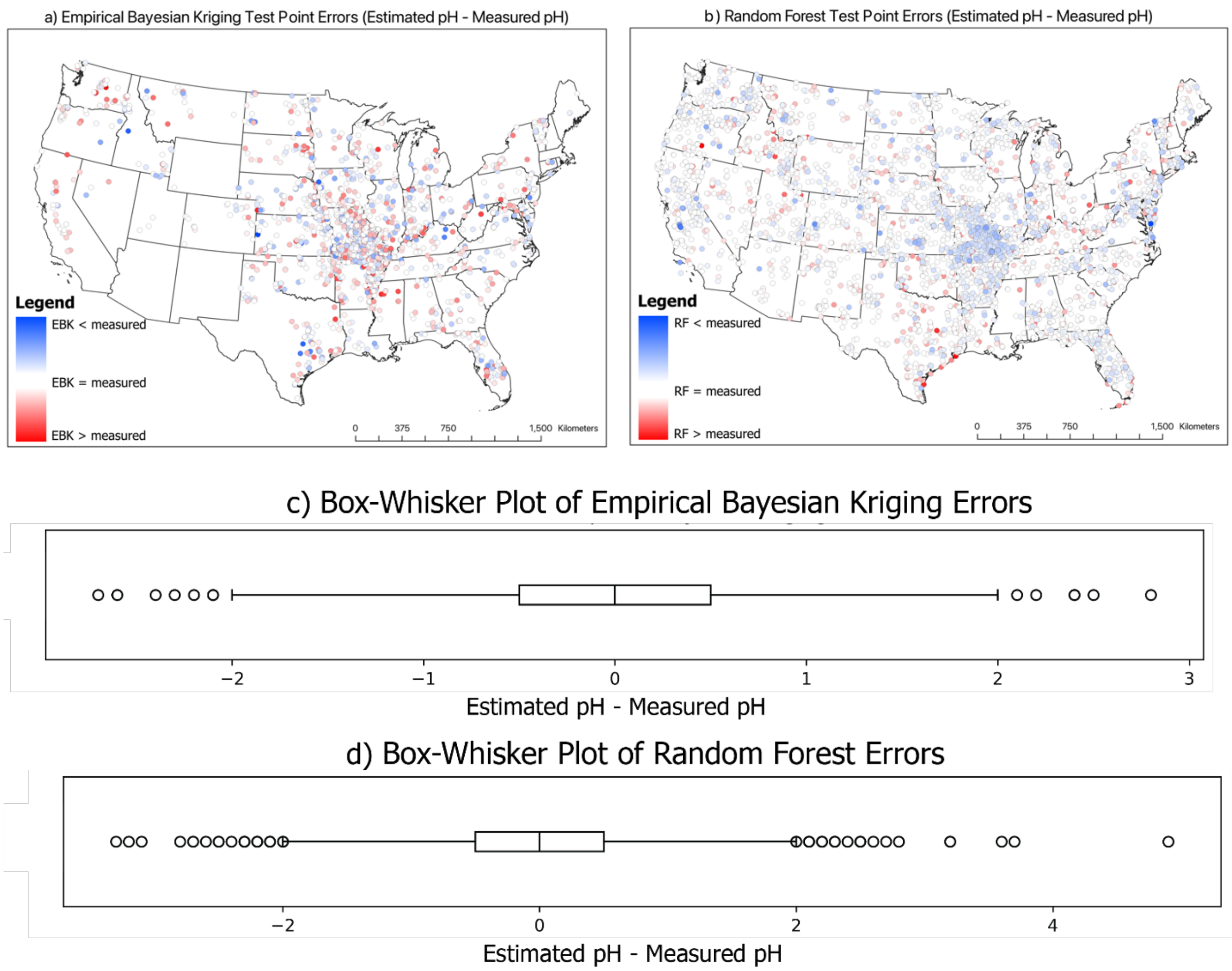


Figure 4(a): Error from n = 2,144 EBK testing points. EBK $\bar{x}$ -0.04, σ 0.75; *(b):* Error from the 5,866 RF testing points. RF $\bar{x}$ 0.003, σ = 0.802; *(c):* box-whisker plot of errors for the 2,144 EBK testing points; *(d):* box-whisker plot of the errors for the 5,866 RF testing points.

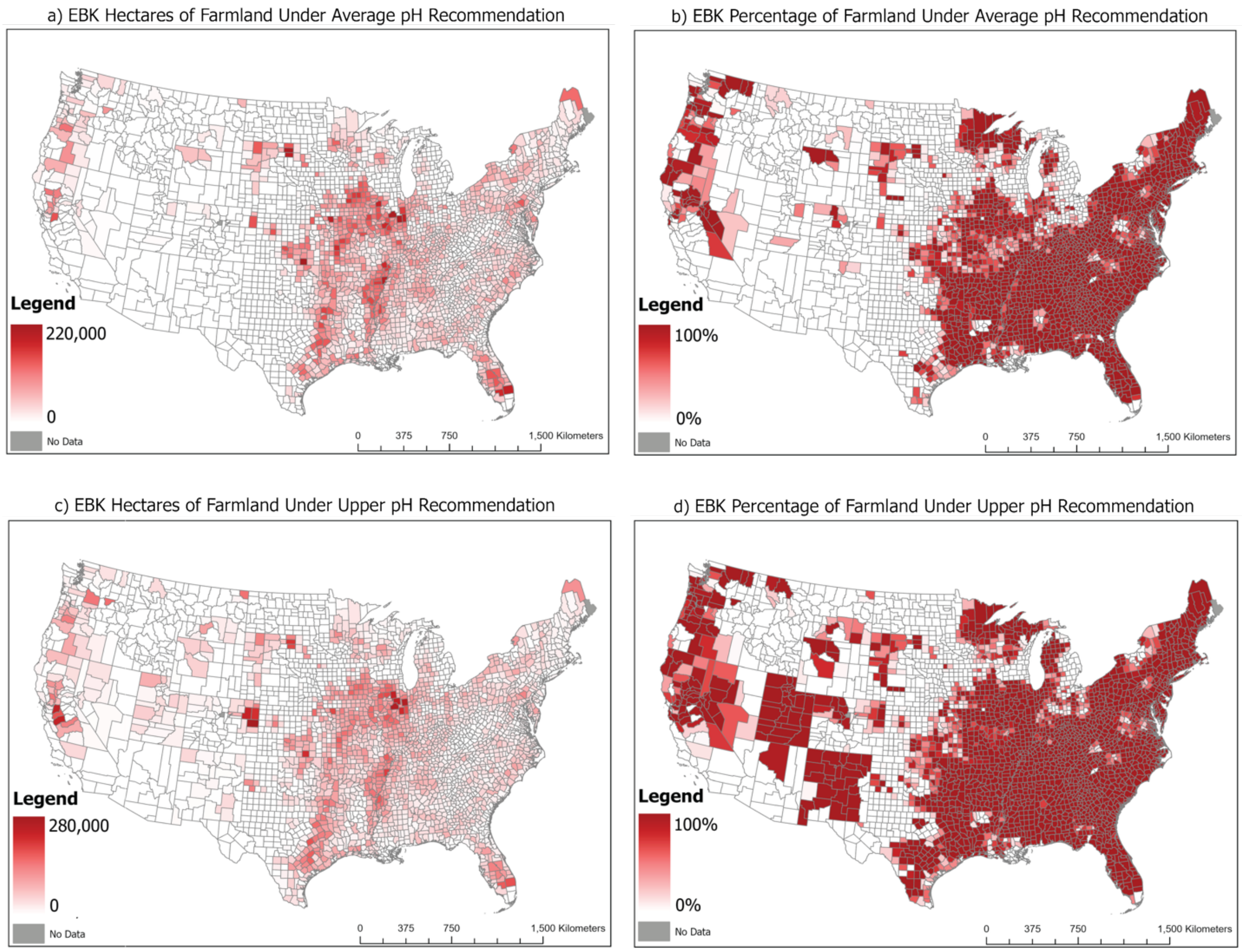
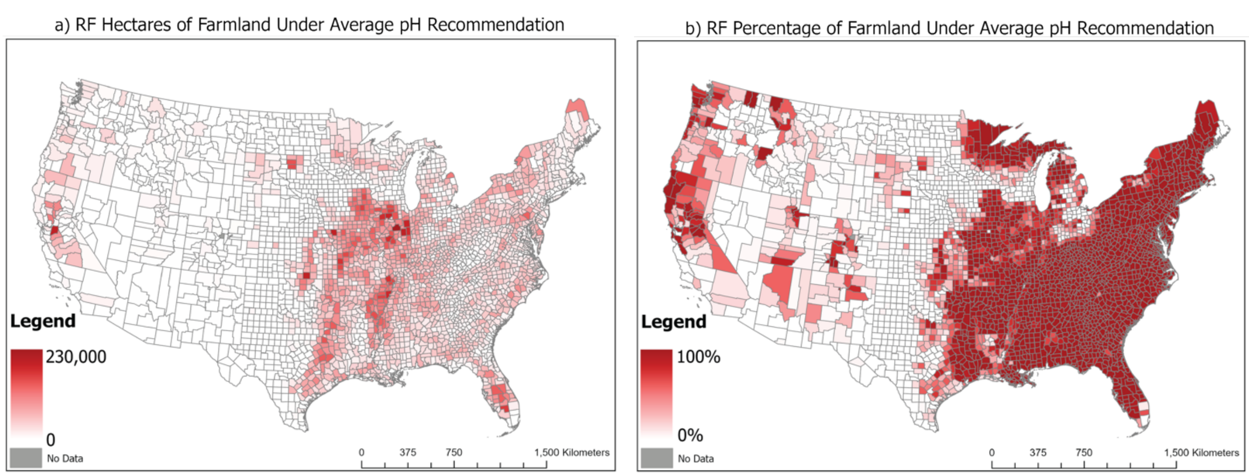


Figure 5(a): Hectares under average pH recommendation using EBK; *(b)*: percentage of cropland under average pH recommendation using EBK; *(c)*: hectares under maximum pH recommendation using EBK; *(d)*: percentage of cropland under maximum pH recommendation using EBK.



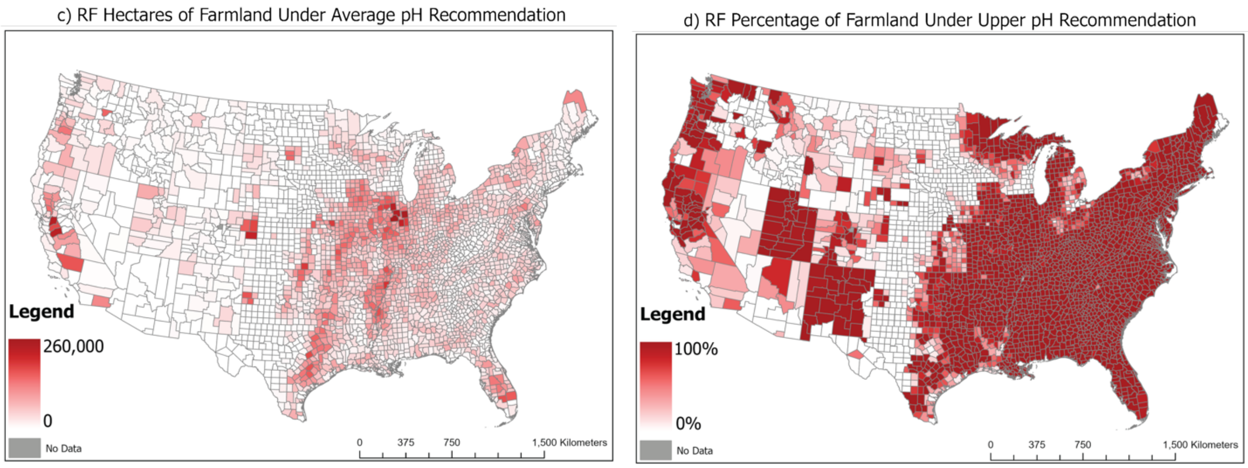


Figure 6(a): Hectares under average pH recommendation using RF; *(b)*: percentage of cropland under average pH recommendation using RF; *(c)*: hectares under maximum pH recommendation using RF; *(d)*: percentage of cropland under maximum pH recommendation using RF.

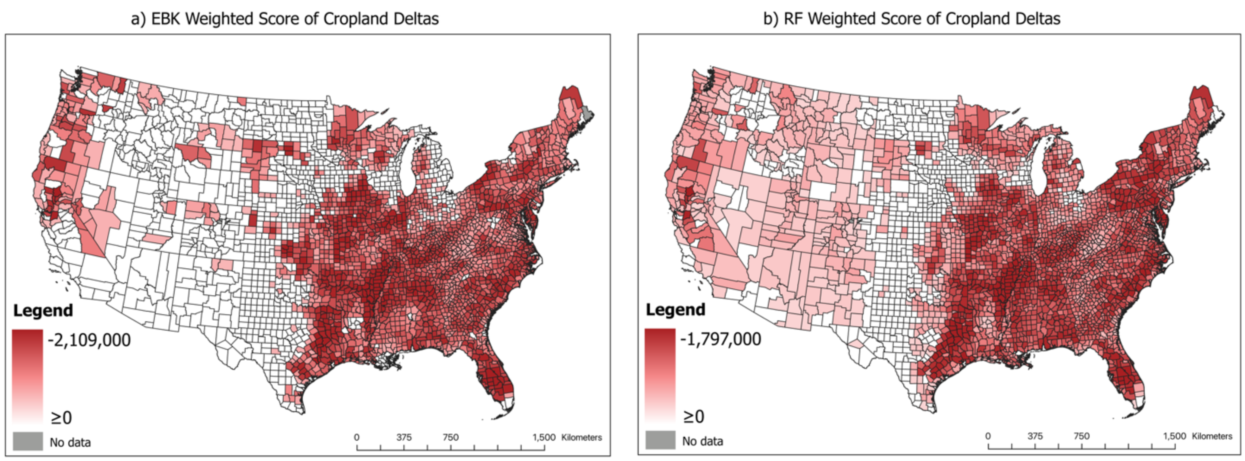


Figure 7(a): Weighted count (W_c) of EBK cells per county under county average extension recommended soil pH, with weights assigned to cells based on the magnitude of EBK minus county recommendation; *(b)*: W_c of RF cells per county under county average extension recommended soil pH, with weights assigned to cells based on the magnitude of RF minus county recommendation.

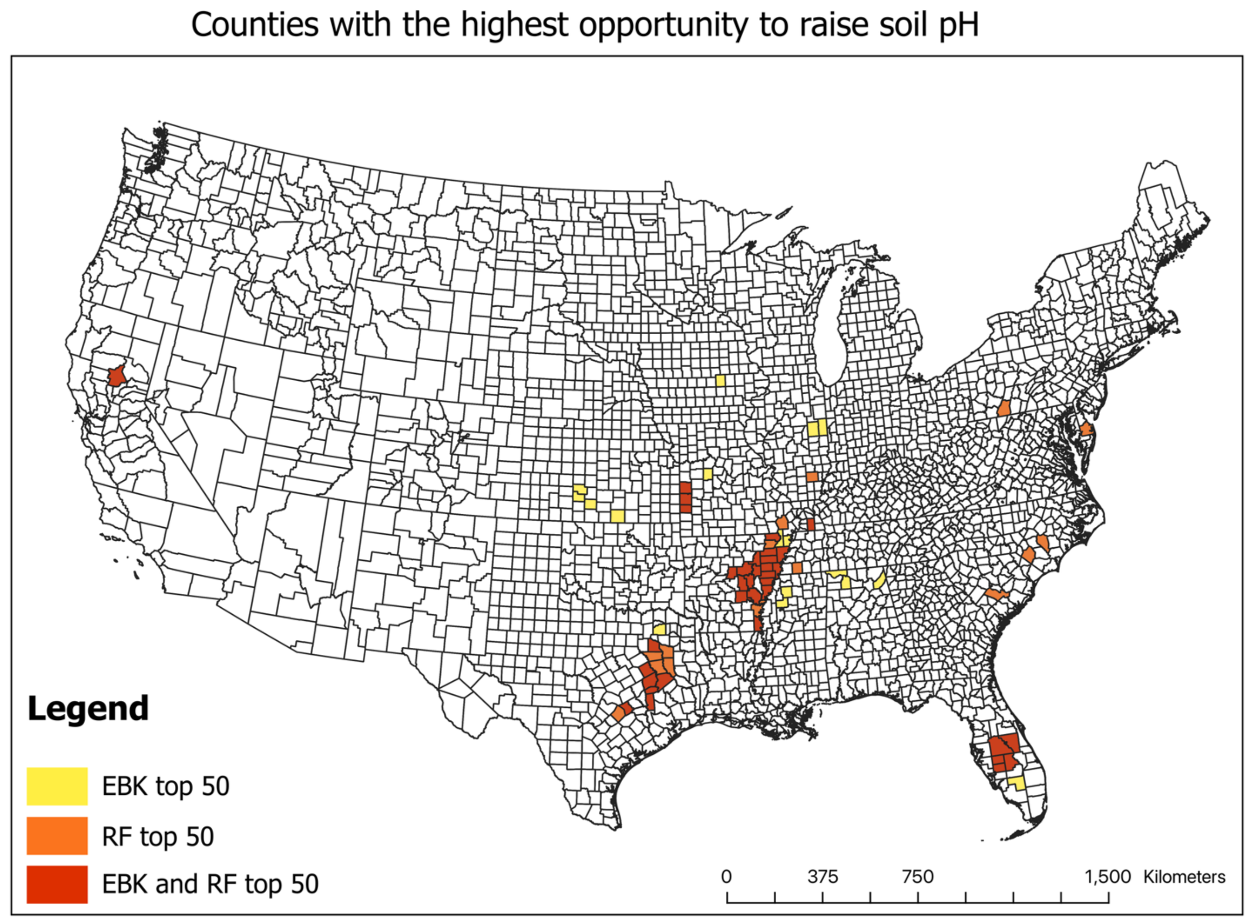
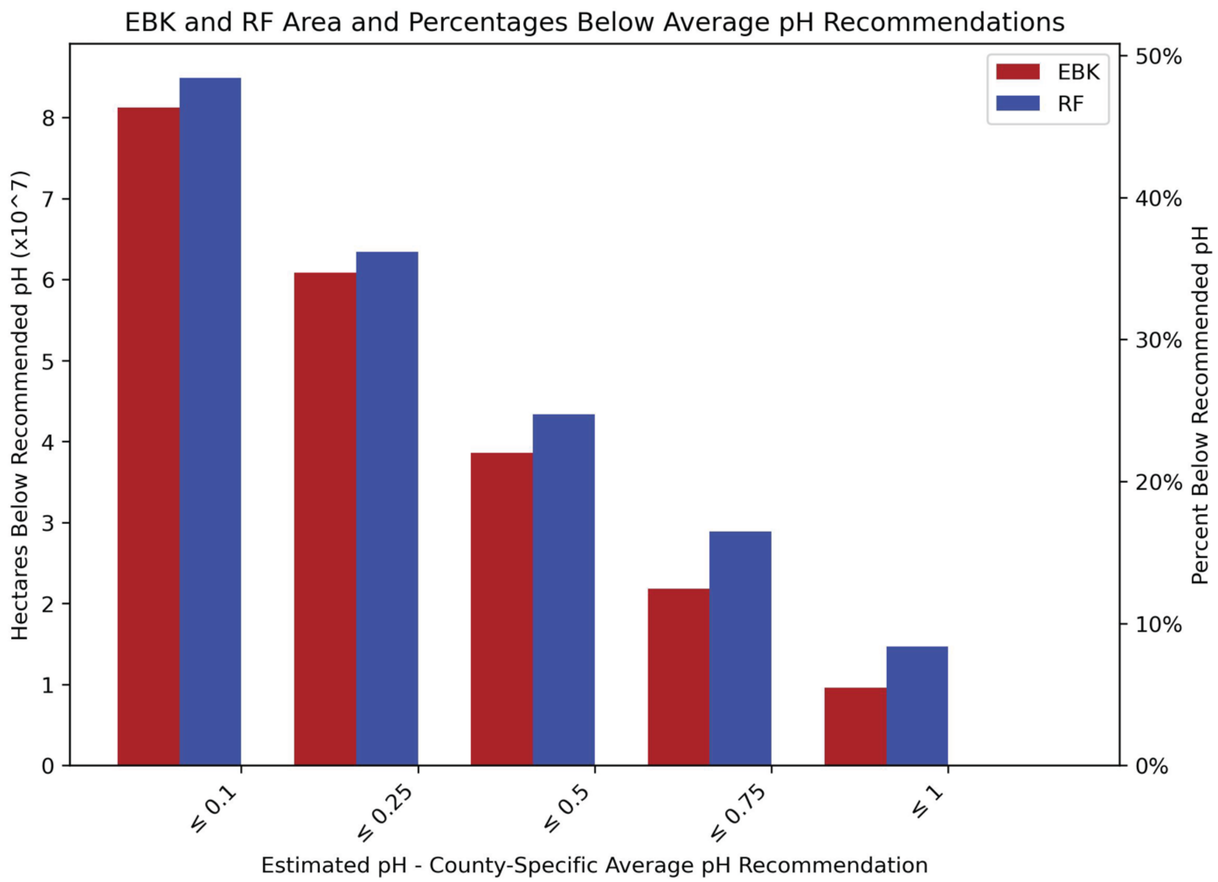


Figure 8: The top 50 counties determined to have the highest opportunity to raise soil pH, as determined using Equation 1. Yellow counties were identified as top 50 highest opportunity by EBK, but not RF; orange counties were identified as highest opportunity by RF, but not EBK; red counties were identified by both EBK and RF as top 50.



	EBK		RF	
Delta	Count (ha)	Percentage	Count (ha)	Percentage
0.1	81,229,711	44.23%	84,895,235	46.22%
0.25	60,823,126	33.12%	63,394,347	34.51%
0.5	38,626,821	21.03%	43,344,636	23.60%
0.75	21,811,654	11.88%	28,881,645	15.72%
1	9,615,026	5.25%	14,709,046	8.01%

Figure 9(a): Box plot of the counts and percentage of cropland hectares below county-specific soil pH recommendations for the most-grown crop in each county; *(b)*: Table containing the counts and percentage of cropland hectares below county-specific soil pH recommendations for the most common crop in each county.

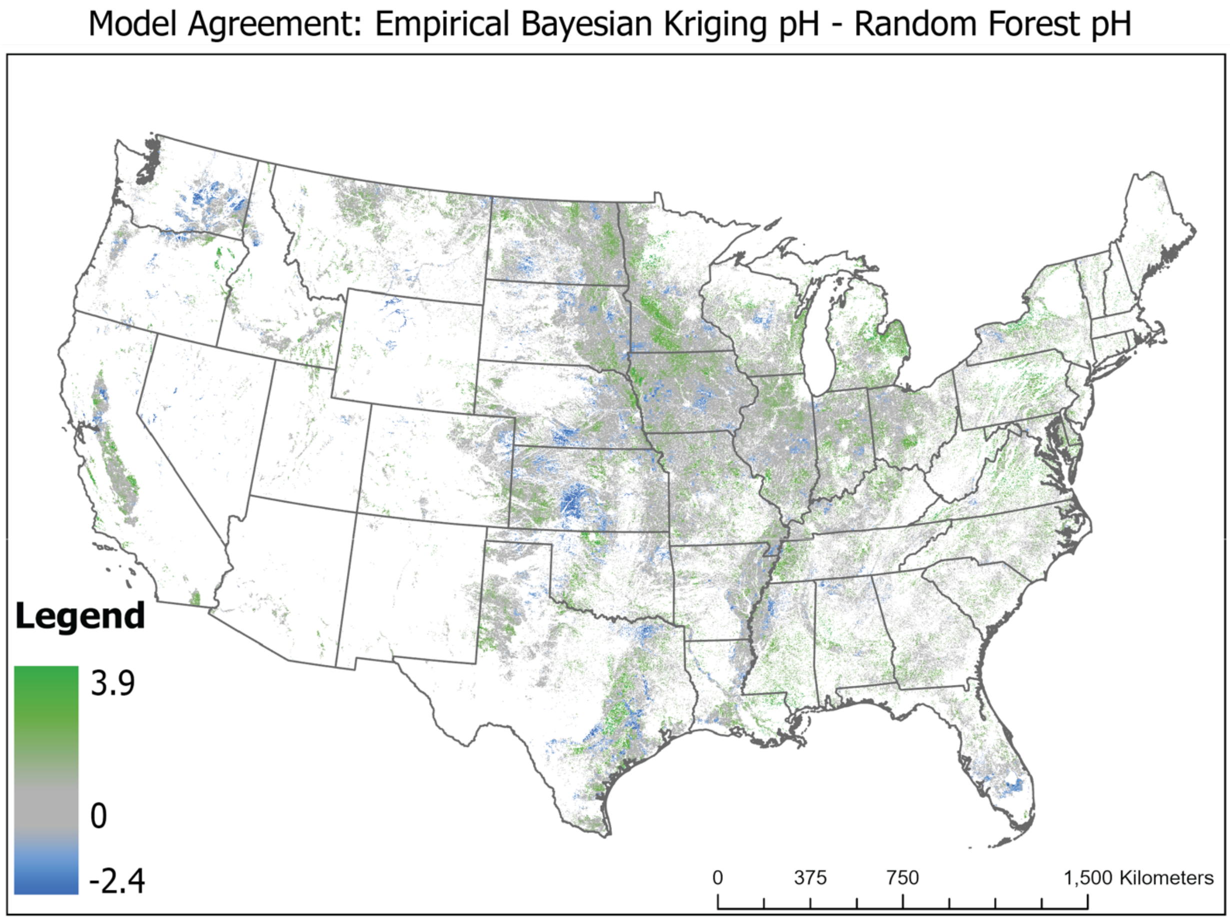


Figure 10: RF interpolation (re-gridded to 30x30m resolution) minus EBK (re-gridded to 30x30m resolution).

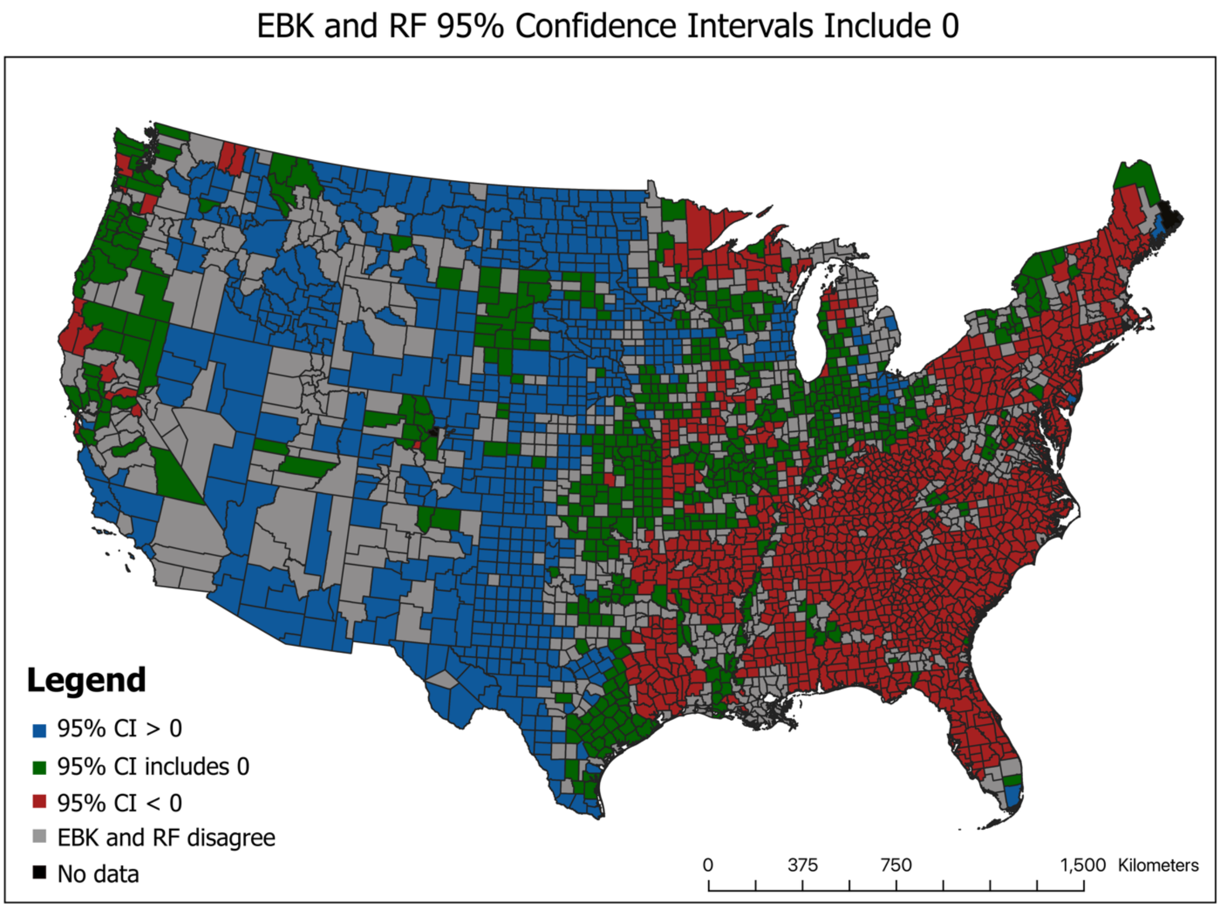


Figure 11: Evaluation of agreement between the mean ±2σ range of county recommended soil pH minus estimated soil pH. In counties colored red, the 2σ is below 0 for both EBK and RF, indicating agreement between the two models that mean county soil pH is below the recommended value. In green, the ±2σ range from mean estimated pH for both EBK and RF overlaps with 0. In counties colored blue, the ±2σ pH range is above 0, indicating agreement that mean county estimated pH is higher than recommended levels. Gray counties indicate disagreement between the EBK and RF.

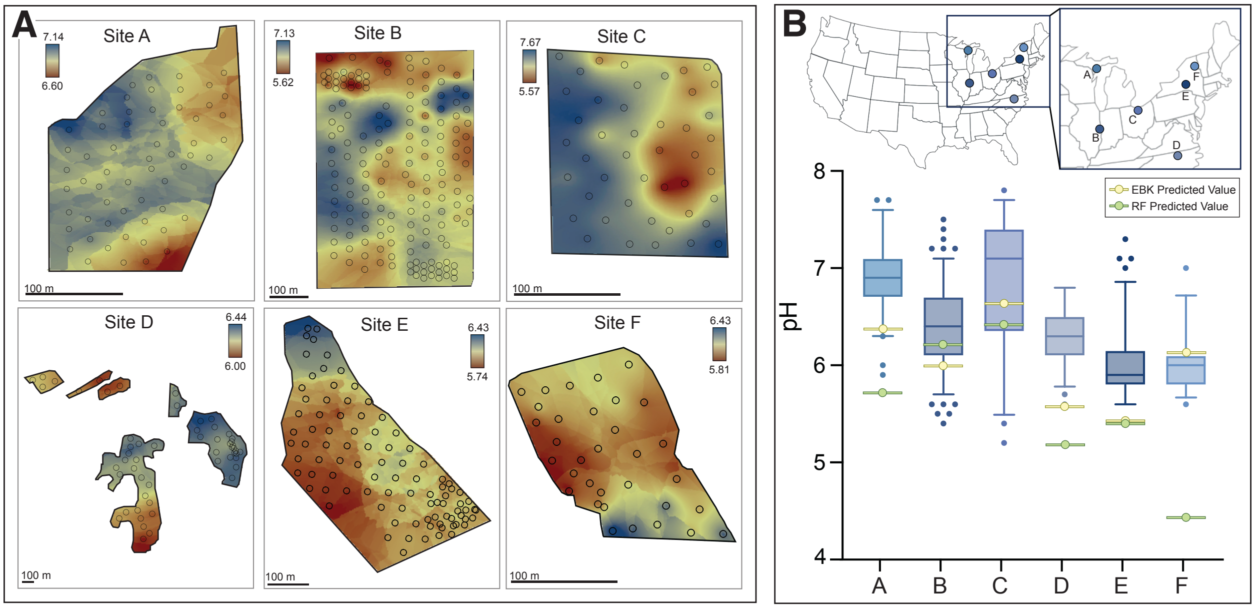


Figure 12(a): EBK-interpolated soil pH across six EW field trial sites located in the US Midwest and East Coast. Additional information on site location and pH analytical approach can be found in Appendix 1; *(b)*: for each included in *(a)*, the 5–95% of data distribution is shown in bars, with outliers plotted as dots.

8. References Cited

Barak, P., Jobe, B. O., Krueger, A. R., Peterson, L. A., & Laird, D. A. (1997). Effects of long-term soil acidification due to nitrogen fertilizer inputs in Wisconsin. *Plant and Soil*, *197*(1), 61–69. https://doi.org/10.1023/A:1004297607070

Beerling, D. J., Kantzas, E. P., Lomas, M. R., Wade, P., Eufrasio, R. M., Renforth, P., et al. (2020). Potential for large-scale CO2 removal via enhanced rock weathering with croplands. *Nature*, *583*(7815), 242–248. https://doi.org/10.1038/s41586-020-2448-9

Beerling, D. J., Epihov, D. Z., Kantola, I. B., Masters, M. D., Reershemius, T., Planavsky, N. J., et al. (2024). Enhanced weathering in the US Corn Belt delivers carbon removal with agronomic benefits. *Proceedings of the National Academy of Sciences*, *121*(9), e2319436121. https://doi.org/10.1073/pnas.2319436121

Bob Nielsen. (2020). Historical Corn Grain Yields In The U.S. [Agronomic extension website]. Retrieved March 26, 2025, from https://extension.entm.purdue.edu/newsletters/pestandcrop/article/historical-corn-grain-yields-in-the-u-s/

Bob Westerfield. (2019). Growing Home Garden Sweet Corn. Retrieved April 6, 2025, from https://extension.uga.edu/publications/detail.html?number=C905&title=growing-home-garden-sweet-

corn#:~:text=Sweet%20corn%20thrives%20best%20in,a%20spade%2C%20plow%20or%20rototiller.

Bodie Pennisi. (2023). Essential pH Management in Greenhouse Crops: pH and Plant Nutrition [Agronomic extension website]. Retrieved March 26, 2025, from https://extension.uga.edu/publications/detail.html?number=B1256&title=essential-ph-management-in-greenhouse-crops-ph-and-plant-nutrition

Brown, B. A., Munsell, R. I., Holt, R. F., & King, A. V. (1956). Soil Reactions at Various Depths as Influenced by Time Since Application and Amounts of Limestone. *Soil Science Society of America Journal*, *20*(4), 518–522. https://doi.org/10.2136/sssaj1956.03615995002000040017x

Cao, P., Lu, C., & Yu, Z. (2018). Historical nitrogen fertilizer use in agricultural ecosystems of the contiguous United States during 1850–2015: application rate, timing, and fertilizer types. *Earth System Science Data*, *10*(2), 969–984. https://doi.org/10.5194/essd-10-969-2018

Chiaravalloti, I., Theunissen, N., Zhang, S., Wang, J., Sun, F., Ahmed, A. A., et al. (2023). Mitigation of soil nitrous oxide emissions during maize production with basalt amendments. *Frontiers in Climate*, *5*, 1203043. https://doi.org/10.3389/fclim.2023.1203043

Dan Undersander. (2001). *Alfalfa Yield and Stand*. Department of Agronomy University of Wisconsin-Extension: University of Wisconsin, Madison. Retrieved from https://fyi.extension.wisc.edu/forage/files/2017/04/Alfalfa-Yield-and-Stand-1.pdf

De Oliveira, E. L., & Pavan, M. A. (1996). Control of soil acidity in no-tillage system for soybean production. *Soil and Tillage Research*, *38*(1–2), 47–57. https://doi.org/10.1016/0167-1987(96)01021-5

Edmund Cecil Shorey. (1940). *The Liming of Soils* (Vol. 1845). Retrieved from https://books.google.com/books?hl=en&lr=&id=psPLMyMGwb4C&oi=fnd&pg=PA3&ots=fxBQYW0qHZ&sig=Re-1bMM92OE-HwpljW25WY05sOw#v=onepage&q&f=false

Gary Schnitkey, Nick Paulson, & Jim Baltz. (2023, June 6). Perspectives on Non-decreasing National Yield Trends [Illinois FarmDocDaily]. Retrieved March 26, 2025, from https://farmdocdaily.illinois.edu/2023/06/perspectives-on-non-decreasing-national-yield-trends.html

Ghimire, R., Machado, S., & Bista, P. (2017). Soil pH, Soil Organic Matter, and Crop Yields in Winter Wheat–Summer Fallow Systems. *Agronomy Journal*, *109*(2), 706–717. https://doi.org/10.2134/agronj2016.08.0462

Gillman, G. P., Burkett, D. C., & Coventry, R. J. (2002). Amending highly weathered soils with finely ground basalt rock. *Applied Geochemistry*, *17*(8), 987–1001. https://doi.org/10.1016/s0883-2927(02)00078-1

Goulding, K. W. T. (2016). Soil acidification and the importance of liming agricultural soils with particular reference to the United Kingdom. *Soil Use and Management*, *32*(3), 390–399. https://doi.org/10.1111/sum.12270

Gribov, A., & Krivoruchko, K. (2020). Empirical Bayesian kriging implementation and usage. *Science of The Total Environment*, *722*, 137290. https://doi.org/10.1016/j.scitotenv.2020.137290

Guo, J. H., Liu, X. J., Zhang, Y., Shen, J. L., Han, W. X., Zhang, W. F., et al. (2010). Significant Acidification in Major Chinese Croplands. *Science*, *327*(5968), 1008–1010. https://doi.org/10.1126/science.1182570

Hartemink, A. E., & Barrow, N. J. (2023). Soil pH - nutrient relationships: the diagram. Plant and Soil, 486(1–2), 209–215. https://doi.org/10.1007/s11104-022-05861-z

Intergovernmental Panel On Climate Change (IPCC). (2023). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Jon Dewitz. (2023). National Land Cover Database (NLCD) 2021 Products [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9JZ7AO3

Kantzas, E. P., Val Martin, M., Lomas, M. R., Eufrasio, R. M., Renforth, P., Lewis, A. L., et al. (2022). Substantial carbon drawdown potential from enhanced rock weathering in the United Kingdom. *Nature Geoscience*, *15*(5), 382–389. https://doi.org/10.1038/s41561-022-00925-2

Kanzaki, Y., Planavsky, N. J., Zhang, S., Jordan, J., Suhrhoff, T. J., & Reinhard, C. T. (2025). Soil cation storage is a key control on the carbon removal dynamics of enhanced weathering. Environmental Research Letters, 20(7), 074055. https://doi.org/10.1088/1748-9326/ade0d5

Krivoruchko, K., & Gribov, A. (2014). Pragmatic Bayesian Kriging for Non-Stationary and Moderately Non-Gaussian Data. In E. Pardo-Igúzquiza, C. Guardiola-Albert, J. Heredia, L. Moreno-Merino, J. J. Durán, & J. A. Vargas-Guzmán (Eds.), *Mathematics of Planet Earth* (pp. 61–64). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-32408-6_15

Kukla, T., Kanzaki, Y., Chay, F., Planavsky, N., & Reinhard, C. (2025, August 23). Swapping carbonate for silicate in agricultural enhanced rock weathering. CDRXIV. https://doi.org/10.70212/cdrxiv.2025304.v1

National Cooperative Soil Survey. (2023). NCSS Characterization Database ESRI File Geodatabase [Data set]. Retrieved from https://ncsslabdatamart.sc.egov.usda.gov/database_download.aspx

Pagani, A., & Mallarino, A. P. (2012). Soil pH and Crop Grain Yield as Affected by the Source and Rate of Lime. *Soil Science Society of America Journal*, *76*(5), 1877–1886. https://doi.org/10.2136/sssaj2012.0119

Pagani, A., & Mallarino, A. P. (2015). On-Farm Evaluation of Corn and Soybean Grain Yield and Soil pH Responses to Liming. *Agronomy Journal*, *107*(1), 71–82. https://doi.org/10.2134/agronj14.0314

Raymond, P., Planavsky, N., & Reinhard, C. T. (2025). Using carbonates for carbon removal. Nature Water, 3(8), 844–847. https://doi.org/10.1038/s44221-025-00473-0

Roberto Mosheim. (2025). Fertilizer Use and Price - Summary of Findings. Retrieved March 30, 2025, from https://www.ers.usda.gov/data-products/fertilizer-use-and-price/summary-of-findings

Sharry, R. L., Smith, M., Moulay, H., Souza, J., Sharma, S., & Arnall, D. B. (2025). Soil pH influence on cotton lint yield and quality. *Agrosystems, Geosciences & Environment*, *8*(1), e70065. https://doi.org/10.1002/agg2.70065

Sirisuntornlak, N., Ullah, H., Sonjaroon, W., Anusontpornperm, S., Arirob, W., & Datta, A. (2021). Interactive Effects of Silicon and Soil pH on Growth, Yield and Nutrient Uptake of Maize. *Silicon*, *13*(2), 289–299. https://doi.org/10.1007/s12633-020-00427-z

Suhrhoff, T. J., Reinhard, C., Kanzaki, Y., Tsao, S. S.-E., Woollen, B., Reershemius, T., et al. (2026, June 19). Agricultural liming is a large carbon sink in the Mississippi River Basin. In Review. https://doi.org/10.21203/rs.3.rs-7068561/v1

USDA NASS. (2023). CroplandCROS 2022 [Data set]. Retrieved from https://croplandcros.scinet.usda.gov/

West, T. O., & McBride, A. C. (2005). The contribution of agricultural lime to carbon dioxide emissions in the United States: dissolution, transport, and net emissions. *Agriculture, Ecosystems & Environment*, *108*(2), 145–154. https://doi.org/10.1016/j.agee.2005.01.002

Wieczorek, M. E. (2019). Area and depth-weighted average of soil pH from STATSGO2 for the conterminous United States and District of Columbia [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P95YOSFQ

Wu, Y.-H., Hung, M.-C., & Patton, J. (2013). Assessment and visualization of spatial interpolation of soil pH values in farmland. *Precision Agriculture*, *14*(6), 565–585. https://doi.org/10.1007/s11119-013-9316-7

9. Supplemental Information

The supplemental information section of this paper is split into four appendices. Appendix I is a compilation of figures referenced in the paper, that provide supporting information on the model development process. Appendix II contains information on different kriging models that were tested on the NCSS data. Appendix III contains additional zonal statistics for the county-level differences between EBK or RF and county-specific average pH recommendations. Appendix IV contains tabular data on the degree to which each state and agronomic region is below county-specific average soil pH recommendations. Appendix V contains additional details about the agronomic extension soil pH recommendations collected for this study. And lastly, Appendix VI contains tabular data on the 50 counties identified by RF and EBK as having the highest opportunity to raise soil pH, using Equation 1.

9.1. Appendix I: Model Development and Comparison

In selecting the EBK approach to soil pH, we first ruled out several alternatives. Compared to ordinary kriging (discussed below), Saffari et al. (2009) found IDW to produce less accurate estimates of soil pH. Additionally, IDW and other deterministic interpolations have been found to produce more extreme estimates of agricultural soil pH (e.g., higher maximums and lower minimums) (Wu et al., 2013). Inferring that managed soils are less likely to gravitate toward extreme pH values, we prioritized interpolations that would tend toward more neutral values (see Figure SI 6). We considered ordinary kriging (OK) which performed very similarly in terms of accuracy to EBK with the NCSS data. We selected the EBK because it had a better-fitting QQ plot and a lower root mean square standardized error (Supplemental Information, Appendix II). The EBK interpolation method is also designed to accommodate data with a clear directional trend in values, exhibiting non-stationarity (Krivoruchko & Gribov, 2014). It is well documented that the Eastern half of CONUS has, in general, more acidic soils (Wieczorek 2019, West & McBride 2005, Shorey 1940) and historically has received more aglime (West & McBride, 2005).

We tested 9 different combinations of neighborhood subset size (50, 100, and 150) and overlap factor (1, 3, and 5). We ran each of these combinations at 100 simulations. Although increasing number of simulations (n) has been found to improve model accuracy (Gribov & Krivoruchko, 2020), we noticed minimal improvements from increased n, up to $n = 300$. The performance of all 9 models tested was very similar. We selected the EBK with a subset size of 100 and an overlap factor of 3 due to a particularly normal QQ plot (Supplemental Information Appendix II, Figure 6).

Because the EBK interpolation is spatially constrained by the farthest NCSS training point in each cardinal direction, the total sum of cropland hectares in the EBK analysis is slightly below the RF. The EBK interpolation does not include cropland in Washington County, Maine, as the easternmost NCSS point in the reclassified NLCD is west of the county boundary. The RF model does not require points to be filtered by cropland (section 2.4), meaning this model does include cropland in Washington County. This difference, however, is negligible at the CONUS scale and does not substantively impact the findings discussed in this analysis.

The first figure below depicts non-stationarity (essentially, discernibly directional heterogeneity) in the input NCSS data that were used to develop the EBK model. For reference, these points were filtered from the NCSS dataset to be on cropland (according to the 2021 NLCD layer), from 1980 or later according to the observation_date variable in the NCSS database (or in cases where observation_date was null, site_obsdate), and each point was the uppermost measurement in cases where samples were taken at multiple depths.

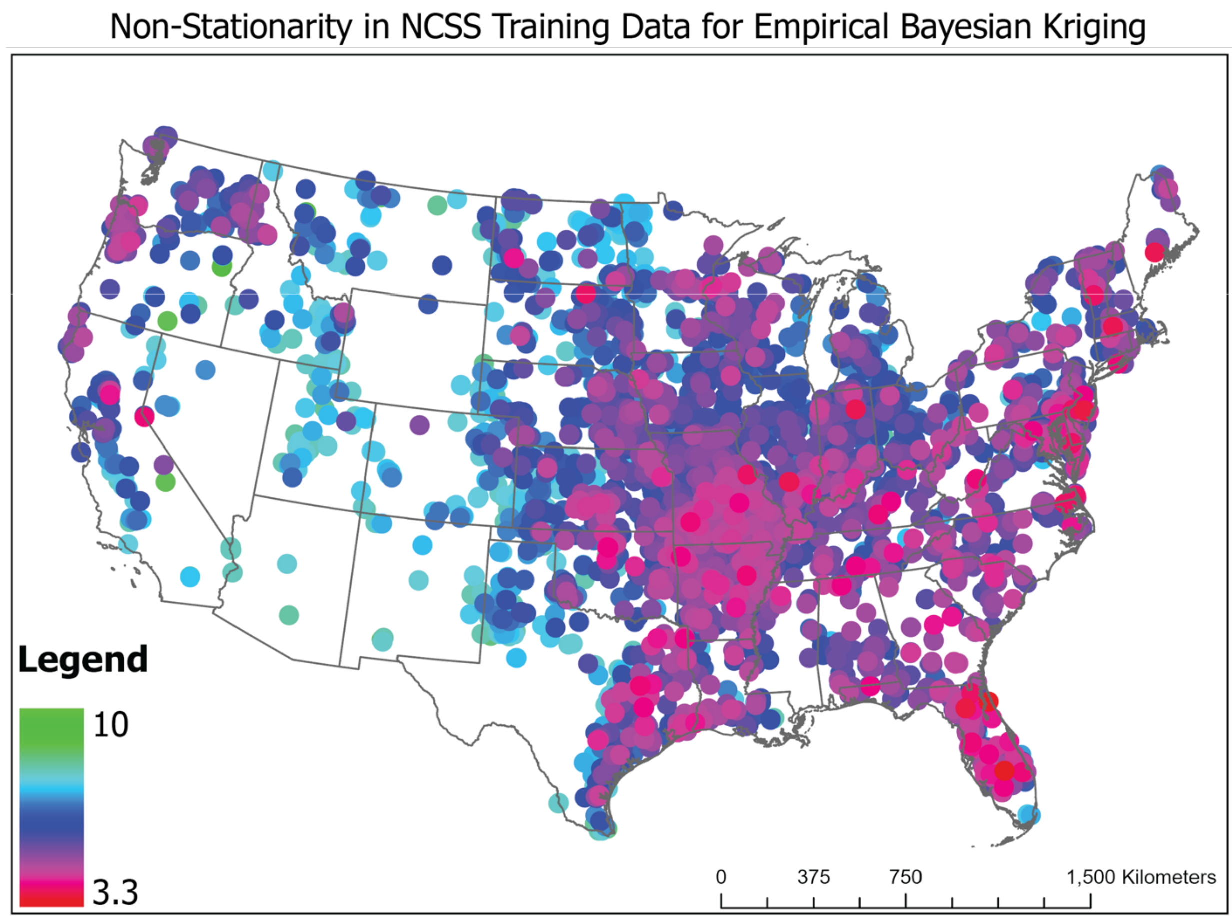


Figure SI 1: NCSS data used to develop the EBK interpolation exhibit considerable non-stationarity, with notably more alkaline, and fewer, samples in the central western region of CONUS. Point size has been increased and color scheme exaggerated to improve figure legibility.

The list below contains the references used to create the RF model used in our analysis. In addition to the variables bulleted below, latitude and longitude were also used as covariates.

- Elevation: Amatulli et al. (2018); aggregated from 30-arc second (~1 km in equator) [constant].
- Land slope: Amatulli et al. (2020); aggregated from 7.5-arc second (~250 m) [constant].
- Basic rock: Hartmann and Moosdorf (2012); rasterized from polygon shapefile [constant].
- Acidic-intermediate rock: Hartmann and Moosdorf (2012); rasterized from polygon shapefile [constant].
- Carbonate rock: Hartmann and Moosdorf (2012); rasterized from polygon shapefile [constant].
- Mixed rock: Hartmann and Moosdorf (2012); rasterized from polygon shapefile [constant].
- Temperature: Karger et al. (2017); aggregated from 30-arc second (~1 km) [1981-2010].

- Temperature seasonality: Karger et al. (2017); aggregated from 30-arc second (~1 km) [1981-2010]; standard deviation of the monthly mean temperatures.
- Precipitation: Karger et al. (2017); aggregated from 30-arc second (~1 km) [1981-2010].
- Precipitation seasonality: Karger et al. (2017); coefficient of variation (in %).
- Quick-flow runoff: Reitz et al. (2017); aggregated from 800 m [2000-2013].
- Evapotranspiration: Reitz et al. (2017); aggregated from 800 m [2000-2013].
- Soil moisture: Vergopolan et al. (2021); aggregated from 30 arc-second (~1 km) [2015-2019].
- Soil porosity: Boiko et al. (2021); aggregated from 3 arc-second (~90 m) [constant].
- Soil CEC: Walkinshaw et al. (2022); aggregated from 800 m [constant].
- Soil respiration rate: Huang et al. (2020); aggregated from 30 arc-second (~1 km) [2010].
- Groundwater table depth: Fan et al. (2013); aggregated from 30 arc-second (~1 km) [constant].
- Net primary production: Zhao et al. (2005; 2010); aggregated from 30 arc-second (~1 km) [2000-2015].
- Cropland proportion: Dewitz and USGS (2023); aggregated from 30 m [2021].
- Pastureland proportion: Dewitz and USGS (2023); aggregated from 30 m [2021].
- Cropland proportion (consensus land-cover): Tuanmu and Jetz (2014); aggregated from 30 arc-second (~1 km) [constant close to 2005].
- Tree proportion (consensus land-cover): Tuanmu and Jetz (2014); aggregated from 30 arc-second (~1 km) [constant close to 2005].
- Vegetation proportion (consensus land-cover): Tuanmu and Jetz (2014); aggregated from 30 arc-second (~1 km) [constant close to 2005].
- Urban land proportion (consensus land-cover): Tuanmu and Jetz (2014); aggregated from 30 arc-second (~1 km) [constant close to 2005].
- Irrigation frequency: LANID-US; Xie et al. (2021); aggregated from 30 m [the number of years irrigated during 1997-2017].
- Area equipped for irrigation (AEI) in percent of land area: Siebert et al. (2015); 5 arc-min [1950, 1960, 1970, 1980, 1985, 1990, 1995, 2000, 2005]; x-y flipped. multiple bands instead of time points. coupling R, gdal, and cdo.
- Fertilizer + manure N application on cropland + pastureland + rangeland: History of anthropogenic Nitrogen inputs (HaNi): Tian et al. (2022); 5-arc minute (~9.25 km) [1970, 1980, 1990, 2000, 2010, 2017]; original data is application rate per grid cell. Coupled with grid cell area to obtain the final result.
- Fertilizer + manure N application on cropland: History of anthropogenic Nitrogen inputs (HaNi): Tian et al. (2022); 5-arc minute (~9.25 km) [1970, 1980, 1990, 2000, 2010, 2017]; original data is application rate per grid cell. Coupled with grid cell area and Hyde (History database of the Global Environment) v3.2 cropland proportion to obtain the application rate per m^2 cropland per year.

- Population: Gridded Population of the World (GPW), v4; Center For International Earth Science Information Network (2016); aggregated from 30-arc second (~1 km in equator) [2010].

Finally, included below is Table SI 1, referenced in section 2.9 of the main text. This table contains field-specific data for six sites measured for pH in EW field trials conducted by the Yale Center for Natural Carbon Capture. Each sample is a composite of 2-10 cores within 1m of the set sampling point. Composite soil samples were dried, sieved, ground, and measured for pH by Waypoint Analytical Labs using 1:1 $CaCl_2$.

Site	Lat	Long	Area (ha)	Cores per sample	Sample Depth (cm)	Total # Samples
A	45.20	-87.61	6.53	2	0-10	59
B	40.79	-87.54	32.41	10	0-10	165
C	41.40	-81.61	4.37	6-8	0-15	45
D	36.40	-78.30	31.64	15	0-15	55
E	42.81	-75.16	15.75	10	0-10	94
F	44.37	-73.71	4.76	10	0-10	33

Table SI 1: Measurement details for six field sites used to evaluate the performance of EBK and RF for field-scale geospatial pH estimates.

9.2. Appendix II: Comparison Between Kriging Models

The EBK model used for this interpolation was developed using the Geostatistical Wizard tool in ArcGIS Pro, which generates a variety of diagnostic values and plots to interpret its output. A selection of the output figures is shown below.

Count	8578
Average CRPS	0.396850421425353
Inside 90 Percent Interval	90.1025880158545
Inside 95 Percent Interval	95.2319888085801
Mean	0.00399766597676479
Root-Mean-Square	0.727225091841044
Mean Standardized	0.00473968734565771
Root-Mean-Square Standardized	0.993299970336893
Average Standard Error	0.724980007307825

Cross validation

Cross validation is a "leave one out" method that allows you to determine how well your interpolation model fits your data. Cross validation works by removing a single point from the dataset and using all remaining points to predict to the location of the point that was removed. The predicted value is then compared to the measured value, and many statistics are generated to determine the accuracy of the prediction.

Figure SI 3: Tabular output generated from the Geostatistical Wizard for our selected EBK model.

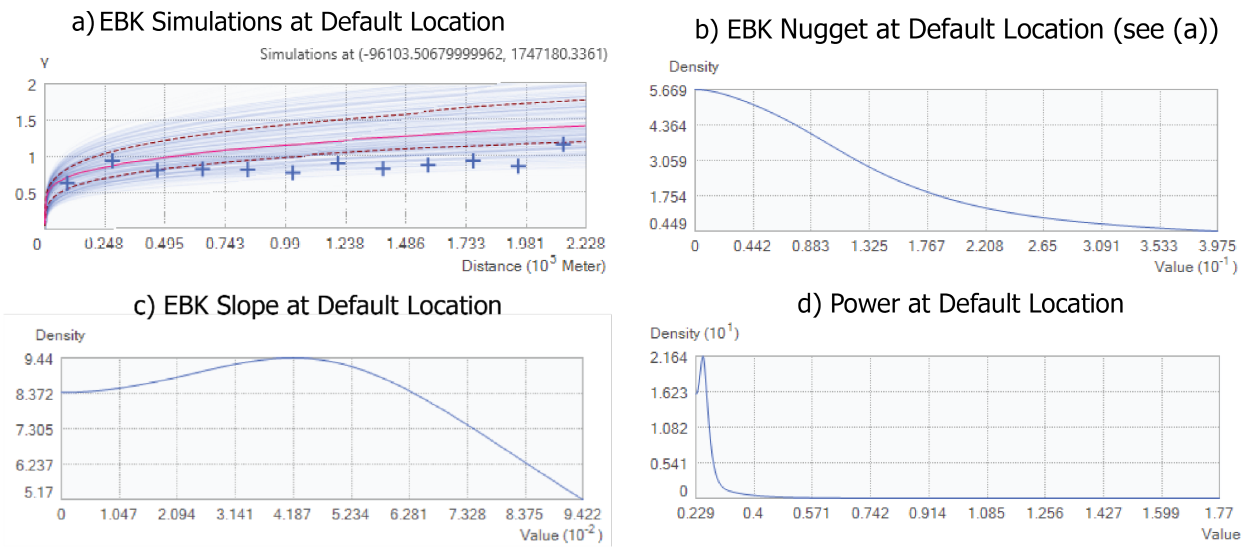


Figure SI 4(a): Simulations at the default location (approximately at the center of CONUS) for our EBK interpolation; *(b):* nugget at the default location; *(c):* slope at the default location; *(d):* power at the default location.

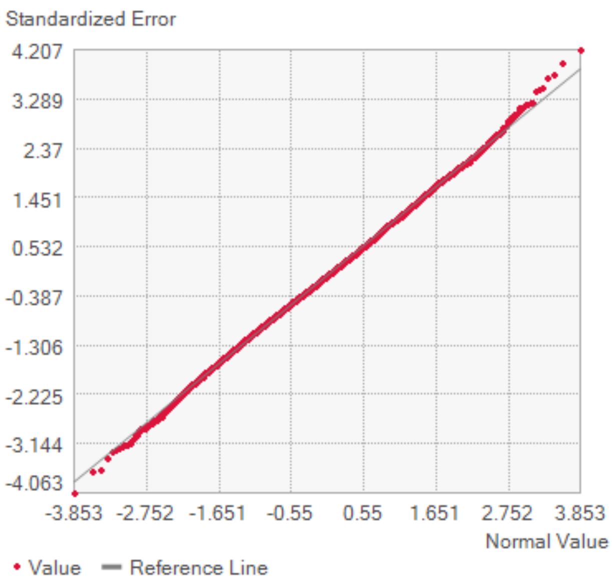


Figure SI 5: Normal QQ Plot for the EBK generated. Of the various interpolations we tested, the combination of parameters we selected (subset size 100, overlap factor of 3, and *n* simulations = 100) had the best-fitting normal QQ plot.

9.3: Appendix III: Exploratory Data Analysis of EBK and RF

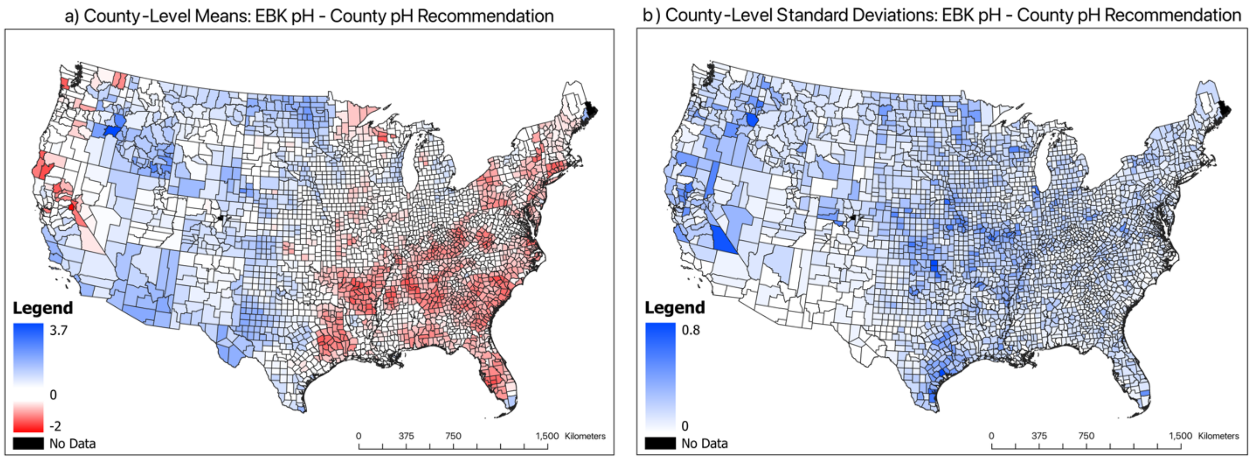


Figure SI 6(a): The mean of EBK interpolation cropland cells minus average pH recommendation in each CONUS county; *(b):* The standard deviation of EBK interpolation cropland cells minus average pH recommendation in each CONUS county.

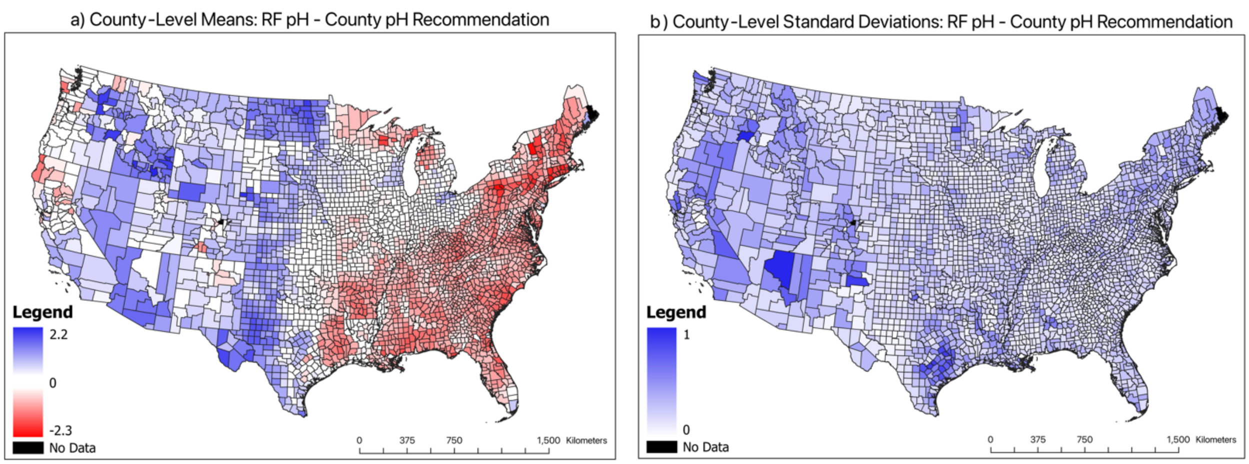


Figure SI 7(a): The mean of RF interpolation cropland cells minus average pH recommendation in each CONUS county; *(b):* The standard deviation of RF interpolation cropland cells minus average pH recommendation in each CONUS county.

9.4: Appendix IV: Tabular Data on States and USDA Agronomic Regions

This appendix contains data on the amount and percentage of each USDA agronomic region below county-specific average extension-recommended, using a 0.1 pH threshold.

USDA Agronomic Region	Total Sum (Ha)	Sum Under Average (Ha)	Percent Under Average
Appalachia	12709698.21	11830137.3	93.07960822
Corn Belt	42681809.61	25217632.17	59.08285614
Delta	11115466.02	9859946.4	88.7047505
Lake	18552146.94	2777092.92	14.96911882
Mountain	14076532.44	527368.68	3.74643885
Northeast	7366032.27	6284414.16	85.31613669

Northern Plains	37897256.61	5810566.14	15.33241891
Pacific	8850810.87	1997162.01	22.56473491
Southeast	9773526.69	9331773.93	95.48010893
Southern Plains	20635903.98	7593616.98	36.79808254

Table SI 2: The amount and percentage of hectares in each USDA agronomic region below county-specific average extension recommendations, according to our EBK estimate.

USDA Agronomic Region	Total Sum (Ha)	Sum Under Average (Ha)	Percent Under Average
Appalachia	12709698.21	12564345.78	98.856366
Corn Belt	42681809.61	28804853.79	67.48742392
Delta	11115466.02	9765398.34	87.85415135
Lake	18552146.94	3671135.28	19.78819644
Mountain	14076532.44	421571.43	2.994852829
Northeast	7376858.73	6927010.74	93.90190315
Northern Plains	37897256.61	3715969.05	9.80537744
Pacific	8853159.24	2000817.18	22.60003605
Southeast	9773527.23	9476054.28	96.95633989
Southern Plains	20646842.22	7548079.14	36.55803178

Table SI 3: The amount and percentage of hectares in each USDA agronomic region below county-specific average extension recommendations, according to our RF estimate.

9.5: Appendix V: Agronomic Data

The following criteria were used to obtain state-specific soil pH recommendations for the most grown crop in each county.

1. Select data from a university-managed agronomic extension source within the state of interest. In the 138 counties where this was unavailable, we entered data from a university-managed agronomic website from another state (either a neighboring state or top result on internet search).
2. When multiple, conflicting pH recommendations exist for the dominant crop within a county, record the recommendation that was more recently published.
3. Include the URL and timestamp of any data sources used in the data entry, along with any relevant notes.

Several states, including Michigan, Wisconsin, and North Carolina, made different recommendations for mineral and organic soils, while most states did not differentiate in the published extension materials we reviewed. The definitions adopted by specific maps and agronomic extensions vary but are generally differentiated by concentration of organic matter. Mineral soils constitute an overwhelming majority of cropland, e.g. (Michigan State University & US Department of Agriculture, 1981). For these cases, we located a government-produced soils map of the state that differentiated mineral and organic soils that had counties delineated. If a county had exclusively mineral or organic soil, the corresponding recommendations were recorded; in cases where a county had both soil categories (even if unequal amounts) an unweighted average of the mineral and organic soil pH recommendations for the dominant crop were calculated and recorded.

9.6: Appendix VI: Counties with the Highest Opportunity to Raise Soil pH

The table below reflects the top 50 counties identified by Equation 1 as having high potential to raise soil pH to extension-recommended levels.

Model	County	State	Rank
EBK	Leon County	Texas	1
EBK	Lonoke County	Arkansas	2
EBK	Highlands County	Florida	3
EBK	Houston County	Texas	4
EBK	Hendry County	Florida	5
EBK	Chicot County	Arkansas	6
EBK	DeSoto County	Florida	7
EBK	Poinsett County	Arkansas	8
EBK	Butte County	California	9
EBK	White County	Arkansas	10
EBK	Polk County	Florida	11
EBK	Craighead County	Arkansas	12
EBK	Sumner County	Kansas	13
EBK	Arkansas County	Arkansas	14
EBK	Cross County	Arkansas	15
EBK	Woodruff County	Arkansas	16
EBK	Jefferson County	Arkansas	17
EBK	Grimes County	Texas	18
EBK	Prairie County	Arkansas	19
EBK	Osceola County	Florida	20
EBK	Hopkins County	Texas	21
EBK	Clay County	Arkansas	22
EBK	Panola County	Mississippi	23
EBK	Jackson County	Arkansas	24
EBK	Crittenden County	Arkansas	25
EBK	Pemiscot County	Missouri	26
EBK	Madison County	Texas	27
EBK	Mississippi County	Arkansas	28

EBK	Graves County	Kentucky	29
EBK	Phillips County	Arkansas	30
EBK	Champaign County	Illinois	31
EBK	St. Francis County	Arkansas	32
EBK	Hardee County	Florida	33
EBK	Lauderdale County	Alabama	34
EBK	Barton County	Missouri	35
EBK	Vernon County	Missouri	36
EBK	Vermilion County	Illinois	37
EBK	Tama County	Iowa	38
EBK	Lawrence County	Alabama	39
EBK	Pratt County	Kansas	40
EBK	Pawnee County	Kansas	41
EBK	Tallahatchie County	Mississippi	42
EBK	Dunklin County	Missouri	43
EBK	Freestone County	Texas	44
EBK	DeKalb County	Alabama	45
EBK	Faulkner County	Arkansas	46
EBK	Pettis County	Missouri	47
EBK	Van Zandt County	Texas	48
EBK	Edwards County	Kansas	49
EBK	Bates County	Missouri	50

Table SI 3: Top 50 counties identified by EBK as having the highest opportunity to raise soil pH using Equation 1.

Model	County	State	Rank
RF	Osceola County	Florida	1
RF	Leon County	Texas	2
RF	White County	Arkansas	3
RF	Lonoke County	Arkansas	4
RF	Highlands County	Florida	5
RF	Arkansas County	Arkansas	6
RF	Polk County	Florida	7
RF	Chicot County	Arkansas	8
RF	Poinsett County	Arkansas	9

RF	Van Zandt County	Texas	10
RF	Houston County	Texas	11
RF	Stoddard County	Missouri	12
RF	Craighead County	Arkansas	13
RF	Cross County	Arkansas	14
RF	Prairie County	Arkansas	15
RF	Woodruff County	Arkansas	16
RF	Jackson County	Arkansas	17
RF	St. Francis County	Arkansas	18
RF	Crittenden County	Arkansas	19
RF	Mississippi County	Arkansas	20
RF	Henderson County	Texas	21
RF	Robeson County	North Carolina	22
RF	Sussex County	Delaware	23
RF	Clay County	Arkansas	24
RF	Barton County	Missouri	25
RF	Phillips County	Arkansas	26
RF	Freestone County	Texas	27
RF	Jefferson County	Arkansas	28
RF	Butte County	California	29
RF	Smith County	Texas	30
RF	DeSoto County	Florida	31
RF	Wayne County	Illinois	32
RF	Lee County	Texas	33
RF	Graves County	Kentucky	34
RF	Vernon County	Missouri	35
RF	Bastrop County	Texas	36
RF	Lee County	Arkansas	37
RF	Hardee County	Florida	38
RF	Madison County	Texas	39
RF	Greene County	Arkansas	40
RF	Bates County	Missouri	41
RF	Desha County	Arkansas	42

RF	Fayette County	Tennessee	43
RF	Anderson County	Texas	44
RF	Cherokee County	Texas	45
RF	Orangeburg County	South Carolina	46
RF	Faulkner County	Arkansas	47
RF	Sampson County	North Carolina	48
RF	Somerset County	Pennsylvania	49
RF	Grimes County	Texas	50

Table SI 4: Top 50 counties identified by RF as having the highest opportunity to raise soil pH using Equation 1.

9.7: Supplemental Information References Cited

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, *5*(1), 180040. https://doi.org/10.1038/sdata.2018.40

Boiko, O., Kagone, S., & Senay, G. (2021). Soil properties dataset in the United States [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9TI3IS8

Center For International Earth Science Information Network-CIESIN-Columbia University. (2016). Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision of UN WPP Country Totals [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4HX19NJ

Edmund Cecil Shorey. (1940). *The Liming of Soils* (Vol. 1845). Retrieved from https://books.google.com/books?hl=en&lr=&id=psPLMyMGwb4C&oi=fnd&pg=PA3&ots=fxBQYW0qHZ&sig=Re-1bMM92OE-HwpljW25WY05sOw#v=onepage&q&f=false

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global Patterns of Groundwater Table Depth. *Science*, *339*(6122), 940–943. https://doi.org/10.1126/science.1229881

Gribov, A., & Krivoruchko, K. (2020). Empirical Bayesian kriging implementation and usage. *Science of The Total Environment*, *722*, 137290. https://doi.org/10.1016/j.scitotenv.2020.137290

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12), 2012GC004370. https://doi.org/10.1029/2012GC004370

Huang, N., Wang, L., Song, X.-P., Black, T. A., Jassal, R. S., Myneni, R. B., et al. (2020). Spatial and temporal variations in global soil respiration and their relationships with climate and land cover. *Science Advances*, *6*(41), eabb8508. https://doi.org/10.1126/sciadv.abb8508

Jafar Yasrebi, Mahboub Saffari, Hamed Fathi, Najafali Karimian, Masome Moazallahi, & Reza Gazni. (2009). Evaluation and Comparison of Ordinary Kriging and Inverse Distance Weighting Methods for Prediction of Spatial Variability of Some Soil Chemical Parameters. *Research Journal of Biological Sciences*, *4*(1), 93–102.

Jon Dewitz. (2023). National Land Cover Database (NLCD) 2021 Products [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9JZ7AO3

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Krivoruchko, K., & Gribov, A. (2014). Pragmatic Bayesian Kriging for Non-Stationary and Moderately Non-Gaussian Data. In E. Pardo-Igúzquiza, C. Guardiola-Albert, J. Heredia, L. Moreno-Merino, J. J. Durán, & J. A. Vargas-Guzmán (Eds.), *Mathematics of Planet Earth* (pp. 61–64). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-32408-6_15

Michigan State University Cooperative Extension Service and Agricultural Experiment Station & U.S. Department of Agriculture Soil Conservation Service. (1981). Soil Association Map of Michigan. Michigan State University Extension Bulletin. Retrieved from https://www.canr.msu.edu/uploads/resources/pdfs/soil_association_map_of_michigan_(e1550).pdf#page=1.00

Mike Walkinshaw, A.T. O'Geen, & D.E. Beaudette. (2022). Soil Properties [Data set]. California Soil Resource Lab. Retrieved from https://casoilresource.lawr.ucdavis.edu/soil-properties/

Reitz, M., Sanford, W. E., Senay, G., & Cazenas, J. G. (2017). Annual estimates of recharge, quick-flow runoff, and ET for the contiguous US using empirical regression equations, 2000-2013 [Data set]. U.S. Geological Survey. https://doi.org/10.5066/F7PN93P0

Scientific Data Curation Team. (2020). Metadata record for: Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.12145791

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., & Scanlon, B. R. (2015). A global data set of the extent of irrigated land from 1900 to 2005. *Hydrology and Earth System Sciences*, *19*(3), 1521–1545. https://doi.org/10.5194/hess-19-1521-2015

Tian, H., Bian, Z., Shi, H., Qin, X., Pan, N., Lu, C., et al. (2022). History of anthropogenic Nitrogen inputs (HaNi) to the terrestrial biosphere: a 5 arcmin resolution annual dataset from 1860 to 2019. *Earth System Science Data*, *14*(10), 4551–4568. https://doi.org/10.5194/essd-14-4551-2022

Tuanmu, M., & Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, *23*(9), 1031–1045. https://doi.org/10.1111/geb.12182

Vergopolan, N., Chaney, N. W., Pan, M., Sheffield, J., Beck, H. E., Ferguson, C. R., et al. (2021). SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US. *Scientific Data*, *8*(1), 264. https://doi.org/10.1038/s41597-021-01050-2

West, T. O., & McBride, A. C. (2005). The contribution of agricultural lime to carbon dioxide emissions in the United States: dissolution, transport, and net emissions. *Agriculture, Ecosystems & Environment*, *108*(2), 145–154. https://doi.org/10.1016/j.agee.2005.01.002

Wieczorek, M. E. (2019). Area and depth-weighted average of soil pH from STATSGO2 for the conterminous United States and District of Columbia [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P95YOSFQ

Wu, Y.-H., Hung, M.-C., & Patton, J. (2013). Assessment and visualization of spatial interpolation of soil pH values in farmland. *Precision Agriculture*, *14*(6), 565–585. https://doi.org/10.1007/s11119-013-9316-7

Xie, Y., Gibbs, H. K., & Lark, T. J. (2021). Landsat-based Irrigation Dataset (LANID): 30 m resolution maps of irrigation distribution, frequency, and change for the US, 1997–2017. *Earth System Science Data*, *13*(12), 5689–5710. https://doi.org/10.5194/essd-13-5689-2021

Zhao, M., & Running, S. W. (2010). Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009. *Science*, *329*(5994), 940–943. https://doi.org/10.1126/science.1192666

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, *95*(2), 164–176. https://doi.org/10.1016/j.rse.2004.12.011