International Journal of Ecology and Environmental Sciences 40 (1): 1-14, 2014
ISSN: 0377-015X (Print); 2320-5199 (Online)

Spatial Structure of Soil Microbial Communities and the Evaluation of the Spatial Structure Analysis based on Relative Dissimilarity

SANGHOON KANG
Department of Biology, Baylor University, One Bear Place, #97388, Waco, TX 76798, U.S.A.
Corresponding author; *Email: sanghoon_kang@baylor.edu*

JOON JIN SONG
Department of Statistical Science, Baylor University, One Bear Place, #97140, Waco, TX 76798, U.S.A.
(Email: Joon_Song@baylor.edu)

ABSTRACT

The spatial structure of soil microbial communities was analyzed across the border of two distinctive grass vegetation patches. Genetic fingerprinting technique was used to assess microbial community structure in binary format data, which is not suitable for conventional semivariogram analysis. An alternative method of geostatistical analysis based on relative dissimilarity provided well fitted models from the experimental "semivariograms". Overall patterns of the models were fairly similar between bacterial and fungal communities, especially fittings (Indicative Goodness of Fit, IGF), sill (C_1) and ranges (a) at comparable scales. A small difference was also noticeable in response to distinctive vegetation settings. The evaluation of the alternative spatial structure analysis was done by comparing the semivariogram and the dissimogram (semivariogram based on dissimilarity) using continuous variables (total soil nitrogen content and C/N ratio from same sampling design). The extremely high fit of the models and prevalent high nugget value from all the dissimograms might be the results of artificiality from using relative dissimilarity instead of semivariance. Thus, one should be careful when using this approach, especially in estimation purposes such as kriging, since prediction is dependent upon the parameters determined in dissimogram models.

Key Words: Soil Microbial Community; Geostatistics; Spatial Structure; Relative Dissimilarity; Dissimogram; Semivariogram

INTRODUCTION

Studies that employ well-designed environmental sampling can provide insight into the variation within either temporal or spatial scales, or both. By analyzing the collected data, relationships among variables can be revealed. However, there is a chance that the relationship might be generated from some common temporal or spatial structure rather than their true interactions (Legendre and Troussellier 1988) i.e., the apparent relationships among variables may be due to factors other than the variables being examined. Spatial autocorrelation, an intrinsic property in terrestrial ecosystem for most variables, might hamper the use of tests of statistical significance due to the spatial dependence of variables measured in proximal samples because conventional statistical inferences are assumed independence among the variables. Generally, one observation accounts for one degree of freedom, but autocorrelation hinders researchers from counting degree of freedom directly from the number of observations. Since positive autocorrelation decreases the degree of freedom, analysis of autocorrelated samples with conventional statistics inflates the degree of freedom (Liebhold and Sharov 1998). Autocorrelation decreases the effective sample size, leading to inflated type I error and decreasing the statistical power (1-β) on statistical hypothesis testing unless the effective sample size is over the critical point which stabilizes statistical power regardless of sample size. This has been confirmed by the Monte Carlo simulation (Pyper and Peterman 1998). In addition, it is well-known that ignoring spatial autocorrelation leads to biased estimates and standard errors for the estimation (Albert and McShane 1995, Keitt et al. 2002).

Standard statistical methods such as correlation analysis, therefore, might not capture the true relationships among variables in the terrestrial ecosystem. The geostatistical approach, which takes into account spatial autocorrelation, is the chosen analytical tool for the questions that will be presented for the current study (Webster and Oliver 2001). Another reason geostatistical analyses are desirable for environmental samples is that traditional statistics lose the information that spatially distributed samples would represent. For example, the mean total microbial abundance (especially with sizable amount of standard deviation) of a field would not be very useful in providing total microbial abundance in certain regions of the field.

Unlike most soil characteristics measurements that were measured as continuous variables, microbial community structures were measured as multivariate binary data using non- or semi-quantitative genetic fingerprinting techniques. The definition of semivariance is the half of the squared difference between two samples separated by a lag distance. This definition has been conceptually extended to the dissimilarity of a lag distance (Franklin and Mills 2003). This definition has been used and accepted in microbial ecology field (Franklin et al. 2002, Lilleskov et al. 2004, Mummey and Stahl 2003, Ritz et al. 2004), however the validity of the approach has never really been explored.

In this paper, we adopted the relative dissimilarity in spatial structure analysis of soil microbial communities and evaluated the approach by comparing it with the semivariogram in the current study. The spatial structure is studied at multiple scales with a systematically nested sampling design from an experimental field with a distinct vegetation cover to provide comprehensive understanding. Two continuous variables of soil chemistry (total nitrogen content and C/N ratio) are used in evaluating the proposed approach.

MATERIALS AND METHODS

Site Description and Sampling Design

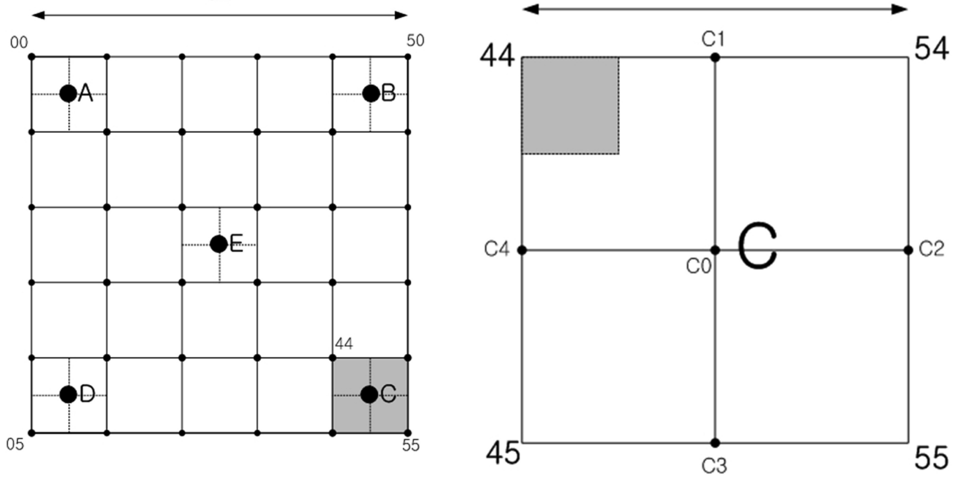
The study was conducted on a field of approximately 14 ha at the Blandy Experimental Farm in the northern Shenandoah Valley of Clarke County, Virginia (78°06 W, 39°06 N), a meadow comprised of native temperate tallgrass species, which is a part of the collection of the Orland E. White Arboretum (Kang and Mills 2004).

The sampling plot was placed at a rather sharp distinction between a nearly monospecific stand of *Andropogon gerardii* and an area with 3 dominant species (*Andropogon virginicus*, *Bouteloua curtipendula*, *Schizachyrium scoparium*) occurs. A spatially hierarchical design in which soil samples were collected at three different distance increments (0.05m, 0.25m and 1m) was used to reveal spatial structure (Figure 1). Each node indicated as a black dot represents a single sampling point. In the Level 1 grid, the nodes were at 1-m intervals so that 36 samples were collected. Each node was named following a column-row (CR) naming rule beginning with 00 at the upper-left corner, moving to 50 in the upper right, 05 in the lower left, and finishing with 55 at bottom right corner. From the 25 Level-1 grids, the 4 corners (indicated as A, B, C and D) and center (E) were chosen for Level-2 sampling. From each grid, 5 samples were collected, as shown in the diagram at 0.5m-intervals for a total of 25 samples at Level 2. Sampling points were named by using the grid id and number, and the indicating sampling position within the grid beginning with 1 at the top and rotating clockwise; 0 indicated the center of the grid. From each Level-2 corner grid (A, B, C and D), a 25-cm by 25-cm square located closest to the center of the Level-1 plot was chosen for Level-3 sampling. Similar to the Level-1 design, a total of 36 sampling points were established at 0.05-cm intervals. One point from each grid had already been collected during the Level-1 sample (for example, Ca00 from Level 3 is the same point as 44 in the Level 1 matrix). Thus, the total number of samples at Level 3 (including those sampled at Level 1) was 144.

Using this design, the spatial autocorrelation of the soil microbial community structure, the relationship of the communities with different overlying plant species, and the influence of soil characteristics on the microbial community structure can be analyzed simultaneously. A total of 199 samples were collected by using a small spatula of approximately 2-cm diameter and 7-cm depth (Scoopula®), because the available soil corer was too large for the Level 3 sampling grids (0.05 m separation). Plants species within a 1.5-cm radius of the sampling point were recorded at the time of the sampling.

Microbial Community Structures Analysis

Whole community DNA from the soil samples were extracted using UltraClean™ soil DNA isolation kit (MoBio Laboratories, Inc., Solana Beach, CA) with a


(A) Level I sampling

(B) Level II sampling

(C) Level III sampling

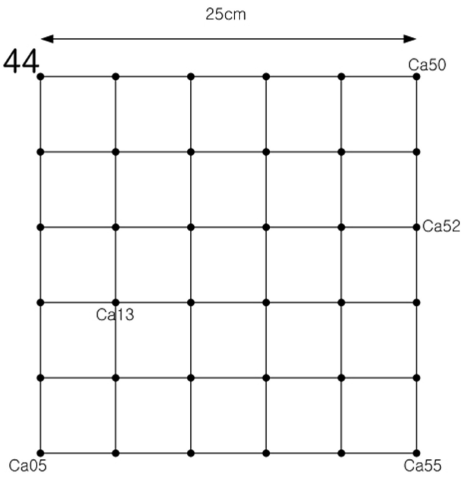


Figure 1. Sampling design used to assess spatial structure of soil microbial community. Level II and III are nested within level I and II sampling design, respectively. The vertical line in level I plot indicates the separation between X1 plot (*A. gerardii* monoculture) and X4 plot (mixture of several species).

slight modification of the manufacture's protocol. DGGE (Heuer and Smalla 1997, Muyzer and Smalla 1998) was the main genetic fingerprinting technique used to study soil microbial community structures throughout this research. The 16S rDNA fragments were obtained by the amplification of 20 to 40 ng of template DNA with a universal primer set for soil bacteria P63f (5'-GC clamp-CAGGCCTAACACATGCAAGTC-3') and P518r (5'-ATTACCGCGGCTGGCTGG-3') (El Fantroussi et al. 1999). A GC clamp of 40 bases was added to prevent complete separation of double stranded DNA and loss of the sequences from the gel (Sheffield et al. 1989). The specific protocols of PCR and the following quantification and polyacrylamide gel electrophoresis were as described previously (Kang and Mills 2004).

In addition to bacteria, the fungal community was studied by using FF390 (5'-CGATAACGAACGAG-ACCT-3') and FR1 (5'-GC clamp-AICCATT-CAATCGGTAIT-3') (Vainio and Hantula 2000). The FF390 and FR1 primer set is designed to amplify 390 bp fragment of 18S rDNA near the 3' end. Evaluation by FastA searches in the GenBank database revealed the matching of high numbers of fungi (Vainio and Hantula 2000). The PCR reaction mixture contained 0.5 mM primers, 2.5 mM $MgCl_2$ 200 mM dNTP mixture, PCR buffer, and two units of AmpliTaq® DNA polymerase.

Amplification was performed as follows: initial denaturation of 8 minutes at 95°C, and 30 cycles of 30 seconds at 95°C, 45 seconds at 50°C, and 2 minutes at 72°C, with a final elongation of 10 minutes at 72 °C. Amplified 18S rDNA fragments were separated on a polyacrylamide gel with a denaturing gradient of 40 to 55% of urea-formamide denaturant. The electrophoretic separation was run for 4 hours at 175V, and the product's visualization was the same with that of the bacterial analysis. Although the original protocols suggested a much longer running time (16-18 hours) at lower voltages (50-75 V), the preliminary running did not provide improved results. The protocols for fungal community PCR and acryl-amide gel electrophoresis were empirically optimized based on those presented by Kowalchuk et al. (2003) and Vainio and Hantula (2000).

Spatial Structure Analysis (Semivariogram)

Geostatistics permit a researcher to incorporate the spatial coordinates of observations into the data set. It then allows the describing of and the modeling of spatial patterns, predicting unobserved values, and assessing the uncertainty of the estimations and predictions. The semivariogram is the beginning step in a series of geostatistical procedures that depict and model spatial structures (spatial continuity of the observations or spatial autocorrelation) (Isaaks and Srivastava 1989). The semivariogram is defined as the semivariance of the increment and is formulated:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}[z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2 \qquad (1)$$

where **h** is the separation distance between any pair of points, also referred to as the lag distance and z is the value of the variable **u** at the sampling location i. Three parameters are normally determined to characterize the semivariogram model: nugget (C_0), sill (C_1) and range (a). Nugget represents the variability which cannot be modeled using a spatial autocorrelation function, i.e., amount of variability at lag distance = 0. Sill is the maximum variance of the variogram, which is the sum of the nugget and the structure variance, C+C_0. Range is the lag distance at which sill is reached. The proportion of sill explained by structure variance, $C/(C+C_0)$, is called the normalized measure of spatial dependence (NMSD) and could indicate the degree of spatial structures of certain models.

The formula (1) can be verbally described as the average dissimilarity between observations separated by vector **h**. In this context, semivariograms of microbial community structures can be created using the 'relative dissimilarity' values calculated from the Jaccard similarity matrix (dissimogram), because DGGE generated multivariate binary data, which cannot be used to create semivariograms in the traditional sense (Franklin and Mills 2003, Lilleskov et al. 2004, Mackas 1984, Meyers et al. 2013, Mistral et al. 2000, Mummey and Stahl 2003).

$$\gamma'(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}[1-J] \qquad (2)$$

where J is the Jaccard similarity coefficient. Although some geostatistical techniques for categorical variables, such as indicator semivariogram and indicator kriging, have been proposed, these techniques are designed for indicator analyses in the basis of estimation of proportion of certain '0's and '1's (in the case of a binary variable). Hence, it would not be very useful in dealing with microbial community composition data, because what matters are the patterns of '0's and '1's in each sample, not the proportions of them in defining identities of each microbial community.

Spatial structure analysis was performed to understand spatial structures of bacterial and fungal communities for comparison to similar measures for the soil characteristics. There were two main purposes of the analysis. First, the overall spatial structure of the sampled field at different scales could be determined. This is necessary due to the efforts of balance between smooth representations with maximum resolution. Simply it is not easy or very intuitive to run semi-variogram modeling analysis for the entire field (7.07 m scale) with a minimum possible lag distance (0.025 m). Second, spatial structures at different locations within the experimental field were described. This is important because the field had distinct separations in plant community composition. For the first purpose, the entire data set was used for an average portrait of the field and the subsets of the data structured at different distance scales were analyzed separately. Therefore, 4 different data sets were available: entire field scale (lag distance up to 7.07 m), medium scale (up to 2.5 m), small scale (up to 1 m) and fine scale (up to 0.25 m). For the second purpose, data from different regions, especially where plant development was distinct (Mono vs. Mixed), were analyzed independently. This approach was used to understand the effect of plants on the spatial structures of microbial communities. It was also used to reveal possible hidden dominant controllers of the microbial community structure other than the plants in the experimental field. In addition to that, four corners with two different scales were also analyzed: fine scale (0.25 m × 0.25 m) and medium scale (2.5 m × 2.5 m). Data were subsampled with new coordinate systems with the lower left corner of each grid being the origin (0, 0).

Spatial structure analysis is initiated by determining the lag distance for all the pairs of the sampling points. Rather than dealing with individual values for lag distances, they are usually distributed into groups (bins) of similar values. Binning, the determination of appropriate lag distance groupings, is a critical part of semivariogram analysis in that it decides the number of bins (lags) and resolution of the semivariogram model. The object of binning was, therefore, obtaining more bins with relatively constant pairs of samples in each of the bins. Multiplication or division of grid spacing was tested for the best representation from all levels of data, and the one with the smaller number of bins possessing all the features was chosen. By looking at the h-scattergrams, some obvious outliers were eliminated for the final semivariogram modeling. To minimize misinterpretation, lags with less than 1% of total pairs were eliminated for the semivariogram modeling.

As one can expect, not all of the sampling points are located on the exact lag spacing without a strict gridded sampling design. Sampling design for this current study employed a grid system with different levels of grid in the field, making it impossible for all the sampling locations to fall into each lag. To solve this prevalent problem, tolerance in both distances and the direction are incorporated in the semivariogram analysis (Isaaks and Srivastava 1989). Distance tolerance (or lag tolerance) was determined to include statistically meaningful number of pairs for calculation. Distance tolerance for the analysis was set to follow the suggestion of Isaaks and Srivastava (1989) to be half the lag distance (Goovaerts 1998). Due to distance tolerance, the actual lag distance varied as it was determined by the average of the distance of all the pairs within the tolerance range.

For the semivariogram modeling in this study, the suggestion of Isaaks and Srivastava (1989) along with the procedures from the software manual for VARIOWIN 2.2 (Pannatier 1996) were used. Modeling began with an omni-directional experimental semivariogram ($\hat{\gamma}(\mathbf{h})$). In VARIOWIN, the nugget was first determined and then the other component models (nested models in VARIOWIN) were chosen based on the best model fit and the principle of parsimony. Modeling with many components model takes more resources without providing significant benefits, and will often generate a model that is more complex than can be explained by the data. One nice feature of VARIOWIN is the existence of a numeric parameter to determine goodness of fit: Indicative Goodness of Fit (IGF) (Goovaerts 1998, Pannatier 1996). The IGF feature is particularly useful in determining structural variance (*C*), which is usually more difficult to estimate reliably compared to the other parameters like range (*a*) and nugget (C_0) (Isaaks and Srivastava 1989). The IGF measures the sum of the squares of the differences between experimental semivariances ($\hat{\gamma}(i)$) and model semivariances ($\gamma(i)$) and replaces traditional visual fits while enabling the testing of several possible models to find the best fit, quantitatively.

In the actual analysis of microbial community structure, .var files (experimental semivariograms) were prepared from the experimental semivariogram analysis based on the dissimilarity, and then modeling was performed by using the Model module of VARIOWIN. A Simple program was prepared using Matlab 6.5 for calculating experimental semivariogram between the distance matrix and the dissimilarity matrix. For regional analyses, separate matrices of dissimilarity, based on the Jaccard coefficient, were prepared.

Evaluation of the Spatial Structure Analysis

Relative dissimilarity, based on the Jaccard similarity coefficients for microbial community structure data, was calculated between pairs of samples and used to construct experimental semivariograms. Although this approach has a historical background as the dissimilarity-area relationship, there has not been any evaluation of the usage of it for a formal geostatistical analysis. Researchers did simple experimental evaluation by calculating continuous variables from the same sample sets as that of the current study (total nitrogen content and C/N ratio) in both the semivariogram and the dissimogram (semivariogram based on relative dissimilarities).

All the conditions were the same between the two approaches for the lag distances of the fine scale (2 cm) and the small scale (10 cm), using the same lag distances for both measurements. The Total nitrogen content used was 70 cm and the C/N ration used was 50 cm in the plot scale analysis. The Medium scale lag distance was set differently for comparison: The semivariogram used 25 cm and the relative dissimilarity used 30 cm. Lag tolerance is half of the lags and angular tolerance is set to 90° because both variables are isotropic.

RESULTS

Spatial Structure of Microbial Communities

Overall patterns of dissimogram models based on the relative were almost identical within certain scales ranges – the field and medium scales were similar, as were the small and fine scales (Figure 2 and Figure 3).

The four larger scale models (field and medium scale), in particular, were almost identical when the scales were ignored. The degree of anisotropy in bacterial and fungal community structures displayed by the field models was supported by comparing the semivariogram models obtained for the experimental field's two sections. The field was comprised of two sections: the Mono section was close to a monoculture of *A. gerardii*, whereas the Mixed section contained a diverse collection of vegetation. The shapes and parameters of the variogram models for the two zones were virtually identical for the bacterial community structures. However, compared to the bacterial community, the spatial structure of the fungal communities showed more distinction between the two zones (Table 1). The summarizing effect, which was found in the field scale models over the region models in terms of a better fit of the model, was apparent with microbial community structures as it was with all other measurements.

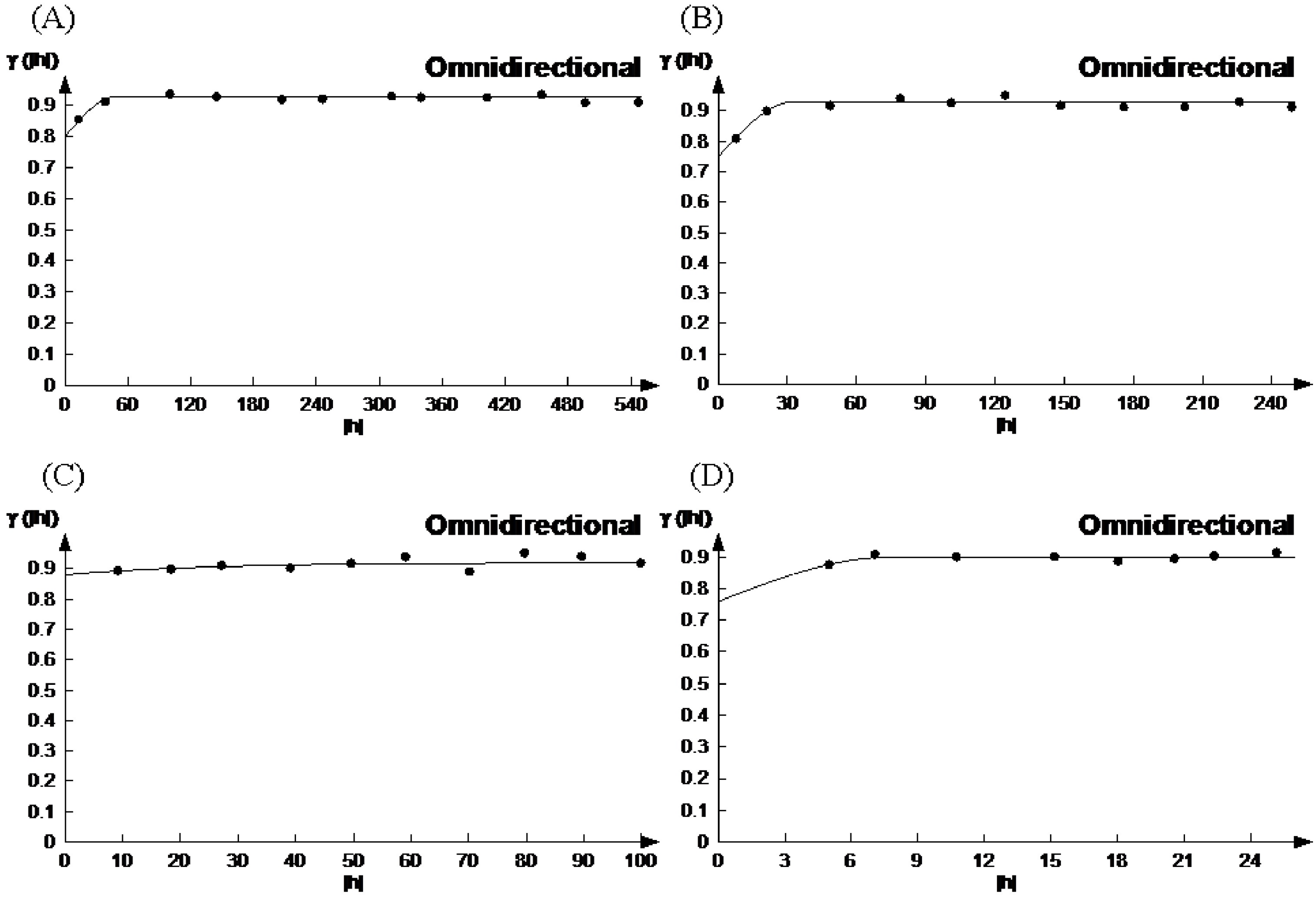


Figure 2. Dissimogram models of bacterial communities in different scales calculated by relative dissimilarity. X axis is lag distance in cm and Y axis is semivarience. A: Field scale, all samples. B: Medium scale, lag distances less than 2.5 m. C: Small scale, lag distances less than 1 m. D: Fine scale, lag distances less than 0.25 m.

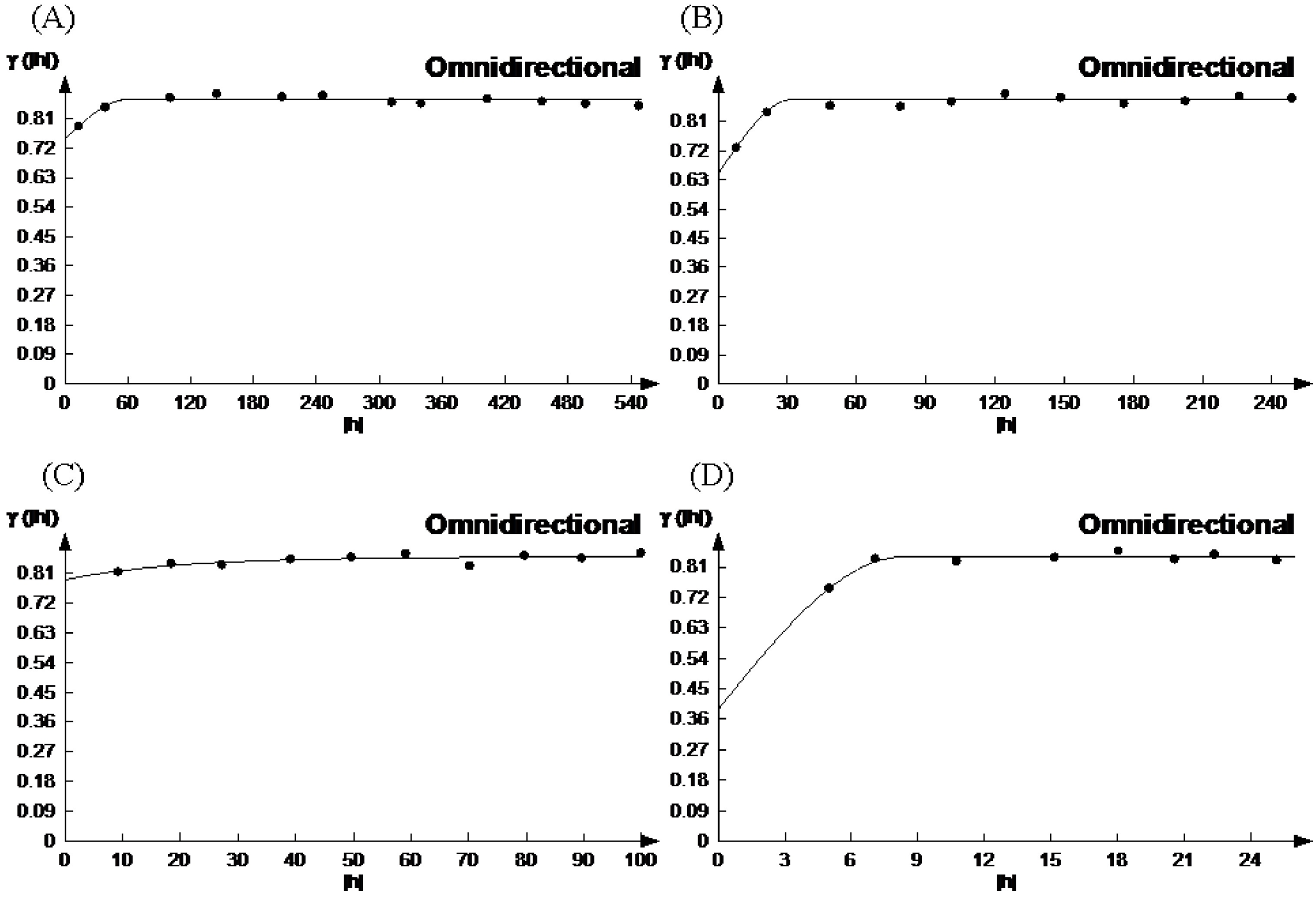


Figure 3. Dissimogram models of fungal communities at different scales calculated by relative dissimilarity. Scales are defined as in Figure 3.

Spatial structures of bacterial communities in the Mono and Mixed sections of the experimental field were very similar to each other. Except for the nugget, which also had a trivial difference, all other parameters (sill, range and NMSD) were identical. Although the region models of the fine scale showed some variances among them, the difference was trivial compared to the previously analyzed variables (Kang 2005). The A-corner region model was most similar to the field model in both overall shape and model parameters (figure not shown). The ranges of the region models at the fine scale were compatible to each other and the field model (7.5 – 10.4 cm). Spatial dependences of the region models were still very low (0.05 – 0.21).

Spatial structures of the Mono and the Mixed sections of the experimental field were less similar to each other for the fungi when compared to the ones from bacterial communities. Both were spherical models and had similar total sill ($C_0 + C$), but the nugget, range, and spatial dependence were distinctive. The region models at the fine scale were more similar to their field model compared to that of the bacterial communities, and the A-corner region model was most similar to the field model, just like that of the bacterial communities. The range of the Mono section was 0.65 m and the Mixed section was 0.5 m. The ranges of region models at the fine scale were compatible when compared with the same types of models, but the B-corner model, which was the only exponential model at the fine scale, showed a much larger range (0.12 m). The spatial dependences were slightly higher than those of bacterial communities in comparison with corresponding models, but the fine scale models' spatial dependence was higher than the others' (0.29 – 0.63).

Evaluation of Relative Dissimilarity Approach

Patterns of semivariogram models between the conventional semivariogram approach and the approach based on relative dissimilarity did not seem to conflict at first sight as all the models were spherical models, and the ranges were compatible between approaches even in

Table 1. Summary of results from dissimogram analysis of microbial community structure relative dissimilarity.

Factor	Scale	Corner	Model[a]	IGF[b]	Nugget (C_0)	Sill (C_1)	Range (a)[c]	NMSD
Bacteria	Field	Entire	1	2.38E-4	0.8	0.13	49.5	0.14
	Medium	Entire	1	2.49E-4	0.75	0.18	32.5	0.19
		Mono[d]	1	6.17E-4	0.8	0.11	45	0.12
		Mixed[d]	1	3.84E-4	0.82	0.11	45	0.12
	Small	Entire	2	1.51E-4	0.88	0.04	70	0.04
	Fine	Entire	1	7.64E-5	0.76	0.14	7.8	0.16
		A	1	1.29E-4	0.75	0.11	7.54	0.13
		B	1	1.48E-3	0.71	0.19	8.32	0.21
		C	1	1.05E-4	0.83	0.1	7.8	0.11
		D	1	3.69E-4	0.86	0.05	10.4	0.05
Fungi	Field	Entire	1	1.65E-4	0.75	0.12	60.5	0.14
	Medium	Entire	1	2.52E-4	0.65	0.23	32.5	0.26
		Mono	1	2.69E-4	0.72	0.14	65	0.16
		Mixed	1	2.27E-4	0.77	0.11	50	0.13
	Small	Entire	2	1.41E-4	0.79	0.07	58	0.08
	Fine	Entire	1	1.71E-4	0.39	0.45	8.32	0.54
		A	1	3.57E-4	0.38	0.46	8.06	0.55
		B	2	8.23E-5	0.6	0.25	12.22	0.29
		C	1	5.37E-4	0.49	0.41	8.58	0.46
		D	1	1.01E-3	0.29	0.5	8.58	0.63

[a] Following convention used in both VARIOWIN: 1. Spherical model, 2. Exponential model, 3. Gaussian model, 4. Power model. For power model, slope and power are recorded under sill and range, respectively and NMSD was not calculated for power model either. By definition, the range of power model was beyond the limit of each scale. 0 is used for pure nugget model.

[b] Indicative goodness of fit [c] Unit of range is cm.

[d] Mono is the left portion of the field with *A. gerardii* as dominant and Mixed is the right portion of the field with diverse plants as co-dominant.

Table 2. Summary of results from semivariogram analysis and relative dissimilarity analysis for total nitrogen content and C/N ratio.

Variable	Scale	Method[a]	ID in Figure 4	IGF	Nugget (C_0)	Sill (C_1)	Range (a)	NMSD
Nitrogen	Field	Semi	A	3.10E-2	41.9	34.1	84 cm	0.45
		Dissim	B	1.31E-2	6.4	3.6	61.6 cm	0.36
	Medium	Semi	C	1.63E-1	23.43	60.35	47.5 cm	0.72
		Dissim	D	2.98E-2	4.8	5.9	43.2 cm	0.55
	Small	Semi	E	1.39E-1	22.7	71	52 cm	0.76
		Dissim	F	3.65E-2	5.8	5.1	50 cm	0.47
	Fine	Semi	G	2.73E-3	2.84	53.3	16.38 cm	0.95
C/N ratio	Field	Semi		1.95E-1	0	0.348	49.5 cm	1.00
		Dissim		2.65E-2	0.31	0.35	66 cm	0.53
	Medium	Semi		2.23E-1	0.044	0.352	70 cm	0.89
		Dissim		1.14E-2	0.26	0.43	60 cm	0.62
	Small	Semi		5.50E-2	0.024	0.344	61 cm	0.93
		Dissim		2.46E-2	0.28	0.42	61 cm	0.60
	Fine	Semi		4.44E-3	0.0040	0.136	13 cm	0.97
		Dissim		6.24E-4	0.17	0.26	12.48 cm	0.61

All semivariograms were modeled as spherical model. Medium scale models were analyzed in different lag distance.

[a] Semi: seminvarience, Dissim: relative dissimilarity

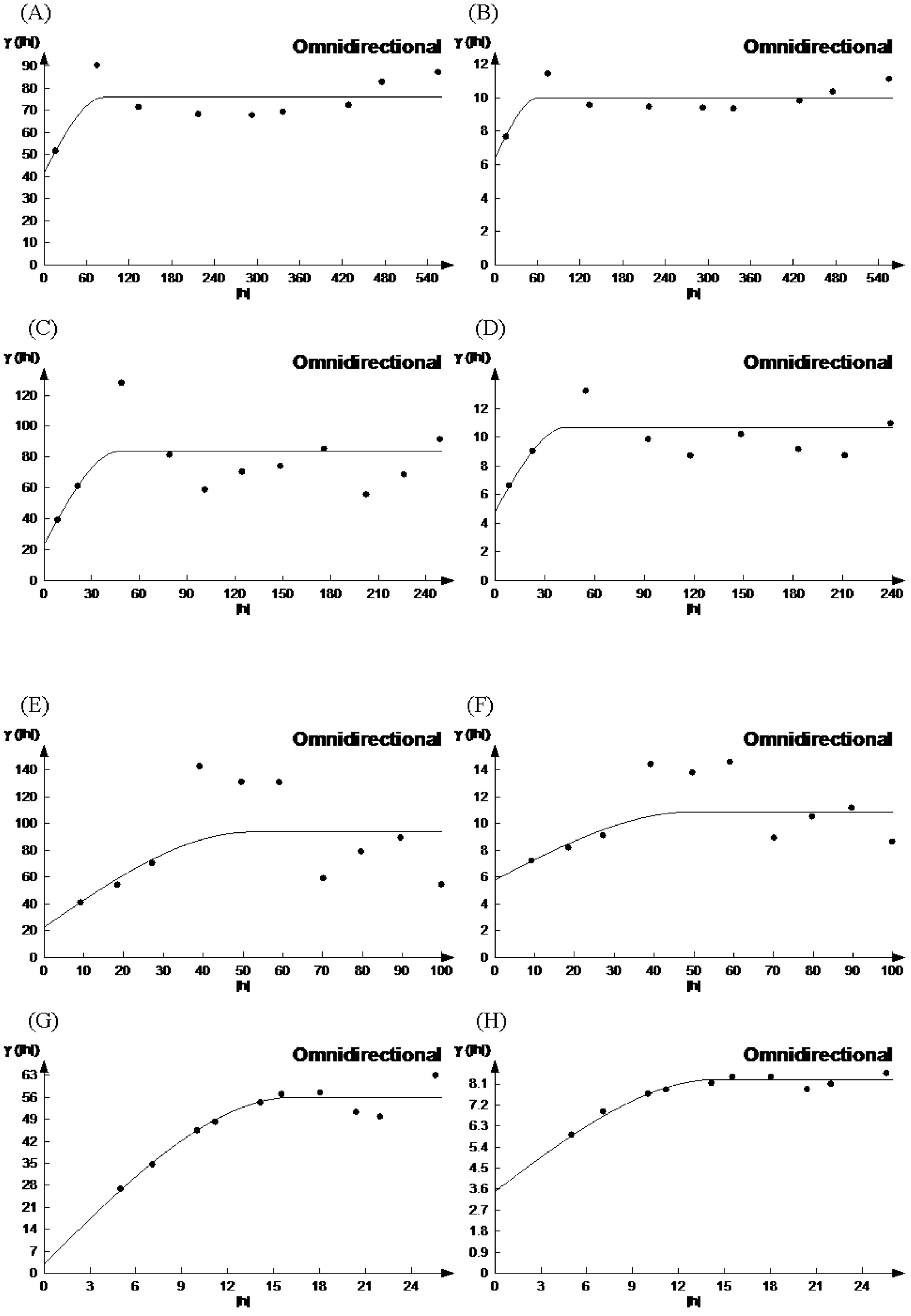


Figure 4. Semivariogram models (left column) and dissimogram models (right column) based on relative dissimilarity in comparison according to different scales of analyses. A,B: Field scale. C,D: Medium scale. E,F: Small scale. G,H: Fine scale. Scales are defined in Figure 3.

a medium scale, which was the one analyzed with different lag distances (Table 2 and Figure 4). However, substantial differences were identified in models for the two approaches in spatial dependence (NMSD): the models by the dissimogram were significantly less spatially structured than the the semivariogram models. The model fittings were better in the analysis by the dissimogram. The overall trends were that the shapes of the experimental semivariograms and models were very similar for the two methods, but the relative position of models at |h|=0 (nugget) were different.

DISCUSSION

While closely related soil characteristics and microbial abundance were determined to be spatially structured (Kang 2005), the microbial communities were analyzed to reveal their spatial structure. As discussed previously, spatial structure analysis was done using relative dissimilarity based on the Jaccard coefficient. Since relative dissimilarity could not be used for the estimation of an unsampled area of the experimental field, PCA (principal component analysis) was used to transform multivariate binary data into continuous variables (Franklin and Mills 2003). However, experimental semivariograms of the first three PCs were too noisy to be modeled and used for estimation (kriging or stochastic simulation). In addition, the sum of explainable variances from the first three PCs was only about 15%, meaning that the obtained PCs were not useful for prediction analysis.

The spatial structures of bacterial and fungal communities were very similar in the experimental field for the current study. But the response to the vegetation setting was a bit different between them as well. Several studies using PLFA (Phospholipid fatty acid) profile indicated certain distinctions between bacterial and fungal communities in their community structure along with different factors (Burke et al. 2003, Clegg et al. 2003, Grayston et al. 2004). For example, Clegg et al., (2003) used PLFA to measure the difference in fungal and bacterial community structures under different soil management treatments on a grassland in the United Kingdom. They found that the most abundant PLFA signals for fungal and bacterial communities differed by management regimes. Pennanen et al., (2004) used DGGE with two separate primer sets for bacterial and fungal communities to see the effect of amendment with a high quality substrate (L-asparagine), and found that the changes in bacterial community structure were more drastic than for that of the fungi, suggesting that the bacterial communities were more responsive to the asparagine. Ranjard et al., (2003) showed the difference of sample size effect on the community structure study between bacterial and fungal community using ARISA (automated ribosomal intergenic spacer analysis), and the overall patterns of different sample sizes on a PC plot were very different between bacterial and fungi. As was found in the present study, others also have shown similar overall responses for the two microbial components (Girvan et al. 2004, Sanyal and Kulshrestha 1999). Sanyal and Kulshrestha (1999) studied herbicide degradation capability between bacterial and fungal communities in soil, and showed that repeated treatment induced adaptation of microbial communities equally for both bacteria and fungi in that both communities were able to degrade over 80% of the applied herbicide. There is a certain difference between bacteria and fungi as was discussed, but it seems that what matters are the methods and the systems in which those differences were represented or not.

The physiology of bacteria and fungi is very different as they are classified in different domains; Bacteria vs. Eukarya (Madigan and Martinko 2006). Their relationships are a little more complex in an ecological aspect. Their habitats are generally segregated due to their size differences, differentiating other aspects – living conditions. Since bacteria reside within aggregates (Foster 1988), they are generally protected from grazing (Heijnen et al. 1993, Postma et al. 1990, Young et al. 1994) and catastrophic events such as, complete drought (van Gestel et al. 1996). Thus, bacterial biomass is relatively stable, and as a result, bacterial communities experience starvation most of the time (Morita 1990) due to limited mobility (Bass et al. 1998). Although the abundance of bacteria is at least several orders of magnitude higher than that of fungi in most soil ecosystem (Foster 1988, Polyanskaya et al. 2000, Tiquia et al. 2002), their relationship in biomass is more variable (Frey et al. 1999, Frostegård and Bååth 1996). The ratios of fungal to bacterial biomass are reported as 1% (Bloem et al. 1994) to 42% (Yang and Insam 1991). In some soil; however, fungal biomass dominates soil biomass, such that it exceeds the biomass of all other constituents in the soil, including bacteria, plants and animals (Anderson and Domsch 1978, Kjøller and Struwe 1982, Schmidt and Bolter 2002). This inconsistency of the relationship between bacterial and fungal communities lead to the conclusion that, while there are always differences between bacteria and fungi, whether or not the differences are observable in any particular study is dependent on the analytical methods used for that study and the systems in which the studies were done.

Nugget effects were apparent in semivariograms describing microbial communities, and the effect was especially prevalent in community composition analyses (~86% of the sill in field scale models) than for the abundance. Similar results that have been obtained throughrelative dissimilarity have been reported by others (Franklin et al. 2002, Franklin and Mills 2003, Lilleskov et al. 2004, Mummey and Stahl 2003). However, it is not known for sure whether the high nuggets are the properties of the distribution of microbial community structure, or whether they are artificially generated in the current approach of relative dissimilarity. The Mantel test results show that the association between microbial community structures and distances among communities was significant ($r = 0.347$, $P < 0.001$ for bacteria, $r = 0.461$, $P < 0.001$ for fungi), thus demonstrating the benefit of employing geostatistics on microbial community structure analysis. but the results from the dissimogram analysis seems to suggest the opposite. Even if high nuggets might simply indicate experimental error or high heterogeneity existing below the sampling resolution (5 cm), the general trends between the semivariogram and the dissimogram of the total nitrogen concentration and C/N ratio could raise a question. As discussed earlier, a high nugget indicates low spatial dependence and a less advantage of using geostatistics over conventional statistics that extract inference out of spatially distributed data.

The results of the comparisons between the two methods of spatial structure analysis confirmed the question that arose from the analysis of bacterial and fungal community structures based on their relative dissimilarity. The extremely good fit of the model and the very low spatial dependence were two features found from the comparisons between the semivariogram and the dissimogram. It can be concluded that at least a certain degree of higher nuggets found in the models of bacterial and fungal communities might be artificially generated from the spatial structure analysis based on the relative dissimilarity rather than from their intrinsic spatial structure. Therefore, one should pay extra attention to the interpretation of spatial structures of microbial communities based on relative dissimilarity measurements. The semivariogram modeling with relative dissimilarity might still be legitimate in the sense of spatial structure analysis that mainly focuses on the range of spatial autocorrelation (Nekola and White 1999), while the employment of further prediction analyses (e.g., kriging or stochastic simulation) based on the results of a semivariogram modeling might require specific justification.

Since microbial community structure data are formed as multivariate binary data, a multivariate generalized prediction model would be a natural and sensible method for studying spatial structure and prediction. Hence, it should be valuable, promising work in microbial ecology and statistics to develop a suitable model for this problem based on those suggestions made above. In addition, it would also be worthwhile to consider the time effect or trend over a given region of interest by using a multivariate space-temporal model.

ACKNOWLEDGEMENTS

We thank Aaron Mills for helpful critical review of this manuscript. SK and JJS were supported by start-up funds from the Office of Vice Provost of Research at Baylor University.

REFERENCES

Albert, P.S. and McShane, L.M. 1995. A generalized estimating equations approach for spatially correlated binary data: an application to the analysis of neuroimaging data. Biometrics 51: 627-638.

Anderson, J.P.E. and Domsch, K.H. 1978. Mineralization of bacteria and fungi in chloroform-fumigated soils. Soil Biology & Biochemistry 10: 207-213.

Bass, C.J.; Davey, R.A. and Lappin-Scott, H.M. 1998. Long-term starvation survival of a thermophilic sulfidogen consortium. Geomicrobiology Journal 15: 29-36.

Bloem, J.; Lebbink, G.; Zwart, K.B.; Bouwman, L.A.; Burgers, S.L.G.E.; Devos, J.A. and Deruiter, P.C. 1994. Dynamics of microorganisms, microbivores and nitrogen mineralization in winter wheat fields under conventional and integrated management. Agriculture Ecosystems & Environment 51: 129-143.

Burke, R.A.; Molina, M.; Cox, J.E.; Osher, L.J. and Piccolo, M.C. 2003. Stable carbon isotope ratio and composition of microbial fatty acids in tropical soils. Journal of Environmental Quality 32: 198-206.

Clegg, C.D.; Lovell, R.D.L. and Hobbs, P.J. 2003. The impact of grassland management regime on the community structure of selected bacterial groups in soils. FEMS Microbiology Ecology 43: 263-270.

El Fantroussi, S.; Verschuere, L.; Verstraete, W. and Top, E.M. 1999. Effect of phenylurea herbicides on soil microbial communities estimated by analysis of 16S rRNA gene fingerprints and community-level physiological profiles. Applied and Environmental Microbiology 65: 982-988.

Foster, R.C. 1988. Microenvironments of soil microorganisms. Biology & Fertility of Soils 6: 189-203.

Franklin, R.B.; Blum, L.K.; McComb, A.C. and Mills, A.L. 2002. A geostatistical analysis of small-scale spatial variability in bacterial abundance and community structure in salt marsh creek bank sediments. FEMS Microbiology Ecology 42: 71-80.

Franklin, R.B. and Mills, A.L. 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. FEMS Microbiology Ecology 44: 335-346.

Frey, S.D.; Elliott, E.T. and Paustian, K. 1999. Bacterial and fungal abundance and biomass in conventional and no-tillage agroecosystems along two climatic gradients. Soil Biology & Biochemistry 31: 573-585.

Frostegård, Å. and Bååth, E. 1996. The use of phospholipid fatty acid analysis to estimate bacterial and fungal biomass in soil. Biology & Fertility of Soils 22: 59-65.

Girvan, M.S.; Bullimore, J.; Ball, A.S.; Pretty, J.N. and Osborn, A.M. 2004. Responses of active bacterial and fungal communities in soils under winter wheat to different fertilizer and pesticide regimens. Applied and Environmental Microbiology 70: 2692-2701.

Goovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biology & Fertility of Soils 27: 315-334.

Grayston, S.J.; Campbell, C.D.; Bardgett, R.D.; Mawdsley, J.L.; Clegg, C.D.; Ritz, K.; Griffiths, B.S.; Rodwell, J.S.; Edwards, S.J.; Davies, W.J.; Elston, D.J. and Millard, P. 2004. Assessing shifts in microbial community structure across a range of grasslands of differing management intensity using CLPP, PLFA and community DNA techniques. Applied Soil Ecology 25: 63-84.

Heijnen, C.E.; Chenu, C. and Robert, M. 1993. Micro-morphological studies on clay-amended and unamended loamy sand, relating survival of introduced bacteria and soil structure. Geoderma 56: 195-207.

Heuer, H. and Smalla, K. 1997. Application of denaturing gradient gel electrophoresis and temperature gradient gel electrophoresis for studying soil microbial communities. Pages 353-373, In: Elsal, J.D.v.; Trevors, J.T. and Wellington, E.M.H. (Editors) Modern Soil Microbiology. Marcel Dekker, New York, NY.

Isaaks, E.H. and Srivastava, R.M. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, NY. 592 pages.

Kang, S. 2005. Spatial Structure of Soil Microbial Communities and Their Controlling Factors. Thesis. Department of Environmental Sciences. University of Virginia, Charlottesville, VA. 253 pages.

Kang, S. and Mills, A.L. 2004. Soil microbial community structure changes following disturbance of the overlying plant community. Soil Science 169: 55-65.

Keitt, T.H.; Bjørnstad, O.N.; Dixon, P.M. and Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions. Ecography 25: 616-625.

Kjøller, A. and Struwe, S. 1982. Microfungi in ecosystems - fungal occurrence and activity in litter and soil. Oikos 39: 391-422.

Kowalchuk, G.A.; Os, G.J.; Aartrijk, J. and Veen, J.A. 2003. Microbial community responses to disease management soil treatment used in flower bulb cultivation. Biology & Fertility of Soils 37: 55-63.

Legendre, P. and Troussellier, M. 1988. Aquatic heterotrophic bacteria: Modeling in the presence of spatial autocorrelation. Limnology and Oceanography 33: 1055-1067.

Liebhold, A.M. and Sharov, A.A. 1998. Analysis of insect count data: testing for correlation in the presence of spatial autocorrelation., pages 111-117, In: Braumgartner, J.; Brandmayer, P.; Manly, B.F.J. (Editors) Population and Community Ecology for Insect Management and Conservation. Swets & Zeitlinger Publishers, Balkema, Rotterdam.

Lilleskov, E.A.; Bruns, T.D.; Horton, T.R.; Taylor, D.L.; Grogan, P. 2004. Detection of forest stand-level spatial structure in ectomycorrhizal fungal communities. FEMS Microbiology Ecology 49: 319-332.

Mackas, D.L. 1984. Spatial autocorrelation of plankton community composition in a continental shelf ecosystem. Limnology and Oceanography 29: 451-471.

Madigan, M.T. and Martinko, J.M. 2006. Biology of Microorganisms, 11th edition. Pearson Prentice Hall, Upper Saddle River, NJ. 992 pages.

Meyers, L.M.; DeKeyser, E.S. and Norland, J.E. 2013. Differences in spatial autocorrelation (SAc), plant species richness and diversity, and plant community composition in grazed and ungrazed grasslands along a moisture gradient, North Dakota, USA. Applied Vegetation Science 17: 53-62.

Mistral, M.; Buck, O.; Meier-Behrmann, D.C.; Burnett, D.A.; Barnfield, T.E.; Scott, A.J.; Anderson, B.J. and Wilson, J.B. 2000. Direct measurement of spatial autocorrelation at the community level in four plant communities. Journal of Vegetation Science 11: 911-916.

Morita, R.Y. 1990. The starvation-survival state of microorganisms in nature and its relationship to the bioavailable energy. Experientia 46: 813-817.

Mummey, D.L. and Stahl, P.D. 2003. Spatial and temporal variability of bacterial 16S rDNA-based T-RFLP patterns derived from soil of two Wyoming grassland ecosystems. FEMS Microbiology Ecology 46: 113-120.

Muyzer, G. and Smalla, K. 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie Van Leeuwenhoek 73: 127-141.

Nekola, J.C. and White, P.S. 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26: 867-878.

Pannatier, Y. 1996. VARIOWIN: Software for Spatial Data Analysis in 2D. Springer-Verlag, New York, NY.

Pennanen, T.; Caul, S.; Daniell, T.J.; Griffiths, B.S.; Ritz, K. and Wheatley, R.E. 2004. Community-level responses of metabolically-active soil microorganisms to the quantity and quality of substrate inputs. Soil Biology & Biochemistry 36: 841-848.

Polyanskaya, L.M.; Orazova, M.K.; Burkanova, O.A. and Zvyagintsev, D.G. 2000. Microsampling of rhizosphere soil and laboratory artifacts. Microbiology 69: 581-585.

Postma, J.; Hokahin, C.H. and Veen, J.A.v. 1990. Role of microniches in protecting introduced rhizobium-leguminosarum *Biovar trifolii* agaginst competition and predataion in soil. Applied and Environmental Microbiology 56: 495-502.

Pyper, B.J. and Peterman, R.M. 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. Canadian Journal of Fisheries and Aquatic Sciences 55: 2127-2140.

Ranjard, L.; Lejon, D.P.H.; Mougel, C.; Schehrer, L.; Merdinoglu, D. and Chaussod, R. 2003. Samling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. Environmental Microbiology 5: 1111-1120.

Ritz, K.; McNicol, J.W.; Nunan, N.; Grayston, S.; Millard, P.; Atkinson, D.; Gollotte, A.; Habeshaw, D.; Boag, B.; Clegg, C.D.; Griffiths, B.S.; Wheatley, R.E.; Glover, L.A.; McCaig, A.E. and Prosser, J.I. 2004. Spatial structure in soil chemical and microbiological properties in an upland grassland. FEMS Microbiology Ecology 49: 191-205.

Sanyal, D. and Kulshrestha, G. 1999. Effects of repeated metolachlor applications on its persistence in field soil and degradation kinetics in mixed microbial cultures. Biology & Fertility of Soils 30: 124-131.

Schmidt, N. and Bolter, M. 2002. Fungal and bacterial biomass in tundra soils along an arctic transect from Taimyr Peninsula, central Siberia. Polar Biology 25: 871-877.

Sheffield, V.C.; Cox, D.R.; Lerman, L.S. and Myers, R.M. 1989. Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. PNAS 86: 232-236.

Tiquia, S.M.; Wan, J.H.C. and Tam, N.F.Y. 2002. Microbial population dynamics and enzyme activities during composting. Compost Sciences & Utilization 10: 150-161.

Vainio, E.J. and Hantula, J. 2000. Direct analysis of wood-inhabiting fungi using denaturing gradient gel electrophoresis of amplified ribosomal DNA. Mycology Research 104: 927-936.

van Gestel, M.; Merckx, R. and Vlassak, K. 1996. Spatial distribution of microbial biomass in microaggregates of a silty-loam soil and the relation with the resisitance of microorganisms to soil drying. Soil Biology & Biochemistry 28: 503-510.

Webster, R. and Oliver, M.A. 2001. Geostatistics for Environmental Scientists. John Wiley, West Sussex, UK. 330 pages.

Yang, J.C. and Insam, H. 1991. Microbial biomass and relative contributions of bacteria and fungi in soil beneath tropical rain forest, Hainan Islang, China. Journal of Tropical Ecology 7: 385-393.

Young, I.M.; Roberts, A.; Griffiths, B.S. and Caul, S. 1994. Growth of a ciliate protozoan in model ballotini systems of different particle sizes. Soil Biology & Biochemistry 26: 1173-1178.

Received 11 November 2013;
Accepted 21 January 2014