The Influence of Inherent Soil Factors And Agricultural Management On Soil Organic Matter

Gregory S. Richardson

University of Wisconsin-Madison

Matthew D. Ruark (✉ mdruark@wisc.edu)

University of Wisconsin-Madison

Timothy Radatz

University of Wisconsin-Madison

Amber Radatz

University of Wisconsin-Madison

Eric Cooley

University of Wisconsin-Madison

Erin M. Silva

University of Wisconsin-Madison

Abigail J Augarten

University of Wisconsin-Madison

Jun Zhu

University of Wisconsin-Madison

Chelsea H. Zegler

University of Wisconsin-Madison

Research Article

Keywords: soil organic matter, total organic carbon, total nitrogen, soil texture, drainage class, tillage, manure

Posted Date: August 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-784637/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The accumulation of soil organic matter (SOM) is vital to the agronomic and environmental functioning of agroecosystems, yet the relative influence of inherent soil properties and agricultural management practices on SOM dynamics are not often addressed in individual studies. Using a network of 218 operating farm fields across Wisconsin and southern Minnesota, USA, this research employs single variable analysis (ANOVA and regression) and regression tree analysis to assess the effects of soil properties (texture, drainage class, pH) and management variables related to crop rotation, tillage, cover cropping, and manure application on SOM, as well as total organic carbon (TOC) and total nitrogen (TN) in the upper 15 cm. Single variable analysis revealed that greater SOM, TOC, and TN were associated with poorly drained soil, tile drained fields, high-clay content soil, and high biomass crop rotations. Soil organic matter (SOM) and TOC were strongly related ($R^2=0.71$), but different regression trees were produced; SOM was most influenced by clay content, while TOC was most influenced by drainage class. Future benchmarking of SOM should be conducted with drainage and texture class categories. The rapid building of data sets thorough unstructured sampling, including an abundance of meta-data, should be a research priority in agricultural science.

Introduction

Soil organic matter (SOM), as well as the total carbon (C) and nitrogen (N) within the SOM, is vital to the agronomic and ecological performance of agroecosystems¹. Among its many functions, SOM serves as the primary source of nutrients for most plants and microorganisms as well as a medium for C sequestration. Soil organic matter formation occurs as plant, animal, and microbial residues decompose into a heterogeneous mix of C-based macromolecules². Once stabilized after initial rapid loss of carbon as CO₂, the remaining SOM cycles at different temporal intervals with the majority remaining in the soil for tens to hundreds and sometimes thousands of years^{3,4}. Consequently, it may take decades to quantify changes in SOM depending on the change of management practice that is implemented⁵.

While major initiatives have been established to increase SOM on agricultural lands – e.g., France's 4 per 1000 Initiative (<https://www.4p1000.org/>) and California's Healthy Soils Initiative (<https://www.cdfa.ca.gov/healthysoils/>) – several foundational assumptions about the mechanisms that control SOM accumulation and loss are still debated. In particular, researchers are still addressing how quality of litter inputs⁶, tillage intensity⁷, and cover cropping⁸ affect SOM levels. As technical understanding about SOM evolves, there is increasing need for empirical inquiry into what soil and management factors are associated with greater amounts of SOM, and in relation with inherent soil properties such as drainage class, texture, and pH.

To study soil dynamics, most agricultural studies use factorial designs that measure the effects of a few variables at a plot-scale and across relatively short time frames⁹. These experiments are critiqued as inadequate representations of true agricultural systems due to their reductionist nature, vulnerability to

edge effects, and inability to capture long-term processes¹⁰. Long-term systems experiments overcome many of these issues¹¹, yet the number of these studies is lacking and this restricts the geographic relevance of their findings. Establishing new long-term systems experiments would overcome this barrier, but the operational costs of these experiments can be barriers to initiation^{12,13}. An alternative approach is participatory research utilizing existing farmer networks which allows for data collection from farm fields whose unique management histories can be studied in aggregate to identify trends. A benefit of this research study is its potential to achieve broader geographic scope while obtaining results more quickly and at lower expense than establishing new experimental plots. As such, the number of multi-site studies is expected to increase significantly in the near future¹³. A downside to this approach is the inherent difficulty of holding variables fixed across sites, thus requiring inclusion of a large number of independent (and potentially confounding) variables, related to both management and soil characteristics.

Machine learning techniques can be applied to multi-site studies on operational farms since the associated methods for handling large datasets have become more sophisticated¹⁴. In particular, regression tree analysis offers several advantages in this context even though this technique is not often utilized in agricultural research. Regression trees can simultaneously handle complex relationships with categorical and numerical independent variables, are robust against outliers, and can handle skewed data¹⁵. They also allow for missing values by identifying closely matching surrogates from the dataset or merely skipping the missing data without discarding the entire observation. The results, which reflect nonlinear regression by a binary decision-making process, are also flexible and relatively easy to interpret by decision makers.

The primary objective of this research is to use single-factor ANOVA and regression approaches along with a simple machine learning approach (regression tree) to identify the relative effects of inherent soil properties and management practices on SOM, total organic carbon (TOC), and total nitrogen (TN). The exploratory approach used here is may be a valuable supplement to long-term research plots. Here we are most interested in if any single inherent or management factor explains most of the variation and how robust of a regression tree can be produced with such data. Inherent soil properties were limited to pH, drainage class, and soil texture. Management practices of interest include crop rotation, cover cropping, manure application, and tillage. We also evaluate soil compare measures of SOM, TOC, and TN.

Results

Field management. Among the fields sampled (Fig. 1), 72% received manure applications within the past 5 years, with 70% of those applications being dairy manure, and 57% of the fields received a manure application within 9 months prior to sampling. Almost half of the fields (49%) were conventionally tilled, with 69% of tilled fields receiving two or more tillage passes since the previous year's harvest and 37% of the fields were under no-till management for at least 4 years. Cover crops were planted at least once in the past 5 years in 31% of the fields historically, and 20% of the fields were planted with cover crops leading into the season when soil samples were taken. Planting cover crops was a new practice for most

of the farmers, as the fields with cover crops had typically only grown them for 1 (44%) or 2 (31%) years. The median fertilizer N rate was 130 kg ha⁻¹, whereas the median manure N application was 47 kg ha⁻¹ (Supplementary Table 5). When factoring in N-credits from alfalfa (*Medicago sativa* L.), the median total N input was 193 kg ha⁻¹.

Soil organic matter, total organic C, and total N. Median values for SOM, TOC, and TN were 27, 19.0, and 1.9 g kg⁻¹, respectively (Table 1). Total organic C (TOC) demonstrated slightly greater variation (CV), skewness, and kurtosis compared to SOM. The relatively large CV values for all variables indicate that a sufficiently wide range of values was obtained in this dataset. Soil organic matter (SOM), TOC, and TN were all significantly and positively related to one another, with SOM and TOC demonstrating the strongest relationship ($R^2 = 0.71$). However, the slope of the relationship indicates a very high TOC to SOM ratio, much greater than typically reported in the literature¹⁶. Visual interpretation of Fig. 2a suggests the relationship is heavily influenced by the high SOM and TOC values. Reanalysis of data using only SOM values less than or equal to 50 g kg⁻¹ (a removal of nine data points) reveals a weaker relationship ($R^2 = 0.55$) but a more appropriate TOC to SOM ratio (0.56). The SOM to TN relationship was weaker ($R^2 = 0.35$), but less influenced by SOM values greater than 50 g kg⁻¹ (Fig. 2b). TOC and TN also had a relatively weak relationship ($R^2 = 0.33$) (Fig. 2c).

Table 1
Univariate statistics for soil organic matter (SOM), total organic carbon (TOC), and total nitrogen (TN).

	SOM	TOC	TN
	g kg ⁻¹		
n	218	216	216
Minimum	7.0	4.0	0.2
First quartile	23	15.6	1.7
Mean	28	20.1	2.0
Median	27	19.0	1.9
Third quartile	32	23.0	2.2
Maximum	68	24.5	4.5
Standard Deviation	10	7.9	0.6
CV	35	39	32
Skewness	1.2	1.6	1.1
Kurtosis	2.6	3.9	2.8

When analyzed separately, all continuous variables produced a statistically significant relationship with at least one response variable, with the exception of total N input, but all R^2 values were less than 0.16 (Table 2). Clay and pH were positively associated with SOM, TOC, and TN, and sand was negatively associated with SOM (Table 2). Manure N input had a very weak positive relationship with TOC and fertilizer N input had a very weak negative relationship with TOC.

Table 2

Linear regression results [direction of slope (positive or negative) and coefficient of determination (R^2)] between independent continuous variables and soil organic matter (SOM) (n = 218), total organic carbon (TOC) (n = 216), or total nitrogen (TN) (n = 216) (ns = not significant, alpha = 0.05).

Predictor variable	SOM			TOC			TN		
	P-value	Slope	R^2	P-value	Slope	R^2	P-value	Slope	R^2
Clay	< 0.001	+	0.162	< 0.001	+	0.092	< 0.001	+	0.081
Sand	0.001	-	0.069	ns			ns		
pH	0.004	+	0.059	< 0.001	+	0.104	0.003	+	0.040
Manure N input	ns			0.026	+	0.024	ns		
Fertilizer N input	ns			0.006	-	0.036	ns		
Total N input	ns			ns			ns		

Sampling region had a significant effect on SOM, TOC, and TN, while sampling year did not. Southeast Minnesota had greater SOM, TOC, and TN values than all other regions (Supplementary Table 8). Dry Run and Elk Creek had significantly lower SOM than Jersey Valley, southeastern Wisconsin, and northeast Wisconsin and had significantly lower TOC than northeast Wisconsin and southeast Wisconsin. The later sampling time (late-June) had greater SOM and TOC values (but not TN values) than the mid-April sampling time. Drainage class, texture class, and tile drainage all had significant effects on SOM, TOC, and TN. Very poorly drained and poorly drained soils had greater SOM, TOC, and TN values than other drainage classes, with very poorly drained soils having greater TOC and TN values compared to poorly drained (Fig. 3). Tile drained soils had greater SOM, TOC, and TN values compared to undrained. Soils with greater clay content (clay loam and silty clay loam) typically had greater SOM, TOC, and TN values compared to sandier soils (sand, fine sandy loam, sandy loam). Silt loam textured soils (n = 159) had lower SOM compared to silty clay loam and clay loam textures, but greater values than sandy soil. However, silt loam textured soils were not different from sandy soils for TN (Fig. 4).

Crop rotation had a significant effect on SOM, TOC, and TN, while previous crop did not (Supplementary Tables 7 and 8). Continuous corn (*Zea mays* L.) rotations had greater SOM than corn-soybean and corn-small grain rotations and corn-alfalfa rotations had greater SOM than corn-small grain rotations. For

TOC, only continuous corn rotations were greater than corn-small grain rotations and for TN corn-small grain rotations were significantly less than the other three. The number of years a cover crop had been used on the past 5 years had a significant effect on SOM and TOC, but the presence of a cover crop in the year prior to sampling did not (Supplementary Tables 7 and 8), although there were not consistent trends of more frequent cover crop use being associated with greater SOM or TOC. The number of tillage passes in the past year and tillage type had significant effects on SOM and TOC (Supplementary Tables 7 and 8). Two or three tillage passes had greater SOM and TOC compared to no tillage practices. Although tillage was not universally different than no tillage, as one and four tillage passes were not different from no-till. Evaluation on tillage type revealed that conventionally tilled fields had greater SOM and TOC compared to both minimum till and no-till fields. In contrast, the number of manure applications in the past 5 years did not have a significant effect on SOM, TOC, or TN.

Regression tree analysis. Regression tree models were developed for SOM (n = 218), TOC (n = 216), and TN (n = 216) using six continuous and 14 categorical variables. The regression tree for SOM had more splits (Fig. 5) than TOC (Fig. 6) and TN (Fig. 7) and had an R^2 of 0.50. For SOM, the primary split was clay content, with greater SOM associated with values of 26% or more (which represented 6% of the data). For soils with clay content less than 26%, soils with continuous corn or corn-alfalfa based rotations (56% of the data) had greater SOM compared to other rotations (39% of the data). Within continuous corn and corn-alfalfa based rotations, there was an additional split, with very poorly drained soils having greater SOM (but this only included two data points). Within corn-soybean and corn-small grain rotations, the texture classes with the greatest sand and clay contents in our study (sand, loamy sand, and clay loam) had less SOM compared to other texture classes. Clay percentage, soil texture class, and drainage class were the top three variables for the primary split. The improvement value, defined as the percentage of improvement in deviance given by the split using a particular covariate, of soil texture class and drainage class was 94 and 93% of the tree improvement compared to clay percentage (Table 3).

Table 3
Ranking of primary splits for soil organic matter (SOM), total organic carbon (TOC), and total nitrogen (TN) regression trees by improvement value.

SOM		TOC		TN	
Variable	Improvement	Variable	Improvement	Variable	Improvement
Clay	0.326	Drainage	0.289	Clay	0.163
Soil type	0.305	Clay	0.255	Soil type	0.162
Drainage	0.305	Soil type	0.254	Drainage	0.127
Region	0.191	Region	0.157	Region	0.108
Sand	0.097	Tillage type	0.095	Sand	0.042

The regression tree for TOC had an R^2 of 0.28 and only one split; soils with very poorly and poorly drained drainage class (representing 8% of the total data) had greater TOC compared to all others. The

improvement value for clay content and soil texture was 88% of the improvement value of drainage class. The regression tree for TN had an R^2 of 0.16 and only split. Soils with clay content of 26% or greater had greater TN values, but only represented 6% of the data. Soil texture class was very close to the top split, providing 99% of the improvement as clay content. For all trees, no other variables provided more than 58% of the improvement as the primary splits.

Discussion

Soil texture was the most influential variable in determining SOM and TN values. Both clay content and soil texture class were the top two variables for the primary split for SOM and TN and the second and third most important splits for TOC, with clay having the strongest single variable regression relationship with SOM ($R^2 = 0.162$). Soils with greater silt and clay content are known to accumulate and stabilize more organic matter than coarse soils^{17,18}. The clay particles not only form organo-mineral complexes whose strong bonds protect SOM from decomposition, but the clay can also improve soil structure in ways that physically protect SOM through aggregation¹⁹. Previous studies in the Midwest have confirmed this positive relationship between clay content and SOM^{20–22}. This dynamic is not universally true, however, as no significant relationship was found between TN or TC and clay content in a global meta-analysis²³, and an inverse relationship was found in Washington on a single farm due to a particular combination of topographic and climatic conditions²⁴.

Soils from fields with poorly drained soils had greater SOM, TOC, and TN. This is supported by well-recognized soil morphological processes, whereby wetter soils accumulate more C over time due to reduced rates of microbial metabolism and corresponding organic matter decomposition. In our data set, poorly drained soils were also more likely to be high in clay content, thus confounding the influence of clay versus drainage on soil C and N values, although it was only the two mostly poorly drained soils that appeared to be the most influential. The average clay contents for very poorly drained and poorly drained soil were 27.6 and 25.6, respectively, which was similar to the primary split of 26 for SOM (Fig. 5). All other drainage classes had an average clay content below that threshold, and had a relatively equal distribution of texture classes. Related, the fields with tile drainage also had higher SOM and TOC levels. Tile is more regularly added to poorly drained soils, such that the slow changing SOM and TOC pools may remain higher than comparable soils that have had better natural drainage for extended periods of time. Fields with tile drainage also had greater SOM compared to fields without, which could either reflect the fact that fields with artificial drainage are typically on poorly drained soils or that drainage can lead to improved yields and thus more C return to the soil.

Soils sampled from fields in southern Minnesota generally contained higher levels of C and N. Given that Minnesota and Wisconsin cropping systems are similar, this is likely the result of the different soil forming factors and associated soil properties found in Minnesota. Southern Minnesota soils are recently glaciated Mollisols formed under tall grass prairies. In contrast, Elk Creek had the lowest values for TOC and SOM values. This region exists in the Driftless Region in SW Wisconsin, which has not had recent

glaciation and is known for having significant and historical topsoil loss due to its hilly terrain²⁵. Since drainage and texture, the inherent soil properties evaluated in this study, had the largest effect on SOM, TOC, and TN, it is clear that these categories should be considered when assessing the potential for C sequestration or as covariates in regional benchmarking.

Crop rotations with greater biomass production in our study (i.e., continuous corn and rotations with alfalfa) resulted in higher soil C and N levels. The findings are particularly evident in the case of continuous corn and is seen repeatedly throughout the literature in conventional cropping systems. Over a 14-year fertilizer rate study in Iowa, for example, the soil C levels under continuous corn increased when fertilized at the agronomically optimal rate, whereas they generally declined in a corn-soy rotation²⁶. The authors attribute this to 22% more crop residue production in the corn system. Similarly, a meta-analysis found that introducing cover crops into continuous soybean rotations increased TC by 11%, but no effect was found in continuous corn rotations²³ – presumably because the corn was the most significant contributor of C inputs. At the Wisconsin Integrated Cropping Systems Trial (WICST), the lower TC and TN found in a corn-soybean-wheat rotation (managed organically) was also credited to a 25% lower estimated annual C input, compared to continuous corn²⁷. The opposite result was found on sandy soil in Michigan, where corn-soy-wheat rotations had higher SOC values than continuous corn across three separate treatments²⁸; plant litter quality has been found to play a larger role in SOM accumulation in sandier soils²⁹. In our study, we can not rule out that farmers may grow continuous corn primarily on fields with high SOM. Continuous corn yields have been shown to be reduced compared to rotated corn yields^{30,31} and SOM and corn yields have been shown to positively associated³², thus it is not unreasonable to conclude that farmers may choose to conduct continuous corn rotation on high producing land, confounding the causal effect.

Other studies have also found that crop rotations containing perennial legumes (i.e., alfalfa) increase soil organic C^{33,34} and TN and TC values²⁷. The mechanism by which alfalfa enhances organic C and N can result from the direct addition of alfalfa residue (mostly as below ground biomass), reduced tillage during the alfalfa phase, enhanced biomass of subsequent crops that benefit from the N fixed by alfalfa, and through year-round ground cover that protects against soil erosion³⁵. In general, perennial crops tend to boost soil C levels relative to annual cropping systems³⁶. Our results corroborate those findings, but alfalfa production and manure application cannot be completely disentangled, as both are prevalent on dairy farms (92% of fields with alfalfa in rotation also had manure applied at least once in the past five years).

No clear effect of cover cropping occurred in this study. This is not necessarily surprising given that only 31% of the fields had cover crops planted at least once in the past 5 years and only 12% of fields had cover crops planted more than once in the past 5 years. A global meta-analysis found that cover crops can lead to consistent but slow increases in soil C over time ($0.32 \text{ Mg ha}^{-1} \text{ yr}^{-1}$ on average and up to $3 \text{ Mg ha}^{-1} \text{ yr}^{-1}$)⁸, but local studies have shown that it will require continual cover crop use for a decade

before changes in SOM can be detected^{37,38}. Additionally, crop rotation and residue management are likely to be more influential on soil C budgets compared to cover crops in the northern US³⁹.

Variables related to manure application had only a minor influence on SOM. Manure applications can measurably increase total soil C and N levels within a decade or two^{40–42}, which is within the timeframes that most fields in this study had been receiving manure. A global meta-analysis on the effects of manure on C stocks highlights that there is often a linear relationship between manure application rate and increases in soil C and N levels⁴³. Nevertheless, the type of manure (e.g., turkey vs. beef), consistency (e.g., liquid vs. solids), timing (winter vs. spring) and application method (e.g., broadcast vs. injected) all affect SOM mineralization and stabilization dynamics. The variation that exists under the concept of manure application and interactions with soil variables is an area that requires further investigation. Long-term manure input information was difficult to obtain from farmer records, limiting our ability to understand the true long-term effect application.

Tillage practices were not featured in any of the regression tree models. Despite widespread findings in the literature that conversion to no-till production increases topsoil SOM, TOC and TN^{21,44–46}, the conventionally tilled fields had higher TOC and SOM values in this study. This may be a relic of site-specific conditions rather than management, however. In the Driftless region of Wisconsin, where 129 (59%) of the fields in this study are located and where 67% of the no-till fields were located, farmers cultivate steeper slopes that are susceptible to erosion. In contrast, the conventionally tilled fields occur more regularly on less sloping land. Additionally, more pronounced increases in soil C and N would likely have been found within long-term no-till fields if soils were sampled from the top 5 cm of the profile versus the top 15 cm^{44,46}.

This research highlights that soil properties unrelated to management, such as drainage class, region, and soil texture, can have significant and lasting effects on SOM, TOC, and TN in cropland. The importance of these variables would not have been discovered using single site research, highlighting the value of regionally focused research that utilizes a multitude of fields in providing valuable agroecological insights. Regional experiments also offer certain advantages over global meta-analyses, which draw from multiple fields but are limited in the number of variables that can be studied simultaneously and from which the findings are more difficult to translate back into local contexts, and ultimately can help compliment meta-analysis studies. Crop rotation was an influential variable explaining differences in SOM, with greater levels associated with crop rotations that had greater biomass return to soil (continuous corn and corn-alfalfa). Tillage practices resulted in opposite effects than expected, with no-till practices associated with lower SOM values; further investigation revealed most no-till sites were associated with regions with highly erodible landscapes. Manure and cover crop use had little to no effect on SOM in this study.

Although TOC is considered a better quantitative measurement compared to SOM⁴⁷, the widespread adoption and relative ease of performing the loss on ignition method for SOM gives it practical advantages for future research. Regression tree analysis demonstrated that SOM produced a more robust

tree with a greater R^2 compared to TOC. From a data acquisition stand point, this is a positive finding as SOM is more likely to be reported in the literature and collected by farmers, making it more easily obtained and cultivated into large data sets. The regression tree models were valuable tools through which to analyze the dataset. The ability to simultaneously evaluate the potentially complex effect of both continuous and categorical variables on a response variable was particularly informative. Attempting to replicate this study using controlled field experiments would be prohibitive due to the sheer number of treatments needed implement a complete factorial design. In this vein, regression tree models allow for more flexible modeling that can more easily incorporate on-farm sampling. While unstructured sampling will not be a replacement for long-term experiments as it does not incorporate the variable of time, it can be used concurrently to expedite our regional understanding of SOM and guide soil management decision making. Thus, it is clear that the rapid building of data sets thorough unstructured sampling, and likely through crowd-sourced approaches, that include an abundance of meta-data, should be a research priority in soil science.

Materials & Methods

Field descriptions and sampling. Soil samples were collected from 218 farm fields across Wisconsin (n = 212) and Minnesota (n = 6) (Fig. 1) between 2015 and 2017. The fields represent a range of cropping systems common in the Upper Midwest. Six distinct regions were sampled and identified by either general region of a state (northeast Wisconsin, southeast Wisconsin, and southern Minnesota) or by watershed (Dry Run, Elk Creek, Jersey Valley) (Fig. 1). All fields were planted into corn the season soil samples were collected. In each field, three composite soil samples were collected that consisted of five 0- to 15-cm soil cores collected with a probe of 2.5- or 7.5-cm internal diameter. Most soil samples (194) were collected prior to fertilizer application and corn planting (mid-April); 24 samples were in late-June (2017 only). Soil sampling was conducted with an area of 36 m² within the dominant soil map unit as identified by the USDA NRCS Web Soil Survey⁴⁸ and from an area identified by the farmer where average crop yields were obtained. The composite samples were stored cold and transferred into a freezer with 1 to 6 hours of sampling to stagnate microbial metabolism and organic matter mineralization. Within 30 days, soil samples were thawed and dried for 1 week at 32°C in a forced-air drier, ground to pass through a 2-mm sieve, and stored at room temperature until analysis.

Inherent soil properties such as texture class, sand and clay content of the surface horizon, and drainage class were obtained from the USDA NRCS Web Soil Survey⁴⁸. Agronomic management information regarding crop rotation, tillage practices, cover crop use, tile drainage, and manure and fertilizer applications were obtained directly from each farmer. Based on the collected data, four categories for crop rotation (continuous corn, corn-soybean, corn with small grain, and corn with alfalfa) and five categories for previous crop were created (Supplementary Table 2). Two categorical data were developed for cover crops: if there was a cover crop planted last fall (yes or no) and the number of times a cover crop was planted in the past 5 years. Tillage practices were categorized by practice [no tillage, minimum tillage (including vertical tillage or strip tillage), and conventional tillage (chisel, disk or moldboard)] and

by the number of tillage passes that occurred between harvest of the previous year's crop and the planting of the current year's crop (0 to 4). Tillage was only considered no-till or minimum tillage if practiced for more than 4 years. Manure was categorized based on the number of manure applications that occurred in the past 5 years (0 to 5) and if when manure was applied in the past year (none, summer, fall, winter, or spring). Tile drainage presence was also noted (yes or no). The manure N, fertilizer N, and total N input (which includes manure, fertilizer, and legume N inputs) (kg ha^{-1}) to the previous corn crop were also collected. If farmers did not have manure analysis, estimates of available N were used⁴⁹; N input from alfalfa biomass was assumed to be 193 kg ha^{-1} ⁴⁹.

Soil analysis. Soil pH and SOM were analyzed by the University of Wisconsin Soil and Forage Analysis Laboratory (Marshfield, Wisconsin). Soil pH was calculated using a 1:1 slurry of 10 g soil and 10 mL of deionized water and measured with a glass electrode⁵⁰. Soil organic matter values were determined through loss on ignition by heating the soil to 360°C for 2 hours⁵¹. Total C (TC) and total N (TN) levels were determined via the dry combustion method using a Flash EA 1112CN Automatic Elemental Analyzer (Thermo Finnigan, Milan, Italy). Between 8 to 10 mg of finely ground soil were packed into a 5 mm by 9 mm tin capsule prior to combustion at temperatures exceeding 1000°C . Soils with pH greater than 7.0 were tested for effervescence using 5% HCl as an indicator if carbonates were present. If carbonates were not observed, TC was assumed to be TOC; if carbonates were observed, they were subject to acid-fumigation prior to dry combustion⁵². Only 25 samples were analyzed for carbonates and 13 of those had carbonate concentrations above the detection limit. There were 218 samples for SOM, but only 2016 for TOC and TN because two samples were accidentally discarded.

Statistical analysis. Descriptive statistics were used to examine the distribution of the data. For continuous variables, univariate summary statistics (e.g., mean, median, skewness, etc.) were calculated (Table 1) using the "describe" function of the psych package in R (3.6.1) on an Rstudio platform^{53,54}. For the 14 categorical variables in the dataset (Supplementary tables 1 and 2), analysis of variance (ANOVA) was performed using Proc GLM (SAS 9.4)⁵⁵ to determine individual effects on SOM, TOC, and TN. Normality assumptions were evaluated for all dependent variables using QQ-plots, histograms, and the Shapiro-Wilk Normality test⁵⁶. The right skew of the data for SOM, TOC and TN resulted in non-normal distributions, but standard transformations were unable to satisfy the normality requirements of the Shapiro-Wilk test. As such, these data were not transformed. When statistically significant differences were found ($p \leq 0.05$), LSMEANS was used to determine mean differences. For the six continuous variables in the data set (Table 2), relationships with SOM, TOC, or TN were determined with linear regression using SAS Proc REG.

Four regression trees were generated using the rpart package in R⁵⁷ for SOM, TOC, and TN. Starting with the complete dataset, the model identifies a point within a single independent variable (e.g., soil pH above or below 6.5) that divides the data into two mutually exclusive "branches." The "leaf" or node at the end of each branch represents its own subset of the data. This division process is performed independently on each new leaf, continuing in a top-down manner and splitting the dataset recursively into new

branches and leaves until a given termination criterion is reached. Each division minimizes the heterogeneity of the response variable values (e.g., SOM) in each subset, relative to other possible splits. The model is deemed “greedy” in that it does this irrespective of how the outcome influences splits further along the process.

The termination criteria used in this study were initially: a minimum of five samples per leaf; a minimum complexity parameter (cp) value of 0.0001; and up to eight levels (i.e., max depth) of branching. Ninety percent of the data points were designated as training data with which to develop the model and the other 10% were used to test its predictive accuracy. This cross-validation process provides a basis from which to prune back the tree and is automatically conducted 10 times by the rpart package to identify the best predictive model. Pruning the regression tree helps to prevent against bias-variance tradeoffs, whereby the model overfits the training data thereby reducing its predictive accuracy and increasing its sensitivity to changes in the dataset – i.e., small changes would have large effects on the composition of the tree⁵⁸. The cp-value reflects the degree by which additional splits improve the relative accuracy of the model. As a rule of thumb, it is best to prune back to the lowest level in the tree where the sum of the relative error and x-standard deviation remain lower than the x-error (i.e., the average error for the 10% of cross validation data)⁶⁰. For each response variable, the cross-validation process was manually repeated 15 times. While the structure of the tree rarely changes during each run, because the training and testing data are randomly selected the cp-values and corresponding pruning decisions can vary. The most commonly produced cp-value was selected as the point at which to prune each regression tree. Resulting trees are presented as are the top five variables for the primary split.

Declarations

Acknowledgements

Funding for the research was provided by USDA NRCS Conservation Innovation Grant number 69-3A75-14-270 and North Central SARE Graduate Student Project GNC17-249, as well as from funds provided by Dairy Farmers of Wisconsin. The project was conducted in collaboration with the UW Discovery Farms staff and the Wisconsin Department of Agriculture, Trade, and Consumer Protection’s Producer Led Watershed Program. Greg Richardson was funded by the University of Wisconsin-Madison’s Graduate School Fellowship.

Author contributions (following CRediT descriptors)

Conceptualization: MR, AR, TR, EC; Data curation: AA; Formal analysis: GR, MR; Funding acquisition: AR, EC, GR, ES; Investigation: GR, AA; Methodology: GR, M; Project administration: MR, AR ; Resources: AR, EC; Software: GR, CZ, JZ; Supervision: MR, ES; Validation: MR; Visualization: GR, CZ; Writing – original draft: GR, MR; Writing – review & editing: TR, AR, EC, JZ, ES

Competing interests

The authors declare no competing interests

References

1. Magdoff, F. & Weil, R. R. (ed.). *Soil Organic Matter in Sustainable Agriculture*. CRC Press, New York. p. 10-28 (2004).
2. Horwath, W. R. Carbon cycling: the dynamics and formation of organic matter. p. 339–359. *In* E. A. Paul (ed.) *Soil Microbiology, Ecology and Biochemistry* (4th ed.). Academic Press, San Diego, California (2015).
3. Hsieh, Y. P. Pool size and mean age of stable soil organic carbon in cropland. *Soil Sci. Soc. Am. J.* **56**(2), 460–464. doi: 10.2136/sssaj1992.03615995005600020020x (1992).
4. Bernoux, M., Cerri, C. C., Neill, C. & De Moraes, J. F. L. 1998. The use of stable carbon isotopes for estimating soil organic matter turnover rates. *Geoderma* **82**, 43–58. doi: 10.1016/S0016-7061(97)00096-7 (1998).
5. Rasmussen, P. E., et al. Review long-term agroecosystem experiments: Assessing agricultural sustainability and global change. *Science* **282**, 893–896. doi: 10.1126/science.282.5390.893 (1998).
6. Schmidt, M.W.I. et al. 2011. Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56. doi: 10.1038/nature10386 (2011).
7. Powlson, D. S., et al. *Limited Potential of No-till Agriculture for Climate Change Mitigation*. Nat. Publ. Gr. 4. doi: 10.1038/NCLIMATE2292 (2014).
8. Poeplau, C. & Don, A. Carbon sequestration in agricultural soils via cultivation of cover crops – A meta-analysis. *Agric. Ecosyst. Environ.* **200**, 33–41. doi: 10.1016/j.agee.2014.10.024 (2015).
9. Drinkwater, L. E., Friedman, D. & Buck, L. *Systems Research for Agriculture: Innovative Solutions to Complex Challenges*. SARE Handb. Ser. 13. Sustain. Agric. Res. & Education (SAR) Progr., College Park, Maryland (2016).
10. Drinkwater, L. E. Cropping systems research: Reconsidering agricultural experimental approaches. *HortTechnology* **12**, 355–361 doi: 10.21273/HORTTECH.12.3.355 (2002).
11. Silva, E. M. & Tchamitchian, M. Long-term systems experiments and long-term agricultural research sites: Tools for overcoming the border problem in agroecological research and design. *Agroecol. Sustain. Food Syst.* **42**, :620–628. doi: 10.1080/21683565.2018.1435434 (2018).
12. Cochran, W. G. Long-term agricultural experiments. *Suppl. J. Royal Stat. Soc.* **6**, 104–148. doi: 10.2307/2983686 (1939).

13. Berti, A., Dalla Marta, A., Mazzoncini, M. & Tei, F. An overview on long-term agro-ecosystem experiments: Present situation and future potential. *Eur. J. Agron.* **77**, 236–241 doi: 10.1016/j.eja.2016.01.004 (2016).
14. Deytieux, V., Munier-Jolain, N. & Caneill, J. Assessing the sustainability of cropping systems in single- and multi-site studies. A review of methods. *Eur. J. Agron.* **72**, 107–126. doi: 10.1016/J.EJA.2015.10.005 (2016).
15. Granered, N. & Bates Prins, S. C. Identifying outlying observations in regression trees. *Missouri J. Math. Sci.* **28**, 76-87 (2016).
16. Pribyl, D. W. A critical review of the conventional SOC to SOM conversion factor. *Geoderma* **156**, 75–83 doi: 10.1016/j.geoderma.2010.02.003 (2010).
17. Six, J., Conant, R.T. Paul, E. A. & Paustian, K. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant Soil* **241**, 155–176 (2002).
18. Chung, H., Grove, J. H. & Six, J. 2008. Indications for soil carbon saturation in a temperate agroecosystem. *Soil Sci. Soc. Am. J.* **72**, 1132. doi: 10.2136/sssaj2007.0265 (2008).
19. Gulde, S., Chung, H., Amelung, W., Chang, C. & Six, J. Soil carbon saturation controls labile and stable carbon pool dynamics. *Soil Sci. Soc. Am. J.* **72**, 605. doi: 10.2136/sssaj2007.0251 (2008).
20. Needelman, B.A., et al. Interaction of tillage and soil texture: Biologically active soil organic matter in Illinois. *Soil Sci. Soc. Am. J.* **31**, 1326–1334. doi: 10.2136/sssaj1999.6351325x (1999).
21. Karlen, D.L., Cambardella, C. A., Kovar, J. L. & Colvin, T. S. Soil quality response to long-term tillage and crop rotation practices. *Soil Tillage Res.* **133**, 54–64. doi: 10.1016/j.still.2013.05.013 (2013).
22. Fine, A. K. & Schindelbeck, R.R. Statistics, scoring functions, and regional analysis of a comprehensive soil health database. *Soil Sci. Soc. Am. J.* **81**, 589–601. doi: 10.2136/sssaj2016.09.0286 (2017).
23. McDaniel, M. D., Tiemann, L.K. & Grandy, A.S. Does agricultural crop diversity enhance soil microbial biomass and organic matter dynamics? A meta-analysis. *Ecol. Appl.* **24**, 560–570 (2014).
24. Collins, D. P. Farm-scale variation of soil quality indices and association with edaphic properties. *Soil Sci. Soc. Am. J.* **75**, 580–590. doi: 10.2136/sssaj2010.0029 (2011).
25. Trimble, S. Fluvial processes, morphology and sediment budgets in the Coon Creek Basin, WI USA, 1975-1993. *Geomorphology* **108**, 8-23 (2009).
26. Poffenbarger, H. J., et al. Maximum soil organic carbon storage in Midwest U.S. cropping systems when crops are optimally nitrogen-fertilized (X Wang, ed.). *PLoS One* **12**, e0172293. doi:

10.1371/journal.pone.0172293 (2017).

27. Jokela, W., et al. Midwest cropping system effects on soil properties and on a soil quality index. *Agron. J.* **103**, 1552. doi: 10.2134/agronj2010.0454 (2011).
28. Culman, S. W., Snapp, S. S., Green, J. M. & Gentry, L.E. 2013. Short-and long-term labile soil carbon and nitrogen dynamics reflect management and predict corn agronomic performance. *Agron. J.* **105**, 493–502. doi: 10.2134/agronj2012.0382 (2013).
29. Castellano, M. J., et al. Integrating plant litter quality, soil organic matter stabilization, and the carbon saturation concept. *Glob. Chang. Biol.* **21**, 3200–3209. doi: 10.1111/gcb.12982 (2015).
30. Stranger, T.F. & Lauer, J.G. Corn grain yield response to crop rotation and nitrogen over 35 years. *Agron J.* 100,643-650. doi: 10.2134/agronj2007.0280 (2008)
31. Pedersen, P. & Lauer, J.G. Corn and soybean response to rotation sequence, row spacing, and tillage systems. *Agron. J.* 95:965-971. doi: 10.2134/agronj2003.0965 (2003)
32. Oldfield, E.E., Bradford, M.A. & Wood, S.A. Global meta-analysis of the relationship between soil organic matter and crop yields. *SOIL* 5, 13-32. doi: 10.5194/soil-5-15-2019 (2019).
33. Gregorich, E. G., Drury, C. F., Baldock, J. A. & Baldock, C.F. Changes in soil carbon under long-term maize in monoculture and legume-based rotation. *Can. J. Soil Sci.* **81**, 21–31 (2000).
34. Su, Y. Z. Soil carbon and nitrogen sequestration following the conversion of cropland to alfalfa forage land in northwest China. *Soil Tillage Res.* **92**, 181–189. doi: 10.1016/j.still.2006.03.001 (2007).
35. Bauer, A. & Black, A. L. Soil carbon, nitrogen, and bulk density comparisons in two cropland tillage systems after 25 years and in virgin grassland. *Soil Sci. Soc. Am. J.* **45**, 1166–1170 (1981).
36. King, A. & Blesh, J. Crop rotations for increased soil carbon: perennality as a guiding principle. *Ecol. Appl.* **28**, 249-261. doi: 10.1002/eap.1648 (2018).
37. Kaspar, T. C., et al. Examining changes in soil organic carbon with oat and rye cover crops using terrain covariates. *Soil Sci. Soc. Am. J.* **70**, 1168–1177. doi: 10.2136/sssaj2005.0095 (2006).
38. Moore, E. B., Wiedenhoef, M. H., Kaspar, T.C. & Cambardella, C. A. 2014. Rye cover crop effects on soil quality in no-till corn silage-soybean cropping systems. *Soil Sci. Soc. Am. J.* **78**, 968–976. doi: 10.2136/sssaj2013.09.0401 (2014).
39. Cates, A. M. & Jackson, R. D. Cover crop effects on net ecosystem carbon balance in grain and silage maize. *Agron. J.* **111**, 30-38. doi:10.2134/agronj2018.01.0045 (2019).
40. Sommerfeldt, T. G., Chang, C. & Entz, T. Long-term annual manure applications increase soil organic matter and nitrogen, and decrease carbon to nitrogen ratio. *Soil Sci. Soc. Am. J.* **52**, 1668. doi:

10.2136/sssaj1988.03615995005200060030x (1988).

41. Hao, X., Chang, C., Travis, G.R. & Zhang, F. Soil carbon and nitrogen response to 25 annual cattle manure applications. *J. Plant Nutr. Soil Sci.* **166**, 239–245 (2003).

42. Zhang, W., et al. *Relative Contribution of Maize and External Manure Amendment to Soil Carbon Sequestration in a Long-term Intensive Maize Cropping System*. Nat. Publ. Gr. doi: 10.1038/srep10791 (2015).

43. Maillard, E. & Angers, D. A. Animal manure application and soil organic carbon stocks: a meta-analysis. *Glob. Chang. Biol.* **20**, 666–679. doi: 10.1111/gcb.12438 (2014).

44. Gal, A., Vyn, T.J., Micheli, E., Kladvko, E.J., & McFee W.W. Soil carbon and nitrogen accumulation with long-term no-till vs moldboard plowing overestimated with tilled-zone sampling depths. *Soil Tillage Res.* 96:42-51 (2007).

45. Sundermeier, A. P., Islam, K. R., Reeder, R. C. & Dick, W. A. Continuous no-till impacts on soil biophysical carbon sequestration. *Soil Sci. Soc. Am. J.* **75**, 1779–1788. doi: 10.2136/sssaj2010.0334 (2011).

46. Van Eerd, L. L., Congreves, K. Hayes, A., Verhalen, A. & Hooker, D. C. Long-term tillage and crop rotation effects on soil quality, organic carbon, and total nitrogen. *Can. J. Soil Sci.* **94**, 303–315. doi: 10.4141/CJSS2013-093 (2014).

47. Nelson, D.W. & L.E. Sommers. Total carbon, organic carbon, and organic matter. *Methods of Soil Analysis, Part 3 – Chemical Methods*. eds. D.L. Sparks et al. Soil Science Society of America, Madison, WI. pp 961-1010. doi: 10.2136/sssabookser5.3.c34 (1996)

48. Soil Survey Staff, Natural Resources Conservation Service, USDA. Web soil survey <http://websoilsurvey.nrcs.gov/> (accessed 25 Sept. 2019) (2019)

49. Laboski, C.A.M. & Peters, J.B. Nutrient application guidelines for field, vegetable, and fruit crops in Wisconsin. University of Wisconsin-Extension pub. A2809.

50. Peters, J., Nathan, M., & Laboski, C. pH and lime requirement. *In Recommended chemical soil test procedures for the North Central Region*, North Central Regional Research Publication No. 221 (revised), Missouri Agricultural Experiment Station SB 1001 (2015).

51. Combs, S.M. & Nathan, M.V. Soil organic matter. *In Recommended chemical soil test procedures for the North Central Region*, North Central Regional Research Publication No. 221 (revised), Missouri Agricultural Experiment Station SB 1001 (2015).

52. Harris, D., Horwath, W.R., & van Kessel, C. Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. *Soil Sci. Soc. Am. J.* 65:1853-1856 (2001)

53. R Development Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/> (2017).
54. R Studio Team. *R Studio: Integrated Development for R* [Online]. <http://www.rstudio.com/> [22 May 2018, date last accessed (2018)].
55. SAS Institute Inc. *SAS® 9.4 Statements: Reference*. SAS Institute Inc., Cary, North Carolina (2013).
56. Shapiro, S. S. & Wilk, M. B. Biometrika trust an analysis of variance test for normality. *Biometrika* **52**, 591–611 (1965).
57. Therneau T., Atkinson, B. & Ripley, B. *Package 'rpart'* [Online]. <https://cran.r-project.org/web/packages/rpart/rpart.pdf>. Accessed March 2018 (2018).
58. De'ath, G. & Fabricius, K. E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* **81**, 3178–3192. doi: 10.1890/0012-9658(2000)081[3178:cartap]2.0.co;2 (2000).

Figures

Figure 1

Location of fields sampled as part of this study in 2015, 2016, and 2017 (ArcMap, ver. 10.7.1, desktop.arcgis.com/en/arcmap). Different labels are used to represent the different geographic regions in the study with number of sites in parenthesis.

Figure 2

Linear relationships between soil organic matter (SOM) and total organic carbon (TOC) (A), SOM and total nitrogen (TN) (B), and TOC and TN (C) with regression equations and R2 presented. Additional relationships are presented between TOC or TN with SOM, with SOM values less than 50 g kg⁻¹.

Figure 3

Box plots for soil organic matter (SOM), total organic carbon (TOC), and total nitrogen (TN) across seven drainage classes as defined by the NRCS Soil Survey and between field with and without artificial tile

drainage. Treatments with the same letter indicate no difference at $\alpha=0.10$.

Figure 4

Box plots for soil organic matter (SOM), total organic carbon (TOC), and total nitrogen (TN) across eight soil texture classes as defined by the NRCS Soil Survey. Treatments with the same letter indicate no difference at $\alpha=0.10$.

Figure 5

Regression tree predicting SOM (g kg⁻¹) from categorical and continual variables described in Table 2, S1 and S2 (n=218).

Figure 6

Regression tree predicting TOC (g kg⁻¹) from categorical and continual variables described in Table 2, S1 and S2 (n=216).

Figure 7

Regression tree predicting TOC (g kg⁻¹) from categorical and continual variables described in Table 2, S1 and S2 (n=216).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AllsupplementaryTables20Jan.docx](#)