Received: 5 March 2025 | Accepted: 28 April 2025

DOI: 10.1002/ael2.70020

Agricultural & Environmental Letters OPEN ACCESS

RESEARCH LETTER

Think outside the plots: Perimeter measurements and spatial modeling mitigate confounding in a 145-year experiment

Eric Potash[1,2] | **Yuhei Nakayama**[3] | **Michael Douglass**[3] | **Guadalupe Gonzalez**[1] | **Andrew J. Margenot**[1,3]

[1]Agroecosystem Sustainability Center, Institute for Sustainability, Energy, and Environment, University of Illinois Urbana-Champaign, Urbana, Illinois, USA

[2]Department of Natural Resources and Environmental Sciences, College of Agricultural, Consumer and Environmental Sciences, University of Illinois Urbana Champaign, Urbana, Illinois, USA

[3]Department of Crop Sciences, College of Agricultural, Consumer and Environmental Sciences, University of Illinois Urbana Champaign, Urbana, Illinois, USA

Correspondence
Eric Potash and Andrew J. Margenot, Agroecosystem Sustainability Center, Institute for Sustainability, Energy, and Environment, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA.
Email: epotash@illinois.edu and margenot@illinois.edu

Assigned to Associate Editor Tulsi Kharel.

Funding information
National Science Foundation, Grant/Award Number: 2125626; Illinois Soybean Association, Grant/Award Number: 23-10-000664-32-466-37; National Institute of Food and Agriculture, Grant/Award Number: 2022-68013-37052

Abstract

Long-term experiments (LTEs) offer unique insights into the effects of agricultural practices on soil organic carbon (SOC). However, early LTEs commonly lack treatment randomization, replication, and initial measurements of SOC. This creates a potential problem of unmeasured confounding. We address this problem using the Morrow Plots (established 1876) as a case study. We start with a standard mixed effects model of SOC and add (i) a spatial kriging component and (ii) SOC measurements in the sod perimeter of the experiment as an additional treatment level. We find that much of the observed SOC variation between treatments after 145 years is not due to treatments but other factors (e.g. initial SOC), attenuating treatment effects by about 50%. Our study demonstrates that creative measurement and innovative modeling can mitigate some deficiencies in early LTEs. However, our improved estimates still have limited precision, suggesting the importance of careful design and measurement in the first place.

Plain Language Summary

To study the effects of farming practices on soil organic carbon (SOC), modern experiments measure SOC on different pieces of land, randomly assign different practices to them, then remeasure SOC. However, older experiments were not randomized and did not have the technology to measure SOC before the experiments started. This is a problem for estimating the long-term effects of farming practices on SOC. In this study, we show a way around this problem for an experiment started in 1876. We measured SOC in the experiment and in the grass surrounding the experiment and combined them in a spatial statistical model. We found that much of the observed variation in SOC in year 145 in the experiment is not due to the experimental practices but other factors such as SOC levels before the experiment, changing our estimates

Abbreviations: CI, (equal tail) credible interval; LTE, long-term experiment; NPK, nitrogen phosphorus potassium; SOC, soil organic carbon.

of the impacts of the practices on SOC. Our findings highlight the importance of carefully analyzing older experiments as well as adhering to best practices in new experiments.

1 | INTRODUCTION

Early long-term experiments (LTEs) offer a unique opportunity to study the long-term effects of agricultural practices on soil organic carbon (SOC). However, early LTEs tend to lack initial measurements of SOC. In the absence of initial measurements, it is still possible to estimate treatment effects on SOC using randomized experiments. This is because in a (well-designed) randomized experiment, factors such as the distribution of initial SOC are independent of treatment assignment (no confounding). However, randomization was also lacking in early agricultural experiments (Fisher, 1928). Thus, the longest LTEs, established in the 19th century, tend to lack both initial SOC measurements and randomization, posing a challenge for estimating long-term treatment effects on SOC using standard approaches such as mixed effects models.

We address this challenge using the second oldest continuous LTE in the world, the Morrow Plots, established in 1876 in Illinois, USA, to study crop rotation and fertilization treatments on crop productivity and soil fertility. The Morrow Plots lacks both initial SOC measurements and randomization, raising the concern of confounding, specifically a spatial pattern in (unmeasured) initial SOC that could be correlated with treatment assignment. This concern was supported by measurements of SOC concentration (0–15 cm) in the sod (established 1904) on the perimeter of the experiment showing a strong spatial pattern (Darmody & Peck, 2019).

Darmody and Peck (2019) attempted to estimate treatment effects on SOC by assuming that current SOC in the sod perimeter reflected initial SOC in the experiment (established 1876). While innovative in identifying the sod perimeter as an important source of information, this assumption has two limitations. First, the assumption is strong: where there is now sod was cropped for the first 30 years of the experiment and, as the authors acknowledge, sod "is a poor analog of the original tallgrass prairie." Second, the assumption is crude: for each location in the experiment, all SOC measurements in the sod perimeter were considered equally plausible initial SOC levels, regardless of proximity. Additionally, they were limited by their measurements to 0- to 15-cm SOC concentrations, rather than deep SOC stocks.

We use sod perimeter measurements and spatial modeling to estimate treatment effects on deep (~90 cm) SOC stocks in year 145 of the Morrow Plots. Our approach addresses the above limitations as follows. First, we consider the sod perimeter as an additional experimental treatment plot so that we do not need to make the questionable assumption that sod SOC stocks reflect initial SOC stocks. Second, the spatial component of our model, combined with the sod perimeter measurements, allows us to account for the spatial pattern of the SOC measurements and overcome confounding. Our study demonstrates the power of new measurements and statistical methods to mitigate the deficiencies of suboptimal agricultural experiments and increase understanding of management practices on soils.

2 | MATERIALS AND METHODS

2.1 | Experimental design

The Morrow Plots has a split plot design with three crop rotation whole-plots established in 1876 (Caldrone et al., 2024): maize, maize-soybean, and maize-oat-alfalfa (maize: *Zea mays*, soybean: *Glycine max*, oat: *Avena sativa*, alfalfa: *Medicago sativa*). Each rotation whole-plot is divided into eight split-plots (10 m × 10 m) with three fertilization treatments: unamended ($n = 3$), manure ($n = 2$; established 1904), and nitrogen phosphorus potassium (NPK) fertilizer ($n = 3$; established 1955). We note that the rotation whole-plots are not replicated and fertilization split-plots are not randomized within each whole-plot (Figure 1). The perimeter of the experiment is bluegrass sod (*Poa pratensis*; established 1904). Manure and NPK were applied by hand or injected (in the case of liquid manure), minimizing the possibility of spillover into the sod. The experiment is situated on a fine, smectitic, mesic Aquic Argiudoll (Flanagan series), developed on loess parent material.

2.2 | Measurements

In November 2021, soils were sampled by hydraulic probe as intact cores (4.4 cm diameter) to a target depth of 100 cm in the experiment (Figure 1). Cores were sliced in 15-cm sections, weighed, air dried at 24°C, and gently ground by mortar and pestle to pass a 2-mm sieve. A subsample was oven dried at 105°C for 16 h to measure oven-dried bulk density given the known soil volume. SOC concentration was measured by dry combustion. In August 2024, sod 1.5 m outside the perimeter of the experiment (Figure 1) was sampled, processed, and measured using the same methodology. To

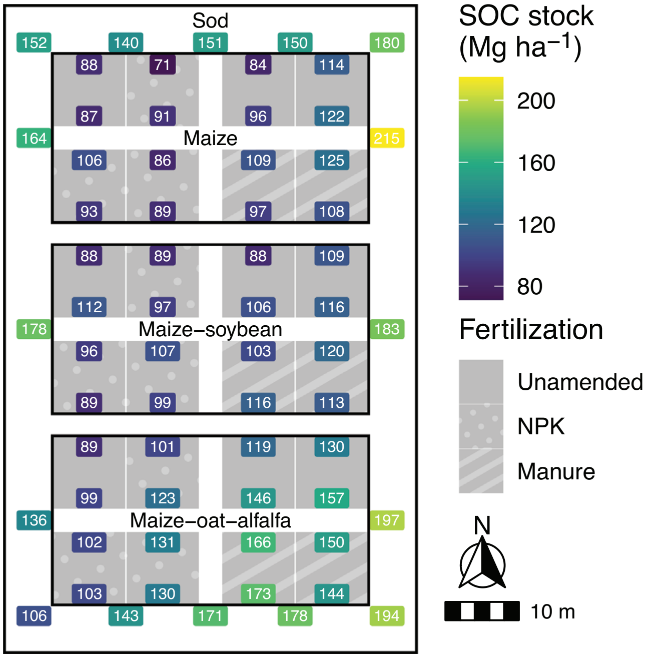

FIGURE 1 Experimental design of the Morrow Plots (established 1876 in east-central Illinois, USA) at the time of sampling in 2021 (year 145). Three crop rotation whole-plots (unreplicated) are subdivided into eight split-plots with three fertilization treatments (non-randomized), for a total of 24 experimental split-plots. Soils were sampled to ~90 cm depth at 48 locations within the experiment ($n = 2$ per split-plot) and at 16 locations in the sod perimeter. Soil organic carbon (SOC) stocks were measured and adjusted to equivalent soil mass.

account for potential differences in bulk density, we expressed SOC stocks on an equivalent soil (mineral) mass basis of 10 Gg ha^{-1} (approximately 90-cm depth) using interpolation by monotonic smoothing splines (von Haden et al., 2020).

2.3 | Treatment effect estimation

To estimate the effect of the treatments (rotation × fertilization) on SOC stocks, we first fit a (log-)linear mixed effects regression model to the measurements from the experiment. We then expanded this model with a spatial component and with measurements from the sod perimeter.

2.3.1 | Unadjusted

We started with a log-linear model of SOC stocks:

$$\log\left(\mathrm{SOC}_{i,Y}\right) = \alpha + \sum_{y=1}^{Y} \theta_{t[i,y],y} + \nu_{j[i]} + \epsilon_i \tag{1}$$

where $\mathrm{SOC}_{i,Y}$ is the measured SOC stock at location i in year Y (i.e. present day), α is the intercept, $t[i, y]$ is the treatment at this location in year y so that $\theta_{t[i,y],y}$ is the coefficient of that treatment in that year, $j[i]$ is the split-plot that the location belongs to so that $\nu_{j[i]}$ is a split-plot effect, and ϵ_i is a residual error. Each treatment t consists of a rotation r and a fertilization f. We decomposed these coefficients into the sum of a rotation coefficient, fertilization coefficient, and rotation–fertilization interaction coefficient:

$$\theta_{t,y} = \theta_{r,y} + \theta_{f,y} + \theta_{r\times f,y}. \tag{2}$$

> **Core Ideas**
>
> - Long-term experiments (LTEs) offer unique insights to agricultural practices.
> - Many of the oldest LTEs are not randomized or replicated and lack initial measurements.
> - Spatial modeling and creative measurement (here sod perimeter) can mitigate these issues.
> - We demonstrate this for soil organic carbon at the Morrow Plots (established 1876).

2.3.2 | Spatial adjustment

One way to account for potential spatial patterns in the regression model is to include row and column effects as well as directional (N-S and E-W) linear trends. We instead opt to use kriging (Gaussian processes) because it is more flexible, accommodating nonlinear patterns. Specifically, we have

$$\log\left(\mathrm{SOC}_{i,Y}\right) = \alpha + \sum_{y=1}^{Y} \theta_{t[i,y],y} + \nu_{j[i]} + \eta_i + \epsilon_i \tag{3}$$

where η_i is a spatial term with exponentiated quadratic variogram (Gelfand et al., 2010).

2.3.3 | Spatial sod adjustment

While the spatial adjustment above is a step in the right direction, the treatment effects are not identified by the experimental data alone due to confounding (see Section 1). To rectify this, we add sod observations in the spatial model, expanding the set of treatment-year effects $\theta_{t,y}$ to include sod post 1904 (when the sod was planted).

2.3.4 | Estimation

Our goal was to estimate the cumulative effect of each treatment over the duration of the experiment. We included

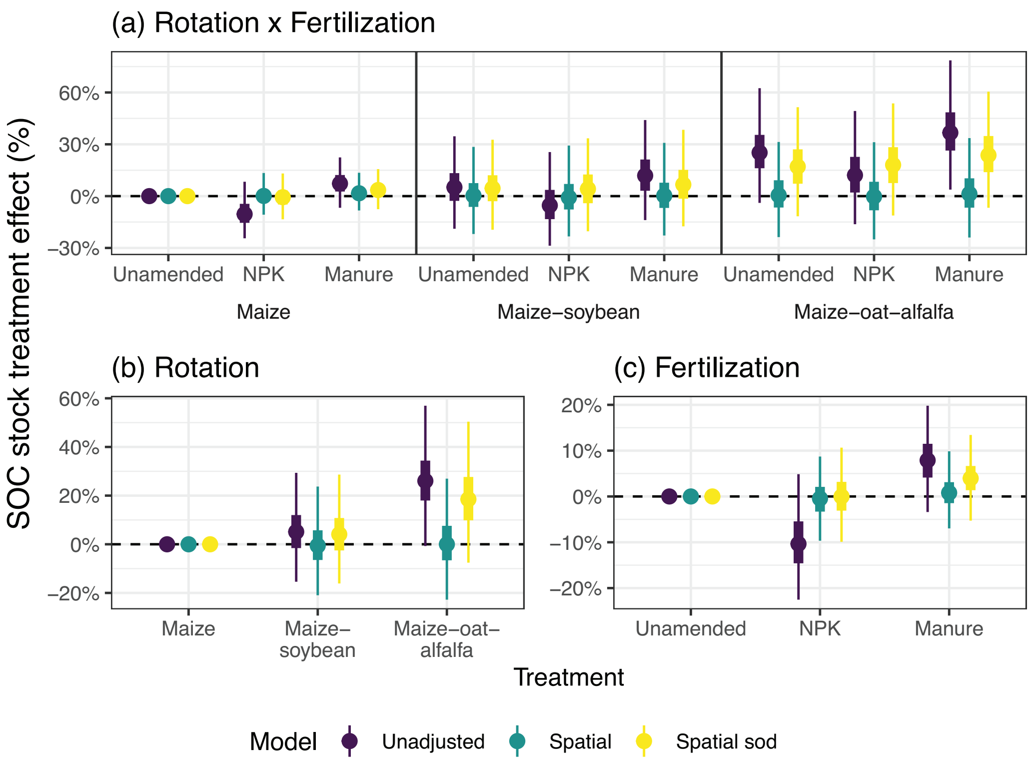

FIGURE 2 Rotation and fertilization treatment effects at the Morrow Plots estimated using three different models: unadjusted, spatial, and spatial sod. (a) Effect of each of the nine treatment pairs compared to the reference of unamended maize. (b and c) Marginal effect of the three levels of each treatment factor, compared to a reference, and averaged across treatments of the other factor (see Section 2). Dots and bars show posterior median, 50%, and 95% intervals. NPK, nitrogen phosphorus potassium.

time-varying effects ☐ to capture changes in treatments that have occurred over the course of the experiment, in particular the introduction of manure (1904) and NPK (1955), as well as sod (1904) which was previously cropped under the adjacent rotation treatment (Aref & Wander, 1997). We used weakly informative priors to estimate these time-varying effects (Supporting Information).

The models were estimated using Markov Chain Monte Carlo (Gelman et al., 2013). Diagnostics for convergence are described in the Supporting Information. We used the probabilistic programming language Stan (Carpenter et al., 2017) via the brms (Bürkner, 2017) interface. Data and R source code is provided.

Treatment effects were extracted from each model using posterior simulation. For each rotation × fertilization treatment, the effect was defined as the relative (%) difference between predicted SOC stock under the treatment compared to the reference treatment of unamended maize. Manure and NPK treatment effects were estimated for 66 years of application since NPK was introduced at the Morrow Plots in 1955. We also computed marginal treatment effects on each factor (rotation or fertilization) by averaging across the other factor. For example, to compute the marginal effect of manure compared to unamended, we averaged three effects: maize manure compared to maize unamended, maize–soybean manure compared to maize–soybean unamended, and maize–oat–alfalfa manure compared to maize–oat–alfalfa unamended.

3 | RESULTS AND DISCUSSION

Measurements of deep (10 Gg mineral soil ha^{-1}; ~90 cm) SOC stocks in the experiment had a median of 106 Mg C ha^{-1} (95% CI 85–165 Mg C ha^{-1}), while in the sod it was 167 Mg C ha^{-1} (95% CI 117–208 Mg C ha^{-1}). Higher SOC on the east and southeast of the sod perimeter, observed by Darmody and Peck (2019) for 0-15 cm SOC concentration, were also seen in our deep SOC stocks (Figure 1).

Using the unadjusted model, we estimate that compared to no amendments, the marginal effect (see Section 2) of NPK on SOC stocks was likely to be large and negative (−10%; 95% CI −23% to 5%), whereas the marginal effect of manure was likely large and positive (8%; 95% CI −4% to 20%). The unadjusted model also estimated a likely large, positive effect of the maize–oat–alfalfa rotation (26%; 95% CI 0%–57%) (Figure 2b,c). Treatment effects under this model are likely to be large, but their precision is low so that the evidence is also consistent with small effects. In addition to the lack of randomization, replication, and initial SOC measurements,

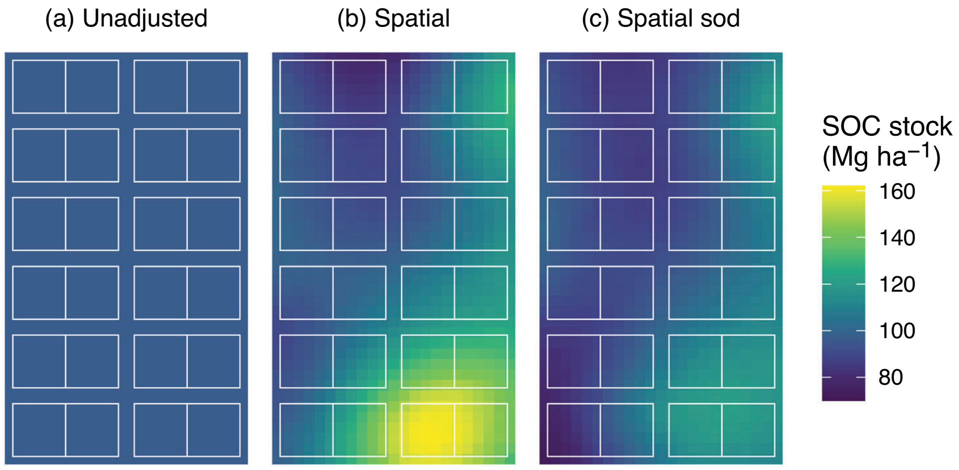

FIGURE 3 Predicted soil organic carbon (SOC) stock maps under the three models considered if Morrow Plots was managed uniformly as unamended continuous maize. Plot boundaries shown for reference. Each point is the posterior predicted median.

this uncertainty is influenced by the plot size and number of measurements.

As hypothesized, the spatial model revealed that treatments were confounded by spatial locations so that the treatment effect estimates are all centered at zero with low precision (Figure 2).

After accounting for spatial variation in SOC stocks in the sod perimeter, treatment effects estimated by our preferred spatial sod model lie between the previous two models (Figure 2). Compared to the unadjusted model, the spatial sod model attenuated treatment effects by about 50% on average. Specifically, we find attenuation of the formerly (i) large negative effect of NPK fertilizer (0%; 95% CI −10% to 11%), (ii) large positive effect of manure (4%; 95% CI −6% to 13%), and (iii) large positive effect of maize-oat-alfalfa (18%; 95% CI −8% to 50%). The precision of our estimates remains low, especially for the maize–oat–alfalfa rotation, which we note had the greatest measured variability in SOC stocks (Figure 1).

We investigated the differences between the models by predicting SOC stock maps under each model. Specifically, we predicted the counterfactual in which the entire experimental area was planted to continuous maize and unamended. The unadjusted model (Figure 3a) showed no spatial pattern because the model does not have a spatial component, and instead attributes all variability in SOC stocks to treatment effects. The spatial model (Figure 3b) showed a significant spatial pattern, overfitting the data and attributing almost all variability to spatial variability. The spatial sod model (Figure 3c) showed a similar spatial pattern to the spatial model (e.g., hotspots in the SE and NE corners) but much smaller in magnitude. The spatial sod model balanced attributing variability in the data between spatial variability and treatment effects.

4 | CONCLUSION

We used perimeter measurements and spatial modeling to mitigate challenges (lack of randomization, replication, and initial measurements) in estimating long-term treatment effects on deep (~90 cm) SOC stocks in year 145 of the Morrow Plots. More generally, this case study demonstrates that creative measurement and careful modeling can mitigate some design and measurement issues common to early LTEs. However, the benefits of new measurements and modeling may be limited, as evidenced by the relatively low precision of our estimates. Thus, we emphasize the importance of careful analysis of early LTEs as well as adhering to best practices in establishing new experiments.

AUTHOR CONTRIBUTIONS

Eric Potash: Conceptualization; formal analysis; methodology; software; writing—original draft. **Yuhei Nakayama**: Data curation; investigation; writing—review and editing. **Michael Douglass**: Investigation. **Guadalupe Gonzalez**: Investigation. **Andrew J. Margenot**: Resources; supervision; writing—review and editing.

ACKNOWLEDGMENTS

This work was supported in part by USDA NIFA award #2022-68013-37052 (Andrew J. Margenot), NSF award #2125626 (Andrew J. Margenot), and Illinois Soybean Association award #23-10-000664-32-466-37 (Andrew J. Margenot).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data and code used to support the findings of this study are available in a figshare repository at https://doi.org/10.6084/m9.figshare.28950647.

ORCID

Eric Potash https://orcid.org/0000-0003-4658-9802
Yuhei Nakayama https://orcid.org/0000-0002-7153-6545
Andrew J. Margenot https://orcid.org/0000-0003-0185-8650

REFERENCES

Aref, S., & Wander, M. M. (1997). Long-term trends of corn yield and soil organic matter in different crop sequences and soil fertility treatments on the Morrow Plots. *Advances in Agronomy*, *62*, 153–197.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Caldrone, S. L., Margenot, A. J., & Morrow Plots Data Curation Working Group. (2024). From complex histories to cohesive data, a long-term agricultural dataset from the Morrow Plots. *Scientific Data*, *11*(1), 1145. https://doi.org/10.1038/s41597-024-03984-9

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1. https://doi.org/10.18637/jss.v076.i01

Darmody, R. G., & Peck, T. R. (2019). Soil organic carbon changes through time at the University of Illinois Morrow Plots. *Soil organic matter in temperate agroecosystems long term experiments in North America* (pp. 161–169). CRC Press.

Fisher, R. A. (1928). *Statistical methods for research workers*. Oliver and Boyd.

Gelfand, A. E., Diggle, P. J., Fuentes, M., & Guttorp, P. (2010). *Handbook of spatial statistics*. CRC Press. https://doi.org/10.1201/9781420072884

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Von Haden, A. C., Yang, W. H., & DeLucia, E. H. (2020). Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon and other mineral soil properties. *Global Change Biology*, *26*(7), 3759–3770. https://doi.org/10.1111/gcb.15124

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Potash, E., Nakayama, Y., Douglass, M., Gonzalez, G., & Margenot, A. J. (2025). Think outside the plots: Perimeter measurements and spatial modeling mitigate confounding in a 145-year experiment. *Agricultural & Environmental Letters*, *10*, e70020. https://doi.org/10.1002/ael2.70020