Supplement to Journal of Nematology 26(4S):626–634. 1994.

Applications of Geostatistics in Plant Nematology[1]

M. K. Wallace[2] and D. M. Hawkins[3]

Abstract: The application of geostatistics to plant nematology was made by evaluating soil and nematode data acquired from 200 soil samples collected from the A_p horizon of a reed canary-grass field in northern Minnesota. Geostatistical concepts relevant to nematology include semi-variogram modelling, kriging, and change of support calculations. Soil and nematode data generally followed a spherical semi-variogram model, with little random variability associated with soil data and large inherent variability for nematode data. Block kriging of soil and nematode data provided useful contour maps of the data. Change of support calculations indicated that most of the random variation in nematode data was due to short-range spatial variability in the nematode population densities.

Key words: change of support, contour mapping, geostatistics, kriging, nematode, plant-parasitic nematode, semi-variogram.

Nematode distribution in the field has been described as aggregated (4), which implies underlying spatial dependence in nematode data. Conventional statistics generally are inadequate to describe data that are spatially correlated. Regionalized variable theory, popularly known as geostatistics, is a methodology for the analysis of spatially correlated data (1). Three geostatistical concepts are relevant to the issue of spatial distribution of nematodes. These are semi-variogram modelling, kriging, and change of support calculations.

Geostatistical analyses were done on soil and nematode data acquired for a related study (7). Our objectives were to quantify the spatial dependence in the soil and nematode data at 10-m intervals by using semi-variogram modelling and kriging and to discuss the implications of change of support calculations on semi-variogram modelling.

Geostatistical Concepts

Semi-variogram: The semi-variogram defines the type and strength of spatial association. At one extreme, there may be no spatial association between measurements at any two points, implying independence of the data. At the other extreme, measurements may show a high degree of continuity between points, with the measurement at any point being highly predictive of those at nearby locations. Most practical phenomena fall between these extremes, showing some purely random variability and some spatial continuity (1). Spatial continuity manifests as a correlation between samples that diminishes as the distance between samples increases and vanishes completely when the distance is great enough that the samples are statistically independent. The distance at which samples become statistically independent is referred to as the range of influence of a sample.

A typical semi-variogram generally rises with increasing distance between samples, then varies about a constant value called the "sill" (1). The distance at which the semi-variogram reaches the sill is called the "range of influence of the sample" or more simply the "range." The sill of the semi-variogram is often characterized as equalling the overall variance from the mean of all sample data over the whole field, but this is only approximately true in moderately sized fields (1).

The semi-variogram quantifies the relationship commonly observed in the field that samples taken close together tend to have more similar values than samples taken farther apart. It is a plot of the semi-variance (one-half the mean squared difference) of paired sample measurements as a function of the distance (and sometimes the direction) between samples (6).

Received for publication 6 April 1993.
[1] Paper No. 20,460 of the contribution series of the Minnesota Agricultural Experiment Station based on research conducted under Project 22-14H.
[2] Former Research Assistant, Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108.
[3] Professor, Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108.

Usually all possible sample pairs are grouped in classes (lags) of approximately equal distance and direction.

The semi-variogram $\gamma(h)$ is defined by:

$$\gamma(h) = 1/2 \text{ the average of } [Z(x + h) - Z(x)]^2.$$

$Z(x)$ is the value of a variable measured at geographical location x. In geostatistical terminology, $Z(x)$ is called a "regionalized variable" (1). The semi-variogram depends on the distance h between x and $x + h$. It may also depend on the direction from x to $x + h$; if this is so, the semi-variogram is anisotropic. If $\gamma(h)$ depends on the distance h but not its orientation, the semi-variogram is isotropic. It is an assumption of variographics that the semi-variogram does not depend on x. In other words, the nature and strength of the relationship of the regionalized variable at any point and at any other point depend on the distance between the points, but not on where the pair of points is located within the field. Regionalized variables that do not satisfy this assumption require more sophisticated geostatistical methods than are discussed here.

Theoretically, the semi-variogram should pass through the origin because samples taken at exactly the same location have equal values; however, there is often a nonzero semi-variance as the distance tends to zero. This nonzero semi-variance is called the "nugget effect" (1). The nugget effect represents the degree of dissimilarity that can be seen between two measurements taken as close as possible to one another, e.g., between two samples from adjacent portions of soil. If there is no spatial association between samples, i.e., the association is entirely random, it is referred to as "pure nugget effect."

For many geostatistical purposes, $\gamma(h)$ is expressed as a mathematical function of h. There are several widely used forms for this. The most common is the spherical model, which has wide applications in field situations (1). It is sufficiently widely used that it is to geostatistics as the normal distribution is to parametric statistics (1).

Kriging: Kriging is a method of optimal unbiased estimation of variables at unsampled locations based on parameters of the semi-variogram and initial data values (1). Predictions are weighted averages of the actual measurements, with values for the weights derived from solution of a set of equations determined by the semi-variogram and the location and orientation of the sample points relative to each other and to the point or area being predicted (1). Weights are chosen to give unbiased estimates and to minimize estimation variance. Kriging supplements predictions with standard errors, permitting quantification of the uncertainty in the predicted value. Because the kriging technique is robust, minor variations in semi-variogram parameter estimates generally make little difference in the interpolated values or their standard errors.

Once the semi-variogram is constructed and its parameters determined, it can be used in kriging. The semi-variogram's form defines the relationship between the regionalized variable at different lags and also leads to the optimal weights of the kriging estimators. There are two types of kriging: point kriging and block kriging (2). Point kriging is used to predict the value of a single measurement at a (generally unsampled) location. Block kriging is used to predict the average of the regionalized variable in some larger "support." The "support" of a sample is the amount of physical material it encompasses, for example, the topsoil in a 1-m^2 portion of the field centered at some location. For prediction at arbitrary locations, both types of kriging problem require an explicit mathematical function for the semi-variogram.

Both point and block kriging can be used to produce contour maps, although the maps differ in purpose and appearance (2). Particularly where there is a substantial nugget effect, the contour maps produced by block kriging will be much smoother than those produced by point kriging.

Change of support: This third area of geostatistics is less commonly used than the other two but nevertheless important. In measuring soil nematode densities, standard practice involves extracting and counting nematodes from a subsample of some fixed volume of soil. Of interest is the relationship between statistics obtained from the subsample and what might be observed using larger or smaller subsamples. Also of interest is the error involved in extrapolating the average nematode number per unit volume of the subsample to the volume of the original sample. The application of kriging contains an indirect change of support calculation. This arises when predictions are made, not of the number of nematodes that might be found if a sample were taken at a particular point, but of the average number of nematodes per unit volume in the topsoil of a particular subarea of a field.

Materials and Methods

A grid of sampling points composed of five east–west transects of 40 points, each 10 m apart, was laid out in a 1.56-ha portion of a reed canary-grass forage field at the University of Minnesota North Central Experiment Station at Grand Rapids, Minnesota, in late June 1989. A single soil core was removed from the A_p horizon at each of the 200 grid intersections with a truck-mounted hydraulic corer and a 7.5-cm-d sampling tube. The A_p horizon (and sampling depth) varied from approximately 20–35 cm.

Plant-parasitic nematodes were extracted from 116-cm^3 soil subsamples with a modified Baermann pan technique (5). An 8-ml aliquant was removed for nematode identification and quantification from an average total extract volume of 103 ml; therefore, actual nematode counts were multiplied by 11.1 to represent nematodes per 100 cm^3 soil.

Measurement error for nematode density was estimated as follows, with data collected in this study for *Pratylenchus penetrans* as an example (Wallace, unpubl.). A total of 3,599 nematodes were counted, for an average from 200 samples of 17.995 nematodes per sample. Poisson counting statistics applied, so the variance was assumed to equal the mean. Multiplying by 11.1 yielded a measurement variance of $17.995 \times 11.1^2 = 2{,}217$, which (making allowance for variation in the actual volume of liquid extracted) was rounded to 2,500.

Twenty-two edaphic parameters were measured as previously described (7). These parameters included total sand and its components (very coarse, coarse, medium 1, medium 2, fine, very fine 1, and very fine 2), total silt and its components (coarse and fine), and clay; effective cation exchange capacity and levels of exchangeable cations (Ca, Mg, K, Na); extractable acidity; organic matter content; pH in water and $CaCl_2$; and field moisture content.

Geostatistical analyses of the data were conducted as follows. Each sample point was given an (x,y) grid coordinate based on distance and direction, with the point of origin located outside the grid of sampling points. Values for each variable were entered with matching grid coordinates, i.e., the computer file for each variable consisted of three columns of numbers: x-coordinate; y-coordinate; and value for the variable.

Semi-variograms for each variable were modelled by the computer program Geostatistical Environmental Assessment Software (GEO-EAS) (7). Computer default settings selected the minimum and maximum variable values from the data file. A pair comparison file was created from lag intervals set at a minimum of 0 m and a maximum of 200 m so as not to exceed a software-imposed limit of 16,384 pairs. The file was used to calculate the semi-variogram values for each variable. These values were displayed as a plot. The curve for the experimental semi-variogram may be fitted by eye (1), but fitting by weighted least squares (3) is both more efficient statistically and more reproducible. Either a spherical or linear model was fitted to each semi-variogram plot by visually estimating the nugget, sill, and range.

Contour maps were constructed and plotted with the computer program Surfer 4.04 (Golden Software, Golden, CO). Ordinary block kriging was used for interpolation. A general equation for this function is: $T^* = w_1g_1 + w_2g_2 + w_3g_3 + \ldots + w_ng_n$, where T^* is the estimator; $w_1, w_2, w_3, \ldots w_n$ are the kriging weights; and $g_1, g_2, g_3 \ldots g_n$ are the sample values. Contour intervals were determined by visually inspecting the data and setting the contour intervals to yield the maximum amount of information while avoiding a cluttered appearance.

Results and Discussion

Semi-variograms: Nematode and soil semi-variograms representative of spherical and linear models were constructed (Figs. 1–4). The general shape of the curves was similar between nematode and soil data sets fitted to the spherical semi-variogram model (Figs. 1–2), but the nugget values (see *y*-intercept) differed notably. The nugget effect in the semi-variogram for *Tylenchorhynchus* spp. was approximately 4,500, which indicated that counts for closely spaced samples would be substantially different. A large non-zero nugget was characteristic of all nematode semi-variograms (Table 1). In contrast, the semi-variogram for coarse sand had no nugget effect, which indicated that measurements of closely spaced samples would be almost identical. Nugget values for soil semi-variograms were zero or very close to zero for all soil tests performed at the site (Table 2), indicating that there was very little small-scale random variability or unpredictability associated with the soil sample values.

The semi-variogram for *Tylenchus maius* (Fig. 3) exhibited pure nugget effect, i.e., there was no discernible relationship between sampling points nearby or distant. This implies that the spatial distribution for *Tylenchus maius* was a purely random phenomenon at the sampling distance used in our study.

In several instances, the linear semi-variogram model applied to our soil data

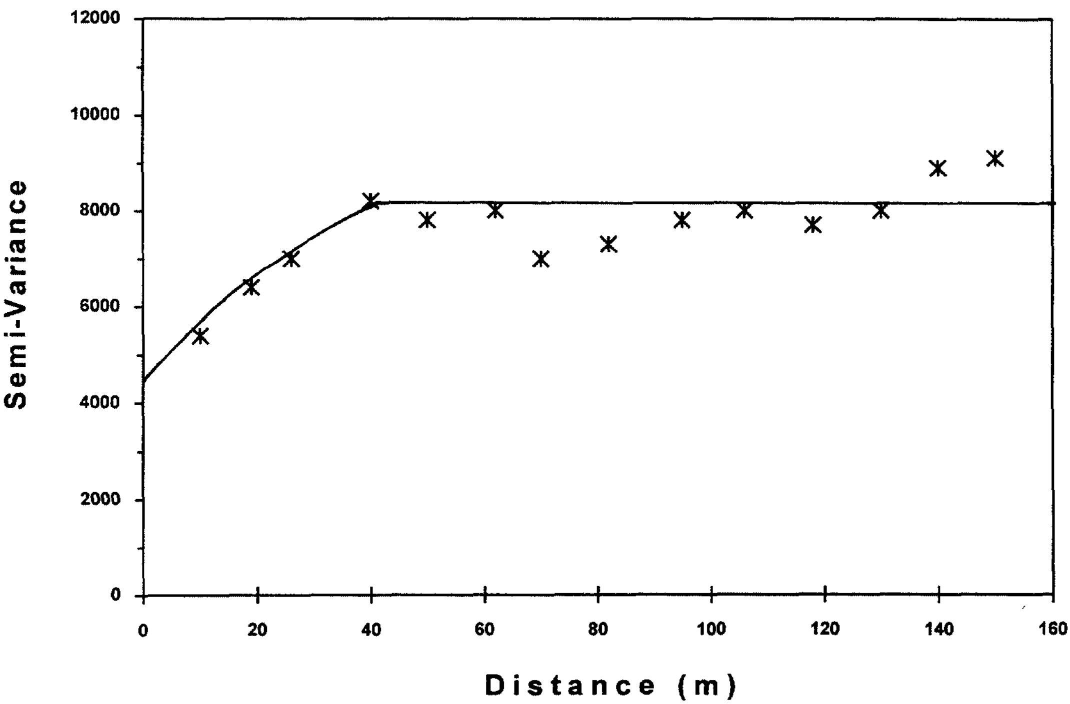

FIG. 1. Semi-variogram (spherical model) for a plant-parasitic nematode, *Tylenchorhynchus* spp., in a reed canary-grass field near Grand Rapids, Minnesota.

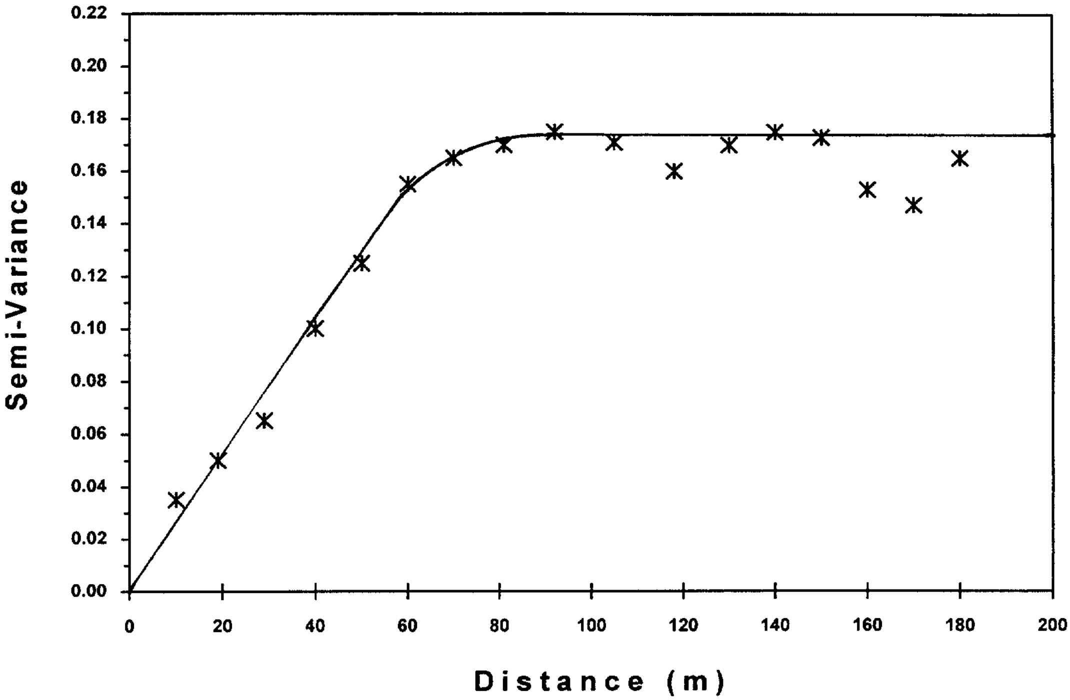

FIG. 2. Semi-variogram (spherical model) for coarse sand in a reed canary-grass field near Grand Rapids, Minnesota.

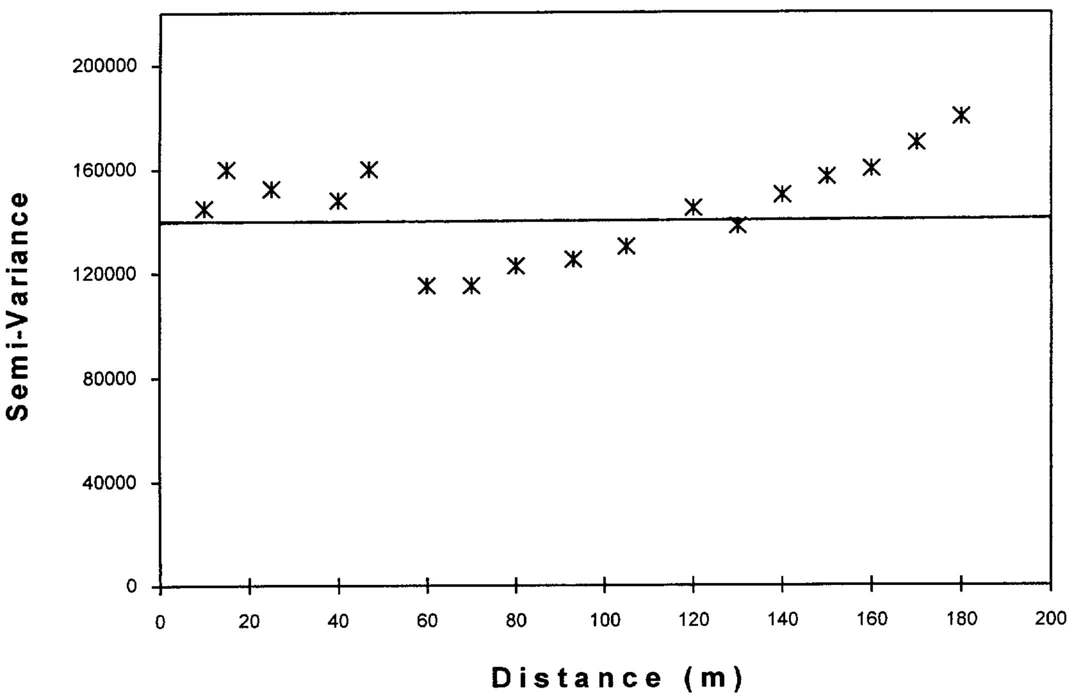

FIG. 3. Semi-variogram (pure nugget effect) for a plant-parasitic nematode, *Tylenchus maius*, in a reed canary-grass field near Grand Rapids, Minnesota.

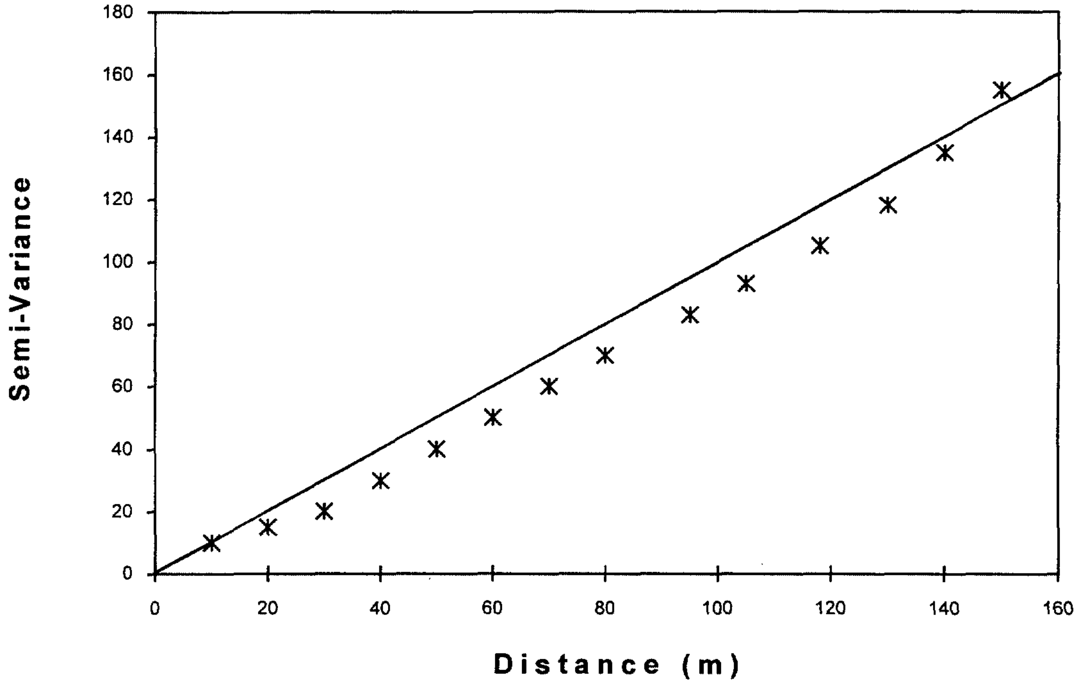

FIG. 4. Semi-variogram (linear model) for very fine sand in a reed canary-grass field near Grand Rapids, Minnesota.

set (e.g., Fig. 4, very fine sand). Again, as was characteristic of soil data, the nugget effect was zero for this type of model.

Kriging: Contour maps for *Tylenchorhynchus* spp., *Tylenchus maius*, coarse sand, and very fine sand were produced by ordinary block kriging (Fig. 5). Because block kriging was used, the contour lines for the nematode maps were smooth despite the large nugget values for the nematode data. This reflects the fact that going up from the point support of the original sample to the concept of the average true nematode density at a point eliminated the random variability of the nugget effect. Contour mapping was useful for visualizing nematode and soil data and for clarifying statistically significant correlations between the nematode and soil data sets (7).

Change of support calculations: In nematology, the nugget effect can be separated into two conceptual parts. One part is any inherent colonizing or clustering tendency between nematodes of the same species, i.e., true short-range spatial variability in the nematode population. This component of the nugget effect is a fundamental function of the biology of the organisms. The second part is measurement error, and it relates to the support issue—the size

TABLE 1. Semi-variogram parameters, including model, nugget, sill, and range, for nematode taxa in a reed canary-grass field near Grand Rapids, Minnesota.

Variable	Model	Nugget	Sill	Range
Pratylenchus penetrans	Spherical	25,000	40,000	60
Aglenchus agricola	Spherical	5,800	8,400	160
Tylenchorhynchus spp.	Spherical	4,500	8,100	45
Tylenchus maius	Pure nugget effect	140,000	—	—
Heterodera trifolii	Spherical	3,400	4,300	40
Paratylenchus spp.	Spherical	9,000	14,000	70
Criconemella sp.	Spherical	40	98	90

TABLE 2. Semi-variogram parameters, including model, nugget, sill, and range, for soil variables from a reed canary-grass field near Grand Rapids, Minnesota.

Variable	Model	Nugget	Sill	Range
Total sand	Spherical	0	500	200
Very coarse sand	Spherical	0.02	0.07	170
Coarse sand	Spherical	0	0.175	90
Medium 1 sand	Spherical	0	0.07	85
Medium 2 sand	Spherical	0	2.8	180
Fine sand	Spherical	0	600	220
Very fine 1 sand	Linear	0	—	—
Very fine 2 sand	Linear	0	—	—
Total silt	Spherical	0	200	200
Coarse silt	Spherical	0	23	190
Fine silt	Spherical	0	90	200
Clay	Spherical	0	70	200
Moisture	Spherical	0.4	3.75	160
Organic matter	Spherical	0.1	1	170
pH in water	Spherical	0.01	0.055	40
pH in $CaCl_2$	Spherical	0.01	0.06	40
Effective cation exchange capacity	Spherical	0	54	200
Calcium	Spherical	0	40	240
Magnesium	Spherical	0.1	1.8	200
Potassium	Spherical	0.01	0.065	65
Sodium	Spherical	0.0002	0.0002	40
Extractable acidity	Spherical	0.00002	0.000033	45

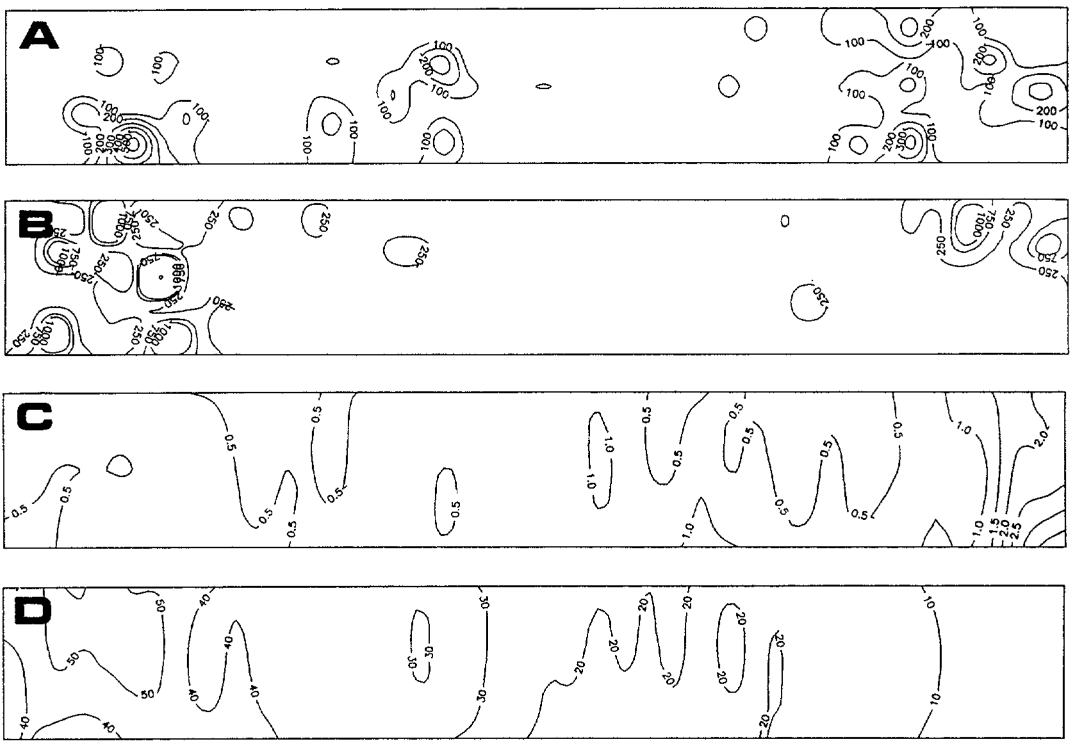

FIG. 5. Contour maps produced by ordinary block kriging of soil and nematode data acquired from a reed canary-grass field near Grand Rapids, Minnesota. Dimensions of the site are 390 × 40 m, with a total area of 1.56 ha. A) *Tylenchorhynchus* spp. Contour units are numbers of nematodes per 100 cm^3 soil. B) *Tylenchus maius*. Contour units are numbers of nematodes per 100 cm^3 soil. C) Coarse sand (particle size is 1–0.5 mm). Contour units are percentages. D) Very fine sand (particle size is 0.1–0.074 mm). Contour units are percentages.

TABLE 3. Partitioning of nugget effects into measurement error and true short-range variability in population density for plant-parasitic nematode taxa in a reed canary-grass field near Grand Rapids, Minnesota.

Nematode	Nugget effect	Measurement error†	True short-range variability‡
Pratylenchus penetrans	25,000	2,500 (10%)	22,500 (90%)
Aglenchus agricola	5,800	1,000 (15%)	4,800 (85%)
Tylenchorhynchus spp.	4,500	650 (15%)	3,850 (85%)
Heterodera trifolii	3,400	300 (10%)	3,100 (90%)
Paratylenchus spp.	9,000	600 (5%)	8,400 (95%)
Criconemella sp.	40	25 (60%)	15 (40%)

† Measurement error includes error associated with nematode sampling, extraction, and counting methodology.
‡ True short-range variability in the nematode population is due to inherent colonizing or clustering tendencies between nematodes of the same species.

of the subsample. The magnitude of this component is inversely proportional to the subsample volume, and it is subject to experimental control. Larger subsamples will lead to proportionately smaller values for this component of the nugget effect.

The total nugget effect for *P. penetrans* was estimated from the semi-variogram at 25,000. Of this, 2,500 (10%) was estimated to be the result of measurement random variation, with the remaining 22,500 (90%) due to true short-range variability in the nematode population density. Estimating nematode measurement error for each remaining nematode taxon led to the following approximate breakdown of the nugget effect into measurement error and true short-range variability (Table 3). For all nematode taxa except *Criconemella* sp., the true short-range variability inherent in nematode populations predominated. For *Criconemella* sp., measurement error predominated, possibly because the modified Baerman pan technique is not an optimal extraction technique for this nematode.

Partitioning the nugget effect into measurement variance and true short-range random variability permits an assessment of how changes in methodology could affect the precision of data gathered. The measurement variance is approximately inversely proportional to the volume of the soil subsample. The use of a larger soil subsample could improve the precision of population density estimates, provided that extraction efficiency did not decline drastically. Another strategy to improve precision of nematode density estimation would be to increase the number of samples (decrease distance between samples) while keeping the grid area constant. A short-range variographic experiment, in which a traverse of touching primary samples was taken and used to derive a variogram, would help to confirm or modify the nugget effects we have calculated from 10-m spacings.

Geostatistics as a means to predict nematode population densities warrants additional investigation. Nematode density at any sampled point has some indicative value for the nematode density at nearby locations (as measured in tens of meters) but there is substantial local-scale random variability superimposed on this spatial persistence. The relative sizes of the nugget effects and the sills for most of the nematode data, coupled with the magnitude of the ranges, support this assertion. Because sources of error in nematode population estimation exist as a result of sampling, extraction, and counting methodology, geostatistical analysis could be used to determine how to adjust such procedures to reduce the measurement random variability associated with nematode densities.

LITERATURE CITED

1. Clark, I. 1979. Practical geostatistics. London: Applied Science Publishers.

2. Cressie, N. 1991. Statistics for spatial data. New York: John Wiley.

3. Cressie, N. 1985. Fitting variogram models by weighted least squares. Mathematical Geology 17: 563–586.

4. Ferris, H., and L. T. Wilson. 1987. Concepts and principles of population dynamics. Pp. 372–376 *in* J. A. Veech and D. W. Dickson, eds. Vistas on nematology: A commemoration of the twenty-fifth anniversary of the Society of Nematologists. Hyattsville, MD: Society of Nematologists.

5. Kable, P. F., and W. F. Mai. 1968. Influence of soil moisture on *Pratylenchus penetrans*. Nematologica 14:101–122.

6. United States Environmental Protection Agency. 1988. Geostatistical Environmental Assessment Software. Las Vegas, NV: Environmental Monitoring Systems Laboratory.

7. Wallace, M. K., R. H. Rust, D. M. Hawkins, and D. H. MacDonald. 1993. Correlation of edaphic factors with plant-parasitic nematode population densities in a forage field. Journal of Nematology 25:641–652.