

RESEARCH ARTICLE

# Differences in soil biological activity by terrain types at the sub-field scale in central Iowa US

Amy L. Kaleita<sup>1</sup>\*, Linda R. Schott<sup>2</sup>, Sarah K. Hargreaves<sup>3,4</sup>, Kirsten S. Hofmockel<sup>4,5</sup>

**1** Agricultural & Biosystems Engineering, Iowa State University, Ames, Iowa, United States of America, **2** Biological Systems Engineering, University of Nebraska, Lincoln, Nebraska, United States of America, **3** Ecological Farmers Association of Ontario, Guelph, Ontario, Canada, **4** Formerly: Ecology, Evolutionary and Organismal Biology, Iowa State University, Ames, Iowa, United States of America, **5** Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington, United States of America

\* These authors contributed equally to this work.

\* [kaleita@iastate.edu](mailto:kaleita@iastate.edu)



## Abstract

Soil microbial communities are structured by biogeochemical processes that occur at many different spatial scales, which makes soil sampling difficult. Because soil microbial communities are important in nutrient cycling and soil fertility, it is important to understand how microbial communities function within the heterogeneous soil landscape. In this study, a self-organizing map was used to determine whether landscape data can be used to characterize the distribution of microbial biomass and activity in order to provide an improved understanding of soil microbial community function. Points within a row crop field in south-central Iowa were clustered via a self-organizing map using six landscape properties into three separate landscape clusters. Twelve sampling locations per cluster were chosen for a total of 36 locations. After the soil samples were collected, the samples were then analysed for various metabolic indicators, such as nitrogen and carbon mineralization, extractable organic carbon, microbial biomass, etc. It was found that sampling locations located in the potholes and toe slope positions had significantly greater microbial biomass nitrogen and carbon, total carbon, total nitrogen and extractable organic carbon than the other two landscape position clusters, while locations located on the upslope did not differ significantly from the other landscape clusters. However, factors such as nitrate, ammonia, and nitrogen and carbon mineralization did not differ significantly across the landscape. Overall, this research demonstrates the effectiveness of a terrain-based clustering method for guiding soil sampling of microbial communities.

## OPEN ACCESS

**Citation:** Kaleita AL, Schott LR, Hargreaves SK, Hofmockel KS (2017) Differences in soil biological activity by terrain types at the sub-field scale in central Iowa US. PLoS ONE 12(7): e0180596. <https://doi.org/10.1371/journal.pone.0180596>

**Editor:** Dafeng Hui, Tennessee State University, UNITED STATES

**Received:** March 8, 2017

**Accepted:** June 16, 2017

**Published:** July 7, 2017

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Microbial biomass and activity are both critical to, and sensitive indicators of, ecosystem health [1]. In production agricultural systems, microbial functions are especially important due to the high nutrient demands of plants.

Precision agriculture is a domain which has focused on techniques for site-specific management of nutrients, water, weeds, and disease, largely without much attention to the microbial

communities that drive many of the plant and soil processes critical to agricultural productivity and agricultural impacts on the environment. Welbaum et al. [2] point out the need for building on the advancements of precision agriculture and integrated pest management by including an emphasis in microbial ecosystem management in production agriculture. The authors call for improved methods of microbial monitoring for this purpose. This is echoed by Shennan [3] who points out the need for considering ecological interactions at a range of scales to inform and optimize agroecosystem management.

Soil microorganisms exhibit high spatial variability, even in uniformly managed agroecosystems [4], making it difficult to capture their abundance and function through standard random plot sampling. Samples are limited to specific points in time and space, which may not be representative or informative. The spatial patterns exhibited can vary significantly depending upon the spatial resolution of the sampling itself (eg. [5–7]). The influence of environmental properties may also be a function of scale [8], and may change over time as ambient conditions, particularly soil water content, change (e.g. [9–10]). In fact, shifts in microbial biomass can occur slowly in response to soil drying, but rapidly—within hours or days—in response to a rainfall event [11]. Hence, capturing microbial responses through field-scale soil sampling is not a trivial task. Ideally, observations would be made frequently (Landesman and Dighton [11] suggest hourly) and with a high spatial density (micrometre—meter scales [12]). However, this is rarely feasible given inevitable limitations of time and/or money. Thus, understanding how patterns in more readily observable properties relate to soil microbial properties may help to guide sampling design.

Topography is known to be a significant control on microbial communities, due primarily to the influence of topography on soil moisture and organic carbon [9, 13, 14]. For example, Florinsky et. al [9] measured the strength of association between individual topographic indices (slope, aspect, curvature, flow contributing area, topographic index, and stream power index) and soil microbial properties (microbial biomass, most probable number, denitrifier enzyme activity, denitrification rate, and microbial respiration rate). By using these relationships in a regression, they were able to develop predictive maps of microbial biomass and denitrifier enzyme activity. The strength of topographic controls varied by soil wetness conditions, decreasing to statistically insignificant relationships between topography and microbial properties under dry conditions.

The use of soil physicochemical characteristics to inform soil microbial sampling or activity mapping have produced differing results. Research has shown that in agricultural soils, microbial biomass and respiration are directly related to spatial patterns in soil pH, dissolved organic carbon, and total organic carbon [4]. In contrast, Katsalirou et al. [15] found weak spatial structure of organic carbon in cultivated soil, and strong positive linear relationships between soil organic carbon and microbial properties in uncultivated soils. Together these studies demonstrate the limitations of traditional geostatistics for soil sampling designs, and the need to use alternative parameters to guide sampling designs. For example, Cavigelli et al. [16] attempted to address how soil microbial property sampling schemes could be guided by soils information: drainage class, soil series, soil map unit, and texture of the Ap horizon. A relatively small proportion of measured soil microbial properties showed soil type effects, suggesting that terrain attributes and other factors that impact soil moisture dynamics may be most important [16]. Because of the high degree of microbial functional diversity and the strong influence of environmental factors, conventional soil classification may not provide a good indicator of microbial activity [17]. Instead, soil structure, which is intimately related to soil moisture and oxygen availability, may be better predictors of microbial biomass and function [12].

Peigné et al. [18] suggested that sampling design should be based upon historical exhaustive sampling of the soil and/or microbial activity to establish their spatial structure and to delineate “zones” of expected high, medium, and low values of the parameter(s) of interest. While useful, the pre-sampling required to design a sampling scheme in this way may be time- and cost-prohibitive in many cases. Van Arkel and Kaleita [19] illustrated that soil moisture sampling could be guided by a clustering analysis on soil and terrain attributes. Given the importance of soil moisture on microbial activity, and the body of literature suggesting that microbial patterns are related to a combination of soil factors and water-related terrain factors, it is reasonable to hypothesize that microbial sampling could be effectively guided with the same or similar approach. A number of studies have investigated “landform segmentation” for delineating zones related to soil microbial activity (enumerated in Pennock [14]), however, these have generally focused only on topographical attributes, and have not emphasized development of a practical methodology for sampling. Tajik et al. [20] considered both soil and terrain variables to predict soil enzyme activity, but the soil variables were derived from extensive soil sampling, which diminishes the practicality of this approach.

In this study, we use the Van Arkel and Kaleita [19] approach to classifying a landscape into moisture-related groups on the basis of readily available terrain and soil attributes, and examine if there is a detectable difference in microbial communities that reside in the different landscape classes within a 60-ha corn/soybean field in central Iowa, US. We use a self-organizing map to determine the number of landscape classes evident in the study area, followed by K-means clustering to partition the landscape into that number of classes. We then analyse soil samples taken from within those classes to determine if there are statistically significant differences in microbial biomass and activity. If so, this would suggest a grouping or clustering algorithm could be useful for guiding soil microbial sampling design.

## Methods

### Site information

The study site is a roughly 77-ha conventionally farmed field in Story County, central Iowa, with a crop rotation of corn and soybeans. The elevation in the field varies by approximately 6 m. While the elevation changes are moderate, the site includes different landscape positions, including ridges, toeslopes, and closed depressions. According to the National Cooperative Soil Survey (NCSS), there are six main soil types in this field: two loams, three clay loams, and one silty clay loam, located in the closed depressions. However, NCSS soil type delineations are not precise at this fine resolution (e.g. [21]).

Elevation data for this field were obtained using a GPS receiver mounted on the all-terrain vehicle that pulled the EMI sled. Using Surfer® (Golden Software, Inc., Golden, Colorado), a 10 m grid of elevation data was interpolated from this elevation data. Slope, planar curvature, and slope aspect were then derived using Surfer®, which uses computational processes based on Moore et al. [22] to calculate terrain attributes. A 10 m grid was used based upon the finding by Yang [23] that this scale was adequate to describe field-scale soil moisture patterns at an adjacent field site; others have also suggested the 10 m scale is appropriate for managing spatially variable soil moisture [24].

In the absence of high-resolution soils data, we use electromagnetic inductance (EMI) as a proxy to identify changing soil properties. This noncontact sensor is sensitive to variations in several characteristics of the soil, including soil texture, soil moisture content, organic matter, and depth of clay pan. Consequently, the EMI data are not direct measures of any single soil property, but variations in EMI do reflect the heterogeneity in soil properties and for this reason are frequently used as a low-cost alternative to extensive soil sampling in

applications where soil spatial variability is of interest [25]. Both horizontal (H-H) and vertical (V-V) conductances in units of milliSiemens/meter were gathered using an EMI sled pulled by an all-terrain vehicle. These data were collected in early spring when the field was saturated with snowmelt and spring rain, in order to minimize EMI differences due to soil moisture spatial variability. EMI data were interpolated with inverse distance weighting from a roughly 20 m average resolution (measurements were more dense in the direction of travel of the sled than in between sled passes) to the same 10 m landscape grid as the elevation data.

Thus, for each of the 7050 locations on the 10m grid across the study area, we had a six-element vector of terrain and soil characteristics that included: elevation, slope, slope aspect, planar curvature, H-H EMI, and V-V EMI. These were used as the inputs to the self-organizing map algorithm as described below. Summary statistics of these data for our site are given in Table 1.

### Sampling location selection

A method of analysis gaining popularity in modeling natural processes is that of computational intelligence, specifically self-organizing feature maps (SOMs), a type of artificial neural network approach. SOM networks learn to arrange occurrences of similar input patterns from a high dimensional input space into a low dimensional lattice of ‘neurons’ in an output layer [26]. The end result is an output network layer (map) with contiguous neurons having similar patterns in their input data [27]. Mele and Crowley [28] successfully used self-organizing maps to demonstrate relationships among different soil biogeochemical properties that can be used to assess soil quality. An advantage of using SOMs to analyse ecological communities is that this network approach allows visualization of associations among data even when the associations occur in different dimensions of the data space [29]. In the case of spatial data, SOM allows for identification of relationships in the data regardless of the spatial proximity of sampling locations. Often, SOMs are combined with a clustering algorithm to find clusters of similarly behaving input data (e.g. [30]). Self-organizing maps may, thus, be useful as a way to identify groups of location in the field, which could be subsampled.

A thorough explanation of the SOM procedure can be found in Kohonen [26]. A set of  $n$  input vectors of  $k$  elements are presented to the SOM algorithm and the vectors then ‘self-organize’ into a two-dimensional map according to the following process. To begin the algorithm produces a random arrangement of the input vectors into a 2-dimensional lattice map. Neurons or nodes on the output map are represented by the vector  $m_i = [m_{i1}, m_{i2}, \dots, m_{ik}]$ , ( $i = 1, 2, \dots, N$ ), where  $N$  is the number of neurons on the output map,  $k$  is the input vector dimension and  $m_{ik}$  represents the value of input data assumed (or later, computed through the updating process described below) for that node. The number of neurons on the map is user-specified and need not be the same as  $n$ .

**Table 1. Terrain and soil characteristics across the study area.**

<b>Attribute (units)</b>	<b>Range</b>	<b>Mean (Std. Dev)</b>
<b>Elevation (m)</b>	309.8–315.9	312.4 (1.3)
<b>Slope (%)</b>	0.0–4.6	1.2 (0.7)
<b>Slope aspect (degrees clockwise from north)</b>	0–360	178 (105)
<b>Plan curvature (<math>m^{-1}</math>)</b>	-4.8–12.1	0.0 (0.2)
<b>H-H EMI</b>	14.9–90.7	47.0 (13.8)
<b>V-V EMI</b>	10.4–68.3	32.6 (11.3)

<https://doi.org/10.1371/journal.pone.0180596.t001>

The SOM is trained iteratively. At each iteration, the algorithm compares a randomly chosen input vector  $\mathbf{x}$  to each node in the lattice, and the winning node,  $\mathbf{m}_c$ , for that vector is chosen based on the formula

$$\|\mathbf{x}_{input} - \mathbf{m}_c\| = \min_i \{\|\mathbf{x}_{input} - \mathbf{m}_i\|\}. \quad (1)$$

Thus, the winning node is the node with the smallest Euclidean distance to the input vector. After finding the winning output node for each input vector, the node's vector is updated according to the following equation:

$$\mathbf{m}_i(t + 1) = \mathbf{m}_i(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \quad (2)$$

where  $t$  denotes the index of the iteration step,  $\mathbf{x}(t)$  is the input sample of  $\mathbf{x}_{input}$  in the iteration  $t$ , and  $h_{ci}(t)$  is called the neighborhood function around the winning node  $\mathbf{m}_c$ . During training,  $h_{ci}(t)$  is a decreasing function of the distance between the  $i^{th}$  and the  $c^{th}$  node. For convergence it is necessary that  $h_{ci}(t)$  goes to 0 when  $t$  goes to infinity. The iteration then proceeds, and node vectors are updated at each iteration, with the end result that the final map has a distribution that represents the characteristics of all  $n$  of the  $k$ -dimensional input vectors in a two-dimensional space of  $N$  nodes.

To produce the input vectors from the elevation, EMI (H-H and V-V), slope, aspect, and curvature grid data, a matrix was built with each parameter in a different column, and rows representing individual field locations in the 10 m grid, for a resulting 7050 x 6 matrix ( $n = 7050$  and  $k = 6$ ). In this way, a vector of terrain characteristics for each location was generated. This matrix was then fed to the SOM algorithm within the MATLAB SOM Toolbox 2.0 [31]. Different process options can be chosen within the SOM algorithm; in this study, the default parameters within the Toolbox were used. Because the magnitudes of the 6 columns of characteristic data varied, SOM development was based upon normalized data for each characteristic (this is an option within the Toolbox).

Using the SOM Toolbox, a unified distance matrix (U-matrix) was generated. The U-matrix shows the distance between the nodes and can be used to identify patterns within the data. Distance between nodes is an inverse representation of similarity; the larger the distance between two nodes, the more dissimilar are the characteristics of the grid points in those nodes. By observing the color difference in the map and using the color scale, one can see the difference in distances between nodes within the U-matrix. Colors corresponding to small numerical values show that the nodes are closely related, whereas colors corresponding to large numerical values show divisions within the input data. Based on this visualization, the number of different groups of data in the field characteristics can be inferred, because large separation suggest large differences in the data. As will be explained in more detail in section 3, three groups were assumed for this site.

A K-means clustering procedure was then used to separate both the SOM map neurons and the temporal and physical data matrices into different clusters containing points with similar characteristics. In the K-means algorithm, the initial 'means' of a  $k$  number of clusters is randomly selected from the input data set. Clusters are created by associating each input vector (grid point data) to the nearest mean. The nearest mean is determined by finding the smallest Euclidean distance between the input vector and the mean vector. The input vectors are then partitioned into clusters depending on their distances from the mean vectors. The geometric centre of each of the clusters becomes the new mean. This process continues until there is no change in cluster membership given additional iterations of the algorithm. Readers are referred to MacQueen [32] for further explanation of the K-means algorithm. The MATLAB

SOM Toolbox contains a K-means clustering algorithm that was used for partitioning the neurons in the output layer of the SOM.

After applying the K-means clustering algorithm to both the SOM neuron data, the centroid vector of each cluster in each method was identified. As the name suggests, the centroids represent the centres (means) of the clusters created. Using the Euclidean distance formula, the input vector (corresponding to a single location) with the smallest distance from each centroid was identified. This input vector (location) was deemed the best matching unit (BMU) to the cluster centroid. These BMUs were then used as the critical sampling locations identified for this field. Final soil sampling locations thus included the BMU for each of the three groups, and an additional eleven sites within each group, distributed spatially across the field, for a total of 36 sites.

## Soil biological sampling and processing

At each of the 36 locations, we collected a single 7.6 cm (3 inch) soil core to a depth of 10 cm and stored it at 4°C until analysis. With permission from the land owner (Iowa State University), we collected soil samples on a single date, October 24, 2011, prior to the field being harvested. Prior to analysis, we homogenized soil samples to a particle size of at least 4mm and sub-sampled individual measurements. We measured gravimetric water content as soil mass loss upon drying at 60°C for 72 h, soil pH using 1:2 soil: water ratio, and total C and N by dry combustion (LECO TruSpec CN, St. Joseph, MI).

We also analysed each sample for soil microbial biomass, mineralization rates and available C and N pools. We measured microbial biomass C (MBC) and microbial biomass N (MBN) as the difference in K<sub>2</sub>SO<sub>4</sub>-extractable dissolved organic C (DOC) and total N (Shimadzu TOC-L CPH/CPN, Kyoto, Japan), respectively, between chloroform fumigated and non-fumigated subsamples using 15 g of soil extracted with 45 mL 0.5M K<sub>2</sub>SO<sub>4</sub>. We used the non-fumigated K<sub>2</sub>SO<sub>4</sub>-extractable DOC value to represent the soil available DOC concentration (modified from [33]). We used a calibration factor of 0.45 for MBC [34] and 0.54 for MBN [35].

We measured potential N mineralization as the difference between extractable inorganic N after a 28-day incubation (modified from [33]), and used non-fumigated samples to determine extractable inorganic pools of N at the beginning of the incubation. We measured extractable nitrate and ammonium concentrations in microplate format via spectrophotometry (BioTek Synergy HT plate reader, BioTek Instruments, Inc., Winooski, VT) [36]. We measured potential C mineralization by sampling CO<sub>2</sub> concentrations in the headspace on days 1, 3, 7, and 10 using an infrared gas absorption analyser (LICOR) (modified from [33]). After each sampling, flasks were vented to prevent CO<sub>2</sub> build-up and possible inhibition of aerobic respiration. C mineralization rates were, therefore, calculated as the concentration per day.

## Statistical analysis

We generated correlation statistics to compare, on a point-basis, the microbial indicators with the terrain attributes for the 36 sampled locations. On a cluster group basis, we used ANOVA to determine if the microbial data from each of the three groups are significantly different. Student's t-tests were performed to determine if the BMU site provides a good estimate of the group mean.

## Results and discussion

Correlation analysis indicated that both terrain and soils attributes have similar trends to microbial indicators (Table 2). Overall, microbial indicators were most strongly correlated with the EMI data, with correlations ranging from 0.39 to 0.83 in magnitude. Specific carbon

**Table 2. Correlations between terrain and soil attributes and microbial indicators: extractable dissolved organic carbon (EOC; ug C g<sup>-1</sup> soil), microbial biomass carbon (MBC; ug C g<sup>-1</sup> soil), percent total soil carbon (%TC), microbial biomass N (MBN; ug N g<sup>-1</sup> soil), percent total soil nitrogen (%TN), specific Carbon mineralization (Cmin), and specific Nitrogen mineralization (Nmin).**

	<i>EOC</i>	<i>MBC</i>	<i>%TC</i>	<i>MBN</i>	<i>%TN</i>	<i>Cmin</i>	<i>Nmin</i>
<i>Elevation</i>	-0.49	-0.64	-0.68	-0.63	-0.63	0.38	0.30
<i>Slope</i>	-0.39	-0.63	-0.68	-0.60	-0.69	0.57	0.28
<i>Slope aspect</i>	-0.17	0.04	-0.07	0.01	0.02	-0.24	-0.24
<i>Plan curvature</i>	0.12	0.04	0.06	0.06	0.04	-0.10	0.04
<i>H-H EMI</i>	0.42	0.81	0.74	0.75	0.79	-0.48	-0.39
<i>V-V EMI</i>	0.45	0.83	0.79	0.79	0.82	-0.53	-0.44

<https://doi.org/10.1371/journal.pone.0180596.t002>

and nitrogen mineralization were negatively correlated with EMI, while the other indicators were positively correlated with EMI. Microbial indicators were also correlated, to a somewhat lesser degree, with elevation and slope. Plan curvature as an individual index was not correlated with any of the microbial indicators. Overall, specific N mineralization showed the least strong correlation with any of the landscape data, with no correlation coefficient greater than 0.39 in magnitude.

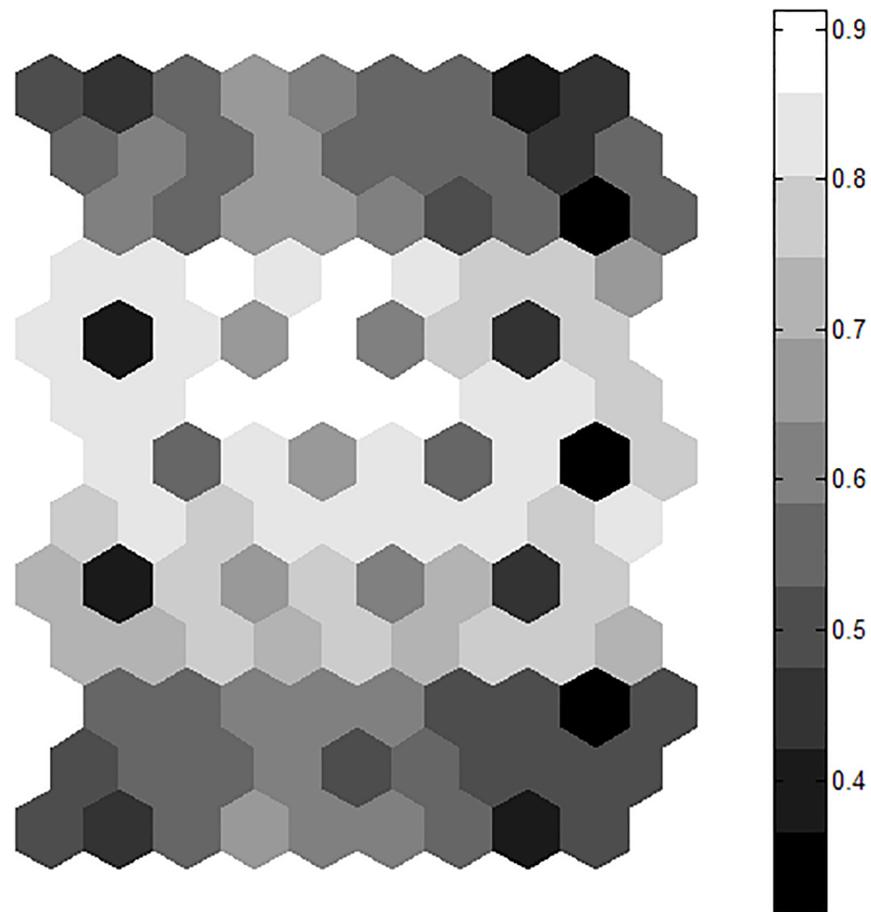
The SOM U-matrix was analysed to determine the appropriate number of groups or clusters in the field, based on the terrain and EMI data. Fig 1 shows the U-matrix. In this case, strong division between top and bottom sections of the U-Matrix suggests two groups. However, the middle of the U-matrix is wide, with some closely related nodes, and may suggest a third group. We thus assumed there were three groups in this field, and used the K-means clustering on the self-organizing map to segment the field into three groups accordingly.

The characteristics and spatial distribution of these groups were then reviewed. Fig 2 shows the spatial distribution in the field of the three groups resulting from the K-means clustering, and Fig 3 shows the characteristics of the three groups. Cluster one is primarily comprised of the potholes and toe slopes with higher EMI soils, which most likely have higher clay content. Cluster two and three are at higher elevations and have lower EMI values. These two groups are quite similar except that they face different directions; cluster two consists mainly of the southern facing slopes, whereas cluster three includes the northeast facing slopes. Cluster two has steeper slopes while slopes in cluster three are close to the field average.

The soil analysis results were then analysed by cluster. Fig 4 shows mean cluster values of microbial analysis. When grouped by cluster, mean values of DOC, MBC, MBN, TC, TN were significantly greater in cluster one than in clusters two and three ( $p < 0.05$ ). Mean DOC for cluster one was more than 1.3 times larger than either of the other two clusters. Mean MBC for cluster 1 was 1.6 times greater than for clusters 2 and 3, and mean MBN was 1.75 greater. Finally for TN and TC, the mean for cluster 1 is at least 1.8 and 1.6 times greater, respectively, than the other two clusters.

Potential rates of C and N mineralization, ammonia and nitrate concentrations, and the ratio of MBC to MBN did not differ among the three clusters. However, when the C and N mineralization rates were computed on a per unit MBC basis, the mass specific rates of C and N mineralization were significantly different ( $p < 4.0e-5$  and  $p < 0.051$ , respectively) between Clusters 1 and 2. Cluster 1 had the lowest rates and cluster 2 had the highest. Cluster 3 was statistically similar to both Cluster 1 and Cluster 2. These are shown in Fig 5.

With respect to microbial activity indicators, these results indicate that only two clusters are evident at this site. This suggests that two strong clusters in the SOM would have been sufficient; the only difference between clusters 2 and 3 in the SOM input data is the primary

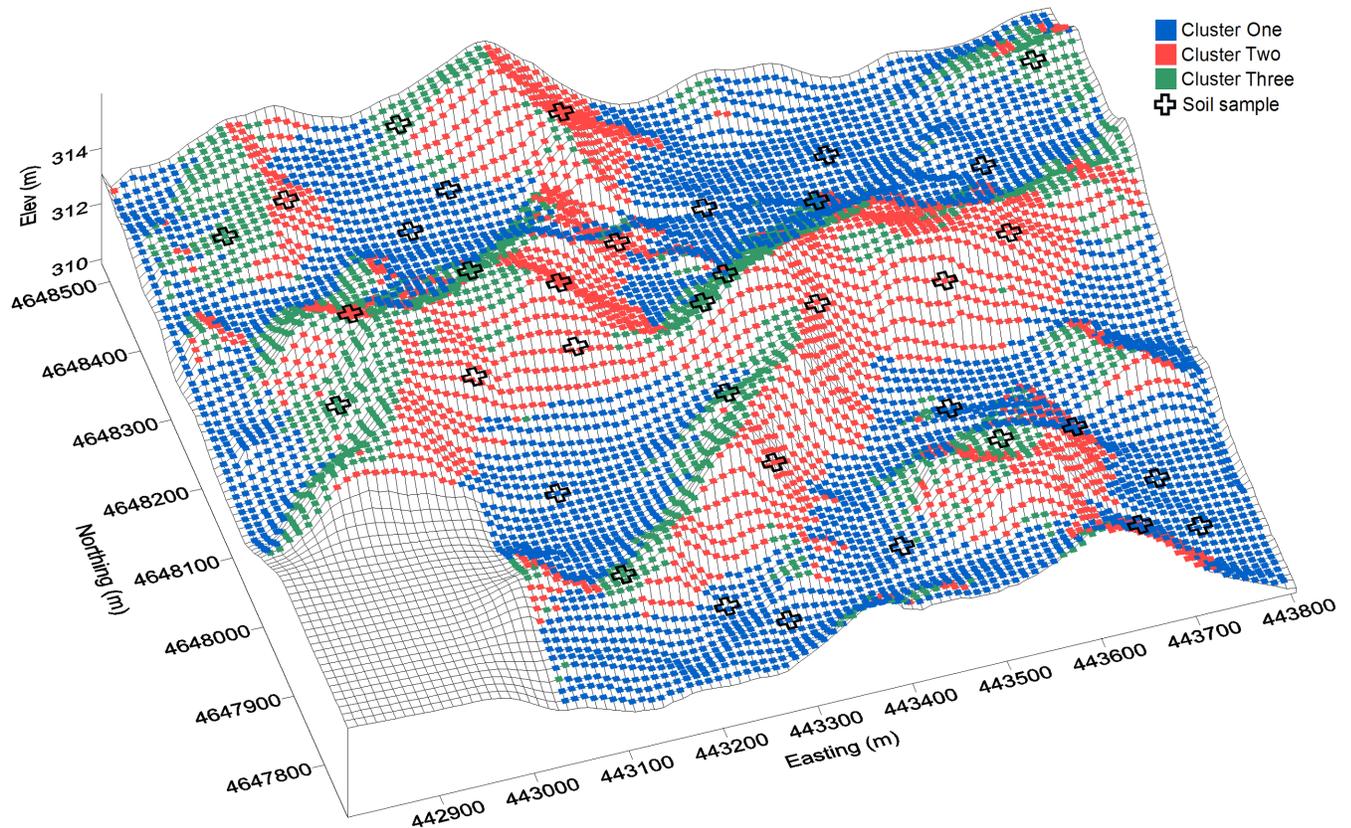


**Fig 1. U-matrix for the terrain and EMI data for the study site.** The legend bar to the right of the U-matrix denotes the Euclidean distance in input data between neighbouring nodes. Lower values indicate more similarity with neighbouring nodes.

<https://doi.org/10.1371/journal.pone.0180596.g001>

direction the hillside faces, and cluster 2 was steeper than cluster 3. The microbial activity within these two clusters was for the most part the same.

The clustering of sample locations based on small-scale topographic and soil features effectively identified differences in microbial abundance and function between two groups (cluster 1 versus clusters 2 and 3 combined). Potholes and toe slopes (cluster 1) had greater organic C, but these areas also had lower specific N mineralization and specific microbial respiration rates compared to the upslope areas (clusters 2 and 3). This indicates that under the wetter conditions of cluster one there is probably leaching of soluble C that is readily available for microbial assimilation and use for microbial growth, as suggested by the greater EOC and MBC in the pothole soils. Corre et al. [37] had similar results and attributed the differences in microbial biomass between the upper and lower slopes to the availability of materials for mineralization. The lower slopes have more redistributed soil material due to surface water flow, so these areas are able so support a higher microbial biomass. It is likely that the wet, nutrient-rich conditions in the potholes allow microbial substrates to diffuse throughout the soil, supporting a larger microbial community (MBC and MBN) compared to the upslope clusters 2 and 3.



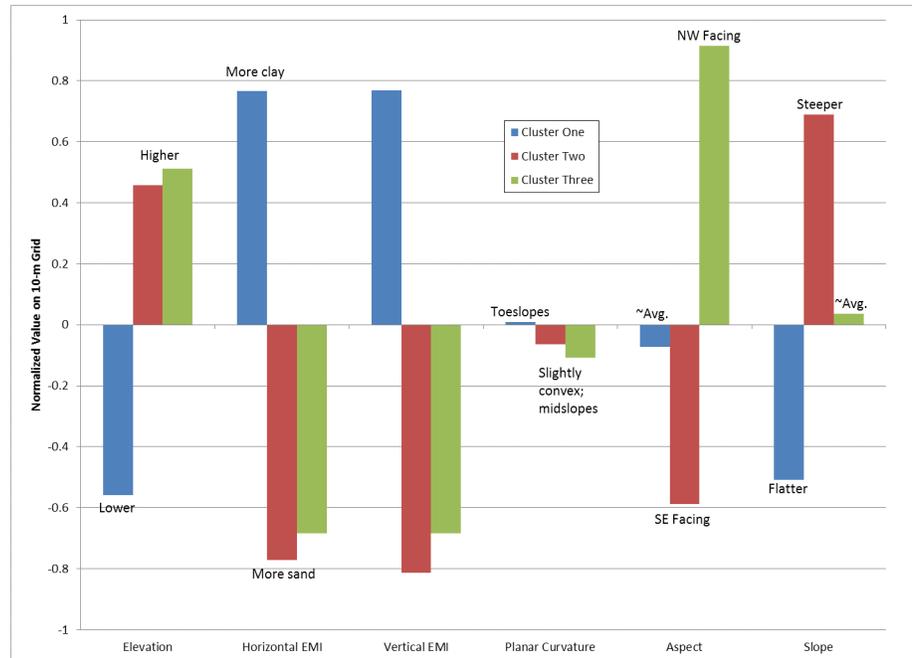
**Fig 2. Distribution of clusters of terrain and soils data in the field.** Soil sampling locations are marked. Each colored dot represents the centre point on a 10 meter grid.

<https://doi.org/10.1371/journal.pone.0180596.g002>

Results of the t-tests indicate that the BMUs are not necessarily good representatives of the typical (mean) microbial activity of each group (Table 3). Van Arkel and Kaleita [19] demonstrated that the BMUs identified through the K-means approach was a useful way to pinpoint individual locations to monitor for soil moisture. On the date of our soil sampling, however, selecting a single BMU from the cluster analysis appears to be less useful for identifying specific sampling locations for observing microbial activity. All of the BMUs had two or more soil microbial activity metrics that were statistically significantly different than the group mean. At the same time, microbial biomass carbon and microbial biomass nitrogen both showed significant agreement between BMU observation and group mean, suggesting that the BMU site was representative of the group behavior for these two metrics.

### Conclusions

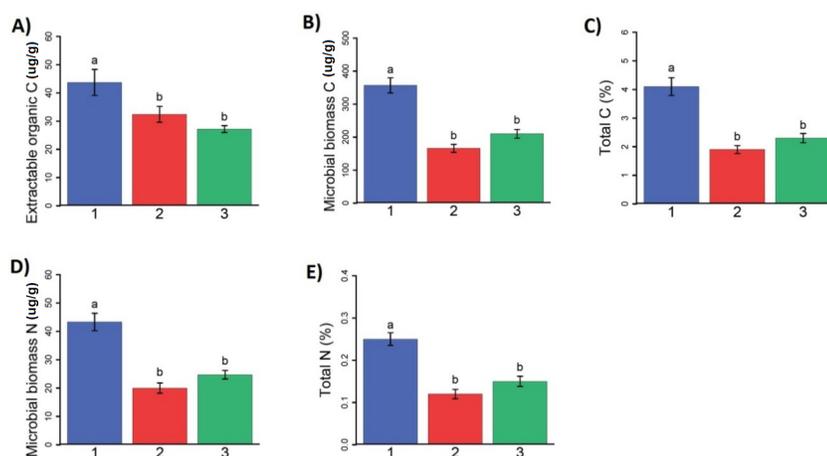
These data suggest that the sampling design for microbial monitoring should consider within-field landscape variability, and that the clustering approach is useful for doing so. While the single BMUs did not necessarily represent the microbial activity for their respective clusters, the soil analysis data showed statistically distinct microbial activity among clusters with different terrain and soils characteristics. The topography data used to generate the clusters can be obtained from high resolution LiDAR, increasingly available through regional government investment, or other surveys.



**Fig 3. Average characteristics of the terrain and EMI data for each cluster and their interpretation.** For development of the SOM, each attribute for all locations on the 10-meter grid was normalized to a mean of zero and standard deviation of one; the normalized values are shown here.

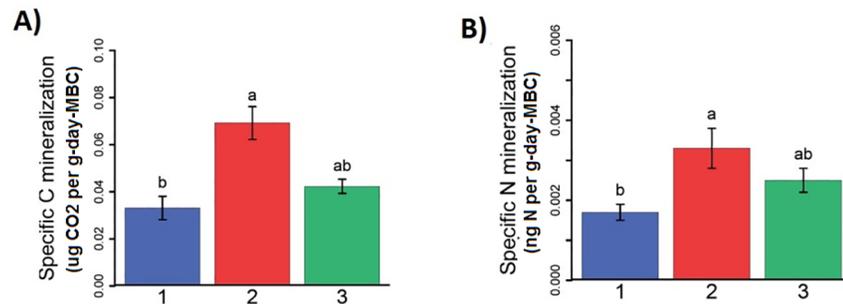
<https://doi.org/10.1371/journal.pone.0180596.g003>

One limitation of this study is that we used soil and microbial samples from just one day, late in the season. Microbial communities in agricultural fields are known to be responsive to management activities, among other factors. For instance, Girvan et al. [38] followed the bacterial and fungal communities in a wheat field over the course of a year. They observed that between-sample microbial heterogeneity decreased under a mature crop but increased following harvesting and ploughing. Our soil samples were collected after harvest, and thus may



**Fig 4. Mean cluster values for (A) extractable dissolved organic carbon ( $\mu\text{g C g}^{-1}$  soil), (B) microbial biomass carbon ( $\mu\text{g C g}^{-1}$  soil), (C) percent total soil carbon, (D) microbial biomass N ( $\mu\text{g N g}^{-1}$  soil), and (E) percent total soil nitrogen.** A significance level of  $p < 0.05$  was used for evaluation. Colors indicate location of soil clusters, where blue is cluster 1, red is cluster 2 and green is cluster 3.

<https://doi.org/10.1371/journal.pone.0180596.g004>



**Fig 5. Mean cluster values for (A) specific carbon mineralization and (B) specific N mineralization.** A significance level of  $p < 0.05$  was used for evaluation. Colors indicate location of soil clusters, where blue is cluster 1, red is cluster 2 and green is cluster 3.

<https://doi.org/10.1371/journal.pone.0180596.g005>

**Table 3. At a significance level of  $p < 0.05$ , whether or not there is a statistically significant difference between the mean of each group and that group's BMU.**

	Cluster 1	Cluster 2	Cluster 3
Extractable dissolved organic carbon ( $\mu\text{g C g}^{-1}$ soil)	No	Yes	No
Microbial biomass carbon ( $\mu\text{g C g}^{-1}$ soil)	No	No	No
Percent total soil carbon	Yes	Yes	No
Microbial biomass N ( $\mu\text{g N g}^{-1}$ soil)	No	No	No
Percent total soil nitrogen	Yes	Yes	Yes
Specific carbon mineralization	Yes	Yes	No
Specific N mineralization	No	Yes	Yes

<https://doi.org/10.1371/journal.pone.0180596.t003>

reflect more heterogeneity than had we sampled at a different time. Furthermore, one might expect temporal or seasonal dynamics to affect the relationship between terrain attributes and microbial activity; this is an area for future work.

Overall, this research demonstrates the value of spatially-explicit sampling for the field of soil microbial ecology, which currently struggles to find meaningful and efficient sampling strategies.

### Supporting information

**S1 File. Soil, terrain, and microbial indicator data.** This file includes easting and northing coordinates for each sample, the assigned cluster from the SOM, the moisture content (g/g), fresh weight to dry weight ratio, pH, MBC ( $\mu\text{g C/g DW}$ ), MBN ( $\mu\text{g N/g DW}$ ), Specific C mineralization, Specific N Mineralization, percent total carbon, percent total nitrogen, elevation (m), slope, aspect (degrees), plan curvature, horizontal-polarization EMI (mS/m), and vertical-polarization EMI (mS/m). (XLSX)

### Acknowledgments

The terrain and EMI data was provided by the USDA ARS National Laboratory for Agriculture and the Environment (Ames, IA, USA).

### Author Contributions

**Conceptualization:** Amy L. Kaleita, Kirsten S. Hofmockel.

**Data curation:** Amy L. Kaleita, Linda R. Schott.

**Formal analysis:** Amy L. Kaleita, Linda R. Schott.

**Investigation:** Linda R. Schott, Sarah K. Hargreaves.

**Methodology:** Amy L. Kaleita, Kirsten S. Hofmockel.

**Project administration:** Amy L. Kaleita.

**Resources:** Amy L. Kaleita, Kirsten S. Hofmockel.

**Software:** Amy L. Kaleita.

**Supervision:** Amy L. Kaleita, Kirsten S. Hofmockel.

**Validation:** Amy L. Kaleita, Sarah K. Hargreaves, Kirsten S. Hofmockel.

**Visualization:** Amy L. Kaleita, Linda R. Schott.

**Writing – original draft:** Amy L. Kaleita, Linda R. Schott, Sarah K. Hargreaves, Kirsten S. Hofmockel.

**Writing – review & editing:** Amy L. Kaleita, Sarah K. Hargreaves.

## References

1. Anderson T, Domsch K. Ratios of microbial biomass carbon to total organic carbon in arable soils. *Soil Biology and Biochemistry*. 1989; 21(4):471–479.
2. Welbaum G, Sturz A, Dong Z, Nowak J. Managing soil microorganisms to improve productivity of Agro-Ecosystems. *CRC Crit Rev Plant Sci*. 2004; 23(2):175–193.
3. Shennan C. Biotic interactions, ecological knowledge and agriculture. *Philosophical Philos Trans R Soc Lond B Biol Sci*. 2008; 363(1492):717–739.
4. Dupuis E, Whalen J. Soil properties related to the spatial pattern of microbial biomass and respiration in agroecosystems. *Can J Soil Sci*. 2007; 87(5):479–484.
5. Ettema C, Wardle D. Spatial soil ecology. *Trends Ecol Evol*. 2002; 17(4):177–183.
6. Franklin R, Mills A. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiol Ecol*. 2003; 44(3):335–346. [https://doi.org/10.1016/S0168-6496\(03\)00074-6](https://doi.org/10.1016/S0168-6496(03)00074-6) PMID: 12830827
7. Hargreaves S, Williams R, Hofmockel K. Environmental filtering of microbial communities in agricultural soil shifts with crop growth. *PLOS ONE*. 2015; 10(7):e0134345. <https://doi.org/10.1371/journal.pone.0134345> PMID: 26226508
8. Fromin N, Saby N, Lensi R, Brunet D, Porte B, Domenach A et al. Spatial variability of soil microbial functioning in a tropical rainforest of French Guiana using nested sampling. *Geoderma*. 2013; 197–198:98–107.
9. Florinsky I, McMahon S, Burton D. Topographic control of soil microbial activity: a case study of denitrifiers. *Geoderma*. 2004; 119(1–2):33–53.
10. Fierer N, Schimel J. Effects of drying–rewetting frequency on soil carbon and nitrogen transformations. *Soil Biol Biochem*. 2002; 34(6):777–787.
11. Landesman W, Dighton J. Shifts in Microbial Biomass and the Bacteria: Fungi Ratio Occur Under Field Conditions Within 3 h After Rainfall. *Microb Ecol*. 2011; 62(1):228–236. <https://doi.org/10.1007/s00248-011-9811-1> PMID: 21360141
12. Enwall K, Throback I, Stenberg M, Soderstrom M, Hallin S. Soil resources influence spatial patterns of denitrifying communities at scales compatible with land management. *Appl Environ Microbiol*. 2010; 76(7):2243–2250. <https://doi.org/10.1128/AEM.02197-09> PMID: 20118364
13. Stenrød M, Charnay M, Benoit P, Eklo O. Spatial variability of glyphosate mineralization and soil microbial characteristics in two Norwegian sandy loam soils as affected by surface topographical features. *Soil Biol Biochem*. 2006; 38(5):962–971.
14. Pennock D. Precision conservation for co-management of carbon and nitrogen in the Canadian prairies. *J Soil Water Conserv*. 2005; 60(6):396–401.

15. Katsalirou E, Deng S, Nofziger D, Gerakis A, Fuhlendorf S. Spatial structure of microbial biomass and activity in prairie soil ecosystems. *Eur J Soil Biol.* 2010; 46(3–4):181–189.
16. Cavigelli M, Lengnick L, Buyer J, Fravel D, Handoo Z, McCarty G et al. Landscape level variation in soil resources and microbial properties in a no-till corn field. *Appl Soil Ecol.* 2005; 29(2):99–123.
17. Simfukwe P, Hill P, Emmett B, Jones D. Soil classification provides a poor indicator of carbon turnover rates in soil. *Soil Biol Biochem.* 2011; 43(8):1688–1696.
18. Peigné J, Vian J, Cannavacciuolo M, Bottollier B, Chaussod R. Soil sampling based on field spatial variability of soil microbial indicators. *Eur J Soil Biol.* 2009; 45(5–6):488–495.
19. Van Arkel Z, Kaleita A. Identifying sampling locations for field-scale soil moisture estimation using K-means clustering. *Water Resour Res.* 2014; 50(8):7050–7057.
20. Tajik S, Ayoubi S, Nourbakhsh F. Prediction of soil enzymes activity by digital terrain analysis: Comparing artificial neural network and multiple linear regression models. *Environ Eng Sci.* 2012; 29(8):798–806.
21. Brevik E, Fenton T, Jaynes D. Evaluation of the accuracy of a central Iowa soil survey and implications for precision soil management. *Precision Agriculture.* 2003; 4(3):71–91.
22. Moore I, Grayson R, Ladson A. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol Process.* 1991; 5(1):3–30.
23. Yang L. Spatio-temporal patterns of field-scale soil moisture and their implications for in situ soil moisture network design [PhD]. Iowa State University; 2010.
24. Hedley C, Roudier P, Yule I, Ekanayake J, Bradbury S. Soil water status and water table depth modelling using electromagnetic surveys for precision irrigation scheduling. *Geoderma.* 2013; 199:22–29.
25. Adamchuk V, Hummel J, Morgan M, Upadhyaya S. On-the-go soil sensors for precision agriculture. *Comput Electron Agric.* 2004; 44(1):71–91.
26. Kohonen T. *Self-Organizing Maps.* 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg; 2001.
27. Kalteh A, Hjorth P, Berndtsson R. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environ Model Softw.* 2008; 23(7):835–845.
28. Mele P, Crowley D. Application of self-organizing maps for assessing soil biological quality. *Agriculture, Ecosystems & Environment.* 2008; 126(3–4):139–152.
29. Giraudel J, Lek S. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecol Modell.* 2001; 146(1–3):329–339.
30. Kiang M. Extending the Kohonen self-organizing map networks for clustering analysis. *Comput Stat Data Anal.* 2001; 38(2):161–180.
31. Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. SOM Toolbox for Matlab 5, Technical Report A57 [Internet]. Helsinki, Finland: Neural Networks Research Centre, Helsinki University of Technology; 2000. Available from: <http://www.cis.hut.fi/somtoolbox/package/papers/techrep.pdf>
32. MacQueen J. Some methods for classification and analysis of multivariate observation, In: *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1967: 1:281–297.
33. Robertson G. *Standard soil methods for long-term ecological research.* 1st ed. New York: Oxford University Press; 2011.
34. Joergensen R, Mueller T. The fumigation-extraction method to estimate soil microbial biomass: Calibration of the kEN value. *Soil Biology and Biochemistry.* 1996; 28(1):33–37.
35. Joergensen R. The fumigation-extraction method to estimate soil microbial biomass: Calibration of the kEC value. *Soil Biology and Biochemistry.* 1996; 28(1):25–31.
36. Hood-Nowotny R, Umana N, Inselbacher E, Oswald-Lachouani P, Wanek W. Alternative Methods for Measuring Inorganic, Organic, and Total Dissolved Nitrogen in Soil. *Soil Science Society of America Journal.* 2010; 74(3):1018.
37. Corre M, Schnabel R, Stout W. Spatial and seasonal variation of gross nitrogen transformations and microbial biomass in a Northeastern US grassland. *Soil Biol Biochem.* 2002; 34(4):445–457.
38. Girvan M, Bullimore J, Ball A, Pretty J, Osborn A. Responses of active bacterial and fungal communities in soils under winter wheat to different fertilizer and pesticide regimens. *Appl Environ Microbiol.* 2004; 70(5):2692–2701. <https://doi.org/10.1128/AEM.70.5.2692-2701.2004> PMID: 15128520