

Published in final edited form as:

Geoderma. 2010 May ; 156(3-4): 243–252. doi:10.1016/j.geoderma.2010.02.024.

Geostatistical modeling of the spatial variability and risk areas of southern root-knot nematodes in relation to soil properties

B.V. Ortiz^{a,*}, C. Perry^b, P. Goovaerts^c, G. Vellidis^d, and D. Sullivan^e

B.V. Ortiz: bortiz@auburn.edu; C. Perry: perrycd@uga.edu; P. Goovaerts: goovaerts@biomedware.com; G. Vellidis: yiorgos@uga.edu; D. Sullivan: dana@turfscout.com

^a Auburn University. 204 Extension Hall, Auburn University, Auburn, AL, 36849, USA

^b Biological and Agricultural Engineering Department, P.O. Box 748, University of Georgia, Tifton, GA, 31743, USA

^c Biomedware, Inc., Ann Arbor, MI, USA

^d Biological and Agricultural Engineering Department, P.O. Box 748, University of Georgia, Tifton, GA, 31743, USA

^e Turfscout., 309 King St Apt 3K - Greensboro, NC 27406, USA

Abstract

Identifying the spatial variability and risk areas for southern root-knot nematode [*Meloidogyne incognita* (Kofoid & White) Chitwood] (RKN) is key for site-specific management (SSM) of cotton (*Gossypium hirsutum* L.) fields. The objectives of this study were to: (i) determine the soil properties that influence RKN occurrence at different scales; and (ii) delineate risk areas of RKN by indicator kriging. The study site was a cotton field located in the southeastern coastal plain region of the USA. Nested semivariograms indicated that RKN samples, collected from a 50×50 m grid, exhibited a local and regional scale of variation describing small and large clusters of RKN population density. Factorial kriging decomposed RKN and soil properties variability into different spatial components. Scale dependent correlations between RKN data showed that the areas with high RKN population remained stable though the growing season. RKN data were strongly correlated with slope (SL) at local scale and with apparent soil electrical conductivity deep (EC_{a-d}) at both local and regional scales, which illustrate the potential of these soil physical properties as surrogate data for RKN population. The correlation between RKN data and soil chemical properties was soil texture mediated. Indicator kriging (IK) maps developed using either RKN, the relation between RKN and soil electrical conductivity or a combination of both, depicted the probability for RKN population to exceed the threshold of 100 second stage juveniles/100 cm³ of soil. Incorporating EC_{a-d} as soft data improved predictions favoring the reduction of the number of RKN observations required to map areas at risk for high RKN population.

Keywords

Cotton; Factorial kriging; Indicator kriging; Logistic regression; Nematodes; Risk map; Semivariogram; Soil properties; Southern root-knot nematode; Spatial variability

*Corresponding author. Tel.: +1 334 844 5534; fax: +1 334 844 4586.

1. Introduction

Southern root-knot nematode [*Meloidogine incognita* (Kofoid & White) Chitwood] (RKN) usually aggregates in irregular patches of coarse sandy soil texture (Goodell and Ferris, 1980; Koenning et al., 1996). In these areas cotton plants typically show no obvious early season above-ground symptoms of nematode damage; hence, it is difficult to identify patches infested by RKN before damage to the crop occurs. Yield losses attributed to southern root-knot nematode (RKN) account for 72% of the total losses caused by different species of nematodes found in U.S. cotton fields.

The assessment of RKN populations is commonly made through collection of soil samples. However high sampling costs can hamper an accurate estimation of its spatial distribution, leading to missed population patches and the reduced efficacy of any management strategy. Besides soil texture, there might be other biotic and abiotic factors associated with nematode reproduction, movement, and distribution within fields. Therefore, if the relationship between RKN and these controlling factors can be established, they can be used for assessing areas at risk for high RKN populations which can be targeted for site-specific management (SSM) and reduce the number of required RKN samples.

Various geostatistical techniques have been applied to nematode population data to determine sampling strategies, levels of infestation, and risk areas. In addition, they have been widely used to describe the spatial variability of nematodes and their spatial relationship with other variables. Through a nested sampling design, Webster and Boag (1992) showed that the spatial dependency of cereal cyst nematode (*Heterodera avenae*) and potato cyst nematode (*Globodera rostochiensis*) in the topsoil ranged from 5 to 50m. Comparison of indicator direct and cross-semivariograms also indicated that the population increased from patch edges towards their centers. Avendaño et al. (2003) found a poorly structured spatial variability for soybean cyst nematodes-SCN (*Heterodera Glycines*) in two Michigan fields (U.S.). For the same fields, Avendaño et al. (2004) reported a positive correlation between SCN population density and percentage of sand. Wyse-Pester et al. (2002) used semivariograms to explain the spatial dependence of three different nematode species within two corn fields. Nematode samples were correlated over distances of 115 to 649 m according to the direction (spatial anisotropy). When they tried to associate nematode population density with soil texture and organic matter content, correlations were inconsistent. Noe and Barker (1985) evaluated 26 different edaphic properties with respect to the spatial distribution of RKN and found that high levels of clay or organic matter, low copper concentrations, and small changes in percent soil moisture were strongly correlated with RKN spatial distributions. Monfort et al. (2007) explained 65–86% of cotton yield variability measured in plots from similar geographic locations using the initial concentrations of RKN and sand content. Other studies have correlated the abundance of RKN with soil pH (Melakeberhan et al., 2004) and soil moisture (Wheeler et al., 1991; Windham and Barker, 1993).

Although previous studies obtained promising results for the characterization of the nematode spatial variability and identification of nematode covariates, none used covariates to identify areas at risk for high RKN populations. Therefore, in this paper we adopted a geostatistical approach to verify the hypothesis that the presence of RKN may be related with soil properties which could then be used as surrogate data to identify high population risk areas. To test these hypotheses, factorial and indicator kriging were used to identify multiple scales of variation and separate them into the corresponding spatial components which can later be used as surrogate data to estimate the probability or risk of encountering population densities above a critical threshold (Goovaerts, 1998; Goovaerts et al., 2005a,b). Factorial kriging analysis (FKA) has been used extensively in soil science (Goovaerts, 1994;

Castrignanò et al., 2000). Typically through a filtering of short-range variation FKA enhances the relation between variables that might otherwise be confounded by mixing all different sources of variation, leading to a better understanding of the physical underlying mechanisms controlling spatial patterns. Castrignanò et al. (2007) used factorial kriging to compute one regionalized factor that summarizes the effect of soil pH, electrical conductivity, exchange sodium percentage, and total clay plus fine silt content on soil salinization. Goovaerts (1994) separated the local and regional variation of soil and vegetation properties using factorial kriging. He attributed local variation to field-to-field differences and regional variation to the presence of different soil types. Moreover, Goovaerts (1998) showed that the probability that an attribute value exceeds a target threshold at an unsampled location can be estimated by indicator kriging which uses a kriging estimator similar to the one developed for continuous variables. Previous studies have adopted indicator kriging (IK) to estimate and map the risk of exceeding threshold values in watershed management (Lyon et al., 2006), soil pollution (Goovaerts et al., 1997; Lin et al., 2002) and groundwater contamination (Goovaerts et al., 2005a,b).

The objectives of this study are: (i) to determine the soil properties that influence RKN occurrence at difference scales; and (ii) to delineate areas at risk for RKN based on indicator kriging (IK) of hard data (i.e., measured RKN population density), soft data (i.e., logistic regression between RKN and soil properties), and the combination of hard and soft data.

2. Materials and methods

2.1. Study fields description and data collection

The study was conducted in a 20 ha irrigated producer field (31° 23' 60 N, -83° 37' 48 E), located in the USA southeastern coastal plain which is characterized by sandy soils, small differences in topographic relief, and a subtropical climate. The field was planted in 2006 with the cotton (*Gossypium hirsutum* L.) cultivar – Delta & Pine Land Company DP 555 BG/RR.

A 50×50 m grid (0.25 ha cell size) was superimposed over the field and a georeferenced sample was collected at the center of each cell (99 samples total). Soil samples for nutrients, RKN, and texture analyses were collected from random locations within a 1.5 m radius of the central node of each grid. Five 30 cm soil cores were collected and combined into a composite sample for phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and soil pH determination. These samples were collected one month after planting. Soil samples for RKN population density determination (second stage juveniles) were collected three times during the growing season — 75, 110, and 167 days after planting (DAP) which coincided with early season (first flower), mid season, and harvest. These sampling events were designated as RKN1, RKN2, and RKN3, respectively. At each sampling event, eight individual subsamples were collected around the center of each cell and composited into a single sample representing RKN population density within each grid cell. The subsamples were collected with a 3 cm diameter sampler which was inserted 15–20 cm deep into the soil adjacent to plant tap roots where the RKN are likely to live (Shurtleff and Averre, 2000). Nematodes were extracted from 100 cm³ of soil by centrifugal flotation (Jenkins, 1964).

Exhaustively sampled data included apparent soil electrical conductivity (EC_a) and elevation (EL), and information derived from these data (i.e., slope derived from elevation). In this study, the VERIS® 3100 implement was used to measure EC_a between 0 and 30 cm (shallow, EC_{a-s}) and 0–90 cm (deep, EC_{a-d}) in 9 m parallel swaths throughout the fields and were georeferenced. Elevation data (EL) were collected at the same time as EC_a data with RTK GPS receiver mounted on the tractor pulling the VERIS® 3100 implement.

2.2. Data processing

Although the EL and EC_a data sets included more than 7000 observations, the parallel swaths used to collect the data were not necessarily collocated with the sampling nodes (grid cell centers). Ordinary punctual kriging was used to estimate the values of EL and EC_a at the sampling nodes (Kerry and Oliver, 2003) using TerraSeer Space-Time Information System (STIS) vr. 1.7.91 (www.terraser.com).

Raster maps of terrain slope (SL) were derived from EL raster maps using the Spatial Analyst extension of ArcVIEW v. 9.0 (ESRI, 2004a). The slope at the sampling node was estimated by averaging the pixel values of slope contained within the 1.5 m radius sampling area surrounding each sampling node. This average slope was then associated with the corresponding RKN data.

2.3. Statistical and multivariate geostatistical analyses

2.3.1. Modeling nested semivariograms—Some of the RKN population density and soil properties data had skewness values above +1 or below −1; hence, failing the assumption of normality. Because the variogram, core of geostatistical analyses, is very sensitive to extreme values, a normal score transformation of the data was performed when necessary. The normal score transform replaced each observation with the corresponding quantile in the standard normal distribution, enabling the normalization of any histogram regardless of its original shape (Goovaerts et al., 2005a,b). Because we are dealing with different types of data sets, all the variables were standardized to a zero mean and unit variance.

The interaction of several biotic or abiotic processes with RKN population density and soil properties might operate differentially at different spatial scales; hence, if the scales at which a pair of variables operate are very different from one another, this should be evidenced by their semivariograms. Each variable can be modeled by a nested semivariogram, which takes the form of a linear combination of different structures $g_l(h)$ with unique ranges of spatial dependence:

$$\gamma(h) = \sum_{l=0}^L b_l g_l(h) \text{ with } b_l \geq 0 \quad (1)$$

where b_l is the variance of the corresponding semivariogram model $g_l(h)$. For $L=2$, the variance b_0 is called nugget effect, $g_1(h)$ is the model with a short range, a_1 (local variability) and $g_2(h)$ is long range model, a_2 (regional variability) (Goovaerts, 1997). The linear model of regionalization (Eq. (1)) assumes that a random function $Z(\mathbf{u})$ with a nested semivariogram $\gamma(h)$ can be interpreted–decomposed as the sum of $L+1$ independent orthogonal random functions or spatial components, Z_l , with semivariogram $b_l g_l(h)$ and a local mean or trend component $m(\mathbf{u})$:

$$Z(\mathbf{u}) = \sum_{l=0}^L Z_l(\mathbf{u}) + m(\mathbf{u}) \quad (2)$$

where $Z_0(\mathbf{u})$ is a micro-scale component, and $Z_1(\mathbf{u})$ and $Z_2(\mathbf{u})$ are the short-range and long-range spatial components associated with the semivariogram models $b_1 g_1(h)$ and $b_2 g_2(h)$, respectively. The best semivariogram model for each data set was chosen based on the maximum coefficient of determination and minimum residual sum of squares for the fit (Isaacs and Srivastava, 1989).

2.3.2. Factorial kriging analysis—Factorial kriging analysis (FKA) has been widely described in the geostatistical literature (Matheron, 1982; Goovaerts, 1997). Thus, only a short section is devoted to it. This geostatistical method allows the estimation and mapping of each source of spatial variability or spatial component $Z_l(\mathbf{u})$ identified in the nested semivariogram, making possible the estimation of scale-dependent correlations among variables. After filtering the noise in the data, spatial components can be grouped into sources of local variability (short range) and regional variability (long range plus local mean or trend). Combining the trend estimate and long-range component was initially proposed by Jaquet (1989) to reduce the impact of the size of the search window on the estimation of the long-range component. Since then, it has been used in several applications of factorial kriging (Oliver et al., 2000; Goovaerts et al., 2005a,b; Oliver, 2010).

Scale-dependent correlations, ρ^l , were calculated directly from kriged spatial components at a specific spatial scale l (i.e., short or regional), without the effect of other scales of variation (e.g., Goovaerts, 1998). In the present application, the spatial distribution of RKN population is likely influenced by soil properties which are expected to be correlated with RKN data at the scale at which they operate. For example, the Pearson's correlation coefficient between RKN2 and slope (SL) is only -0.37 while the same correlation is -0.59 after noise and local components were filtered out.

Statistical and geostatistical analyses were performed using SAS (SAS Institute, 2007), TerraSeer STIS v. 1.7.91 (www.terraseer.com), ISATIS (Geovariances, 2007), and the Geostatistical Analyst extension of ArcVIEW v. 9.0 (ESRI, 2004b).

2.4. Indicator approach to delineate high risk areas

The identification of areas at risk for high populations of RKN is required for management purposes. A threshold of 100 second stage juveniles/100 cm³ of soil is typically used by cotton producers in the Southeast U.S. to trigger nematicide applications. Three geostatistical approaches were evaluated for mapping the probability that this threshold is exceeded: *a*) probabilities calculated from indicator kriging (IK) of RKN population data – *Hard data*, *b*) probabilities derived from ordered logistic regression between RKN population and ancillary data (McCullagh, 1980) – *Soft data*, and *c*) logistic regression probabilities updated using kriged residuals calculated from the *a* and *b* techniques (Goovaerts et al., 1997) — *Hard and Soft data*. Indicator kriging uses the RKN population data only; hence, the probability of exceedence depends on the spatial structure of the sparsely sampled RKN population. However, the estimation can be improved by including densely sampled secondary data related to RKN population. Therefore, probability estimates will not only depend on the RKN population alone but also on the spatial variability of the secondary variables and their relationships with the primary variable. An indicator kriging approach has been selected because: 1) nematode population data typically exhibit a skewed distribution which makes hazardous any transformation to achieve a Gaussian distribution, and 2) the grower is mainly interested in a map of the probability that the RKN population density will exceed an advisory threshold.

2.4.1. Indicator kriging — hard data—The modeling of the spatial distribution of z values above or below a given threshold value z_k , 100 RKN second stage juveniles/100 cm³ of soil, required a prior coding of each observation $z(\mathbf{u})$ (RKN population density) into a new binary or hard indicator variable: 0 (below threshold) and 1 (above threshold). The indicator semivariogram was calculated using six lags of 50m and modeled with ISATIS (Geovariances, 2007).

2.4.2. Logistic regression — soft data—A *soft indicator* equivalent to a probability (i.e. valued between 0 and 1) can be derived from the relationship between the hard indicator variable and a secondary variable which is more densely sampled. In this study the secondary variable(s) was selected from a set of soil properties after the structural correlation analysis. The response variable, expressed in terms of probability, was modeled as a linear function of one or multiple soil properties (X_1, X_2, \dots, X_u) (Kleinschmidt et al., 2000; Lyon et al., 2006):

$$\log(p_u/1 - p_u) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_u X_u \quad (3)$$

where p_u is the probability of having a RKN population density above the threshold; β_u are the parameters estimated by logistic regression; and X_u represents the soil properties or explanatory variables. Therefore, the *soft* or prior probability of observing a population of RKN above the threshold based on RKN surrogate data are estimated as:

$$p^*(u) = e^{(\alpha + \beta^* X)} / (\log + e^{(\alpha + \beta^* X)}). \quad (4)$$

The significance of the logistic regression model was evaluated using a likelihood ratio (-2LogL) with an approximated chi-square distribution. In this study the explanatory variable(s) in Eq. (4) was the local mean of the soil property with the highest and most stable spatial correlation with RKN as determined from the scale-dependent correlation analysis. These local means were estimated at the RKN sampling locations by factorial kriging. The prediction formula (Eq. (4)) was applied to the ordinary kriging map of secondary data using the TerraSeer STIS software to produce a prior probability map based on soft data.

2.4.3. Indicator kriging — hard and soft data—Soft indicators derived from the logistic regression can be considered as an intermediate step in the delineation of probability maps when surrogate data are used (Kleinschmidt et al., 2000; Grunwald et al., 2006; Lyon et al., 2006). At any location u , prior probabilities $p^*(u)$ can be updated into posterior probability using the residuals calculated between the probabilities calculated from hard and soft data. The indicator maps from hard and soft data were combined following the method suggested by Goovaerts et al. (1997): 1) residuals are computed by subtracting, at each RKN sampling location, prior probabilities (soft data) from the indicator variable (hard data), 2) the semivariogram of the residuals was calculated and the model was used to interpolate residuals through simple kriging, and 3) the final probability map was obtained by adding the prior probability map to the kriged residuals.

To simulate the impact of a smaller number of RKN data on the accuracy of indicator maps and to assess the benefit of using soft data, the procedure described above was repeated using a training data set composed of 65% and 37% of the initial RKN sampling observations which were selected randomly. The accuracy of the probability maps was quantified using jack-knife whereby the original data set is split into two samples, one for mapping and the other for validation. For each geostatistical approach (i.e., hard data, hard and soft data) and sampling density, map accuracy was determined by the number of false positives for a series of probability classes: the more false positives in the highest risk class, the lower the map accuracy.

3. Results and discussion

3.1. Descriptive statistics

The RKN population data, with skewness >1 , exhibited both spatial and seasonal variability. Using the coefficient of variation (CV) as an index of dispersion, it was possible to establish the within-field relative variability of RKN population and some soil properties. While the average population during RKN1 was low, it shows the highest dispersion around the mean relatively to RKN2 and RKN3. The highest mean RKN population density was observed during RKN2 which contradicts the generally accepted rule that the highest population density occurs near harvest (Table 1). Even though the cotton was irrigated, this finding might relate to drought conditions experienced towards the end of the growing season (September and October in 2006) and with limited availability of infection sites as a result of root decay (Stanton, 1986). The soil chemical properties exhibited less skewness, except for K and Mg, and the dispersion around the mean and the CV were lower than for the RKN population density and the soil physical properties (Table 1). For the soil physical properties, small changes in EL described by the low SD and CV values, evidenced its low spatial variability (Table 1). Therefore, this variable was not included for subsequent analyses in this study. In contrast, the CV of slope and EC_{a-d} (also having skewness values >1) were fairly high compared to soil chemical properties which might indicate that these soil physical properties had higher variation than most soil chemical properties.

3.2. Modeling nested semivariograms

The existence of several sources of variation operating at different scales resulted in nested semivariograms for most of the variables (Fig. 1). Common features in the experimental semivariograms (square symbol) for the RKN population at RKN1, RKN2, and RKN3 suggested their spatial correlation. Therefore, the spatial variability was modeled as a combination of three different spatial structures: *i*) nugget effect — associated with measurement errors due to the sampling or to micro-variation not captured by the 50 m sampling interval, *ii*) short scale of variation or short-range structure around 70 m (for RKN2, and RKN3), describing small clusters; and *iii*) large scale of variation or long-range structure, with a range of 216 m and 481 m for RKN2 and RKN3 respectively, indicating large clusters of RKN population (Fig. 1a–c).

The semivariograms of EC_{a-d} , SL, soil pH, and Ca shared similar features with the semivariograms of RKN2, and RKN3, which suggested their spatial correlation at short and long scales in most cases (Fig. 1b, c, d, e, f, and i). Among soil properties, the semivariograms of EC_{a-d} , soil pH, K and Ca had similar short scale ranges, fluctuating between 102 and 131 m which might indicate that the spatial variability of EC_{a-d} could be used to explain the variability of other variables; being then a good option as a covariate for RKN population (Fig. 1).

3.3. Scale-dependent correlation analyses

Table 2 lists the aspatial (Pearson's) and scale-dependent correlation coefficients between RKN population densities (RKN1, RKN2 and RKN3) and soil physical and chemical properties. Filtering the noise by factorial kriging generally increased the correlation among variables. For example, while the Pearson's correlation coefficient between RKN2 and EC_{a-d} was only -0.50 , a much larger correlation of -0.77 was observed when the noise and local components for both variables were filtered out. This strong regional correlation was blurred by the weak correlation at the local scale (Table 2).

Despite potential changes in the size of RKN clusters between the three sampling events, short scale correlations of 0.20 and 0.18, and long scale correlations of 0.46 and 0.51

between RKN1–RKN2 and RKN2–RKN3 (data not shown), respectively, indicated that the locations of high RKN population density, especially the large clusters, remained stable until the end of the growing season.

The correlation between RKN population density and soil physical properties changed as a function of spatial scale (Table 2). At local scale SL exhibited the most consistent negative seasonal correlation with the RKN population, which reflects the similarity in the short-range structures of both nested models (Fig. 1a–c, e). At a regional scale, this negative correlation exceeded -0.42 suggesting that high population density of RKN could be found in areas of a field with little or no slope. A strong negative correlation between each three population densities and EC_{a-d} was also observed at regional scale (Table 2). Previous studies have related sandy or coarse-textured soil with low EC_{a-d} values (Khalilian et al., 2001; Perry et al., 2006) and sandy areas with high RKN population density (Monfort et al., 2007). A correlation analysis between sand fraction, measured at this field and four other nearby fields showed a strong negative correlation between EC_{a-d} and soil particle size. Therefore, the negative spatial correlation between RKN population density and EC_{a-d} found here indicates that high RKN population are more likely in large areas with low values of EC_{a-d} which usually characterize sandy areas.

Soil chemical properties were more strongly correlated with RKN population density at the regional scale than at the local scale (Table 2). The strong negative correlation of K, Ca and Mg with RKN population density, especially at the regional scale, indicated that large RKN clusters of high population were spatially correlated with large areas of low K, Ca and Mg levels. Although the Ca levels (mean of 897 kg ha^{-1}), measured one month after planting, did not indicate any nutritional deficiency ($>247 \text{ kg ha}^{-1}$ of Ca is adequate for coastal plain soils), the spatial correlation of soil pH, K, Ca and Mg with EC_{a-d} indicated a strong positive correlation at the local scale (0.44 in average) and at regional scale (0.77 in average). Therefore, the spatial variability of EC_{a-d} could explain the variability of soil pH, K, Ca and Mg.

The relationship between soil chemical properties and RKN population can be considered direct or indirect or both (plant mediated). Oka et al. (2006, 2007) stated that the relation between RKN population and soil pH could related to the loss of nematicide activity by ammonia-releasing organic and inorganic fertilizers due to low levels of soil pH. Other studies have shown that mineral salts such as NaCl, NaNO_3 , KCl, KNO_3 , CaCl_2 , $\text{Ca}(\text{NO}_3)_2$, MgCl_2 , MgSO_4 , FeCl_2 , and FeSO_4 exhibit a degree of repellency toward *M. javanica* and *M. incognita* (Castro et al., 1990; Cadet et al., 2004). Again, this indicates that special attention must be paid to sandy areas where salts might leach preferentially.

The scale-dependent correlation analysis demonstrated the strong spatial relationship of EC_{a-d} with RKN population and most of the soil physical and chemical properties. This result suggests that EC_{a-d} can be a good covariate for mapping many features on this field specially RKN population. The advantage of EC_{a-d} over other variables is its low cost, allowing the collection of large volume of data to facilitate the characterization of RKN spatial variability.

3.4. Delineating high risk areas by indicator kriging

RKN2 data were selected for the IK analysis because of their strong spatial correlation with the soil physical and chemical properties (Table 2). While three different geostatistical approaches were evaluated for mapping the probability that RKN2 population density exceeds the threshold of 100 second stage juveniles/ 100 cm^3 , indicator kriging combining hard and soft data showed the most promising results.

3.4.1. Indicator kriging — hard data—The presence of two scales of variation on the indicator semivariogram calculated from 99 RKN observations suggested the existence of two cluster sizes with high risk (probability >75%) for RKN population to exceed the threshold (Fig. 2a). The short-range structure (117 m) and the long-range structure (257 m) describe the high-risk clusters located in the center and the northwest part of the field, respectively (Fig. 3a). Thirty-four (87%) of RKN samples above the threshold (white square symbols on the maps) coincided with high-risk areas on the map (probability >75%) and the other 13% were located in moderately high risk areas (50–75% probability range of risk) (Fig. 3a, Table 3).

The delineation accuracy of high-risk areas (probability >75%) dropped when 64 RKN observations were used to create the IK map (Fig. 3b). Now, 76% of the observations above the threshold were in the predicted high-risk area and 21% were predicted to be in the moderately high risk areas (Fig. 3b).

Indicator kriging of the smaller data set (35 RKN observations) predicted a few scattered spots of high risk for RKN (Fig. 3c). Out of the randomly selected RKN observations above the threshold, only 15% fell in the predicted high-risk area, while the remaining were located in the moderately high risk areas (Table 3). This decrease in accuracy might be associated with a reduction in the short range of spatial correlation (from 139 to 88 m) of the indicator semivariogram (Fig. 2c). This should serve as a warning for what could happen if producers collect too few soil samples to quantify nematode infestations in their fields.

3.4.2. Logistic regression — soft data—Soil EC_{a-d} was used as the *soft* data because of its strong spatial correlation with the RKN2 population and its low measurement cost which also provide large amount of data. The use of only *soft* data decreased the accuracy in the delineation of high risk areas (probability >75%) relative to the use of *hard* data alone. This could be associated with the low significance of the logistic regression models even though the contribution of EC_{a-d} was significant ($\alpha = 0.05$) (data not shown). On the soft map created from 99 RKN observations (Fig. 3d), only 18% of the RKN2 observations above the threshold agreed with the predicted high risk areas (Table 3).

When the number of observations decreased from 99 to 64, only 31% of the RKN observations above the threshold were classified as being in areas with high risk (Fig. 3e). In contrast, the soft map created from 35 observations did not exhibit a high risk zone (Fig. 3f, Table 3). In all three maps, however, the area delineated as having moderately high risk (50–75%) remained relatively stable and included between 31% and 43% of the observations exceeding the threshold.

Jack-knife using 35 RKN observations indicated that the percentage agreement in predicted high risk areas dropped by 66%, when soft data are used instead of hard data (Table 4).

The probability maps generated from soft data were not as accurate as the IK maps created from the hard data in terms of delineating the areas with the highest level of risk. However, a cotton producer might be able to use maps created this way to target the areas with more than 50% probability of having RKN population density above the threshold.

3.4.3. Indicator kriging — hard and soft data—Combining *hard* (RKN population) and *soft* (EC_a) data improved the accuracy of the IK maps created from only *soft* data, especially in the high risk areas. Fifty nine (59%) of RKN samples exceeding the threshold coincided with high-risk areas on the map (probability >75%) and the other 33% were located in moderately high risk areas (50–75% probability range of risk) (Fig. 3g–i, Table 3). The semivariograms of the residuals calculated for data sets of 99, 64, and 35 RKN

observations indicated the presence of spatial correlation with a range of spatial dependence of around 120 m (Fig. 2d–f). Although one can capitalize on the residual's spatial correlation to improve the map, the spatial correlation of the 35 residual values here might be a random event associated with the spatial location and values of those observations.

The advantage of combining hard and soft data was truly tested by reducing the initial number of RKN observations (99) available for mapping. When the IK map was delineated from 64 RKN observations, all the RKN observations above the threshold coincided with the areas predicted to have a high risk of exceeding that threshold. In contrast, only 76% and 31% of these observations were classified within that range in the maps created with the hard data alone and soft data alone, respectively (Table 3). The jack-knife validation method over 35 RKN observations indicated that the percentage agreement on predicted high risk areas improved 60% and 160% when the IK map-hard and soft data was evaluated respect to the IK map delineated from hard data or soft data alone, respectively.

Reducing the data set to 35 observations emphasized the benefit of combining hard and soft data for delineating zones at risk for RKN population density above the threshold value. The IK map created from the 35 hard data alone (Fig. 3c) greatly under predicted the extent of the high risk area compared to the map created from the combined data set (Fig. 3i). All the RKN observations above the threshold in the reduced data set were predicted as high risk zone while only 15% of the observations coincided in the hard data map and none coincided in the soft data map (Fig. 3c, f, and i, Table 4).

The percentage agreement in predicted high risk areas increased 850% and 357%, when the jack-knife validation method over 64 RKN observations was used to evaluate IK maps-hard and soft data respect to the IK maps delineated from hard data or soft data alone, respectively (Table 4). Accuracy was also improved in the lower risk areas.

The results presented here demonstrate the advantage of using EC_{a-d} as a covariate for improving mapping accuracy of areas at risk for high RKN populations while reducing the number of RKN samples, in this case by 35% and 64%.

If we assume that the distribution of RKN in this field is best represented by the dense data set used to create the hard data map in Fig. 3a, then the most striking difference is that the combined map (Fig. 3i) under estimates by about 40% the area identified as being at high risk by the hard data map (Fig. 3a). Nevertheless, the combined data map does well in predicting areas with at least a moderately high (probability >50%) risk of RKN population density exceeding the established threshold. In light of the earlier discussion on the cost and difficulty of collecting samples for RKN analyses, the slight loss in accuracy using 64 or 35 RKN observations is justified by the reduced cost of creating a combined hard and soft data map with a reduced number of observations.

Delineating different levels of risk for high RKN population on a map might help cotton producers to identify the zones and establish different management strategies for subsequent growing seasons. Target sampling based on the different levels of risk could provide a better spatial assessment of the RKN population present in the field which might results on site-specific application of different inputs including nematicide rates.

4. Summary and conclusions

The spatial aggregation pattern and temporal stability of RKN population density observed in the field under study throughout the 2006 growing season met some of the requirements for site specific management (SSM). The short range of spatial dependence can be used as a guideline for sampling RKN population density in fields with low topographic relief. Scale-

dependent correlations, derived from the spatial components estimated by factorial kriging, allowed the identification of covariates for mapping RKN risk areas and delineation of RKN management zones for SSM. The moderate to strong spatial dependence of RKN population density observed at mid season (RKN2) and the spatial stability of areas with high populations throughout the growing season favored their high correlation with soil physical properties. Although the relationship between RKN population density and soil chemical properties was at best weak, it pointed out that site specific management of soil nutrients might reduce the risk for having high population density of RKN. The relationship with soil physical properties was stronger. In particular, EC_{a-d} is a good covariate for RKN population density because the correlation is strong at both short and long scales and is stable with respect to space, time and location. The spatial aggregation of RKN data facilitated the segregation of RKN risk areas based on low values of EC_{a-d} through the development of indicator kriging maps combining RKN observations (hard data) and EC_{a-d} data (soft data).

Validation demonstrated the benefit of incorporating EC_{a-d} as soft data in the predictions. Indeed EC_{a-d} data are much densely sampled since they are less expensive and easier for a producer to collect than the RKN samples. In the absence of soil movement or the addition of large volumes of soil amendments, EC_a data need only be collected once.

Logistic regression showed that EC_{a-d} alone might not capture the total spatial distribution of RKN population density and predict the areas at risk for high populations. The advantage of combining hard and soft data was emphasized when using a reduced data set (64 or 35 RKN observations) with EC_{a-d} data resulted in a similar precision accuracy than a much large data set of RKN observations alone. The integration of other surrogate data for soil texture, such as slope and elevation along with EC_{a-d} , might improve the mapping accuracy even more.

The biggest advantage of combining soft data with hard data to develop probability maps is the fewer RKN observations required to assess the areas at risk for high population of RKN. Additionally, these maps might be used to target zones for additional sampling or application of nematicides, or both. However, it should be noted that the identification of surrogate data and estimation of areas at risk might be difficult if the RKN population density follows a random pattern of spatial variation or if RKN are not present.

The fact that RKN population density increase in areas of coarse textured soils where leaching of fertilizers is most likely to occur stresses the importance of mapping RKN risk not only for RKN population management but also for soil fertility management. The strong spatial correlation between the RKN and EC_{a-d} , a relatively stable variable in time, indicates that EC_{a-d} can be used to delineate management zones for RKN which will capture most of the variation of RKN. Future research must be focused on the effect that soil chemical properties have on the reproduction and survival of nematodes.

Acknowledgments

This work was supported by grant funds from Cotton Inc., the Georgia Cotton Commission, the Flint River Water Planning and Policy Center, Hatch and State funds allocated to the Georgia Agricultural Experiment Stations and by USDA-ARS CRIS project funds. Many thanks to the cotton producers who participated in this study and shared their time and records with us. Many thanks also to Dr. Robert Nichols who supported this project. Finally, special thanks to Dewayne Dales, Rodney Hill, Gary Murphy, Andrea Milton, and Katia Rizzardi who assisted with the extensive field work required to complete this project. The research by Dr. Goovaerts was funded by grant R43-CA135814-01 from the National Cancer Institute. The views stated in this publication are those of the author and do not necessarily represent the official views of the NCI.

Mention of commercially available products is for information only and does not imply endorsement.

References

- Avendaño F, Schabenberger O, Pierce FJ, Melakeberhan H. Geostatistical analysis of field spatial distribution patterns of soybean cyst nematode. *Agron J* 2003;95:936–948.
- Avendaño F, Pierce FJ, Schabenberger O, Melakeberhan H. The spatial distribution of soybean cyst nematode in relation to soil texture and soil map unit. *Agron J* 2004;96:181–194.
- Cadet P, Berry S, Spaull V. Mapping of interactions between soil factors and nematodes. *Eur J Soil Biol* 2004;40 (2):77–86.
- Castrignanò A, Giugliarini L, Risaliti R, Martinelli N. Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics. *Geoderma* 2000;97:39–60.
- Castrignanò, A.; Buttafuoco, G.; Puddu, R.; Fiorentino, C. A multivariate geostatistical approach to delineate areas at soil salinization risk. In: Stafford, J.; Werner, A., editors. *Proc Sixth European Conference of Precision Agriculture (6ECPA)*. Skiathos; Greece: 2007. p. 199–205.
- Castro CE, Belser NO, McKinney HE, Thomasson IJ. Strong repellency of the root knot nematode, *Meloidogyne incognita* by specific inorganic ions. *J Chem Ecol* 1990;16 (4):1199–1205.
- ESRI. ArcGIS Geostatistical Analyst. ESRI; Redlands, CA: 2004b.
- ESRI. ArcGIS Spatial Analyst 9.0. ESRI; Redlands, CA: 2004a.
- Geovariances. ISATIS technical references, version 7.0.6. 2007
- Goodell PB, Ferris H. Plant-parasitic nematode distributions in an alfalfa field. *J Nematol* 1980;12:136–141. [PubMed: 19300685]
- Goovaerts P. Study of spatial relations between two set of variables using multivariate geostatistics. *Geoderma* 1994;62:93–107.
- Goovaerts, P. *Geostatistics for Natural Resources*. Oxford University Press; New York, USA: 1997.
- Goovaerts P. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biol Fertil Soils* 1998;27:315–334.
- Goovaerts P, Webster R, Dubois JP. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environ Ecol Stat* 1997;4:31–48.
- Goovaerts P, AvRusking G, Meliker J, Slotnick M, Jacquez G, Nriagu J. Geostatistical modeling of the spatial variability of arsenic in ground water of southeast Michigan. *Water Resour Res* 2005a; 41:1–19.
- Goovaerts P, Jacquez GGM, Greiling D. Exploring scale-dependent correlations between cancer mortality rates using factorial kriging and population-weighted semivariograms. *Geogr Anal* 2005b;37 (2):152–182. [PubMed: 16915345]
- Grunwald S, Goovaerts P, Bliss CM, Comerford NB, Lamsal S. Incorporation of auxiliary information in the geostatistical simulation of soil nitrate nitrogen. *Vadose Zone J* 2006;5:391–404.
- Isaacs, EH.; Srivastava, M. *An Introduction to Applied Geostatistics*. Oxford University Press; New York, USA: 1989. p. 146
- Stanton, MA. MS thesis. Gainesville, FL: University of Florida, Agronomy Department; 1986. Effects of root-knot nematode (*Meloidogyne* spp.) on growth and yield of ‘Cobb’ soybean (*Glycine max* (L.) merrill).
- Jaquet O. Factorial kriging analysis applied to geological data from petroleum exploration. *Math Geol* 1989;21 (7):683–691.
- Jenkins WR. A rapid centrifugal flotation technique for separating nematodes from soil. *Plant Dis* 1964;48:692.
- Kerry, R.; Oliver, MA. Co-kriging when soil and ancillary data are not co-located. *Proc. 4th European Conference of Precision Agriculture*; Wageningen: Academic Publishers; 2003. p. 303–308.
- Khalilian, A.; Mueller, JD.; Blackville, SC.; Han, YJ.; Wolak, FJ. Predicting cotton nematodes distribution utilizing soil electrical conductivity. *Proc. 2001 Beltwide Cotton Conference*; Memphis, TN: National Cotton Council; 2001. p. 146–149.
- Kleinschmidt I, Bagayoko M, Clarke GPY, Craig M, Le Sueur D. A spatial statistical approach to malaria mapping. *Int J Epidemiol* 2000;29:355–361. [PubMed: 10817136]

- Koenning SR, Walters SA, Barker KR. Impact of soil texture on the reproductive and damage potential of *Rotylenchulus reniformis* and *Meloidogyne incognita* on cotton. *J Nematol* 1996;28:527–536. [PubMed: 19277171]
- Lin Y, Chang T, Shih C, Tseng C. Factorial and indicator kriging methods using a geographic information system to delineate spatial variation and pollution sources of soil heavy metals. *Environ Geol* 2002;42:900–909.
- Lyon SW, Lembo AJ Jr, Walter MT, Steenhuis TS. Defining probability of saturation with indicator kriging on hard and soft data. *Adv Water Resour* 2006;29:181–193.
- Matheron, G. Pour une analyse krigéante de données régionalices. Centre de Géostatistique; Fontainebleau: 1982. Report N-732
- McCullagh P. Regression models for ordinal data (with discussion). *J R Stat Soc, Ser B* 1980;42:109–142.
- Melakeberhan H, Dey J, Baligar VC, Carter TE Jr. Effect of soil pH on the pathogenesis of *Heterodera glycines* and *Meloidogyne incognita* on glycine max genotypes. *Nematology* 2004;6:585–592.
- Monfort WS, Kirkpatrick TL, Rothrock CS, Mauromoustakos A. Potential for site-specific management of *Meloidogyne incognita* in cotton using soil textural zones. *J Nematol* 2007;39 (1): 1–8. [PubMed: 19259468]
- Noe JP, Barker KR. Relation of within-field spatial variation of plant-parasitic nematode population densities and edaphic factors. *Phytopathology* 1985;75:247–252.
- Oka Y, Tkachi N, Shuker S, Rosenberg R, Suriano S, Roded L, Fine P. Field studies on the enhancement of nematicidal activity of ammonia-releasing fertilizers by alkaline amendments. *Nematology* 2006;8 (6):881–893.
- Oka Y, Shapira N, Fine P. Control of root-knot nematodes in organic farming systems by organic amendments and soil solarization. *Crop Prot* 2007;26 (10):1556–1565.
- Oliver, MA. The variogram and kriging. In: Fischer, MM.; Getis, A., editors. *Handbook of Applied Spatial Analysis: Software Tools. Methods and Applications*. Springer-Verlag; Berlin, Germany: 2010. p. 319-352.
- Oliver MA, Webster R, Slocum K. Filtering SPOT imagery by kriging analysis. *Int J Remote Sens* 2000;21 (4):735–752.
- Perry, C.; Vellidis, G.; Sullivan, DG.; Rucker, K.; Kemeraite, R. Predicting nematode hotspots using soil EC_a data. Proc. 2006 Beltwide Cotton Conference; San Antonio, TX: National Cotton Council; 2006.
- SAS Institute Inc. SAS OnlineDoc (R) 9. 1. 3. Cary, NC: SAS Institute Inc; 2007.
- Shurtleff, MC.; Averre, CW. Methods. Diagnosing plant diseases caused by nematodes. The American Phytopathological Society; St. Paul, MN, USA: 2000.
- Webster R, Boag B. Geostatistical analysis of cyst nematodes in soil. *J Soil Sci* 1992;43:583–595.
- Wheeler TA, Barker KR, Schneider SM. Yield-loss models for tobacco infected with *Meloidogyne incognita* as affected by soil-moisture. *J Nematol* 1991;23:365–371. [PubMed: 19283140]
- Windham GL, Barker KR. Spatial and temporal interactions of *Meloidogyne incognita* and soybean. *J Nematol* 1993;25(Suppl S):738–745. [PubMed: 19279833]
- Wyse-Pester DY, Wiles LJ, Westra P. The potential for mapping nematode distribution for site-specific management. *J Nematol* 2002;34:80–87. [PubMed: 19265913]

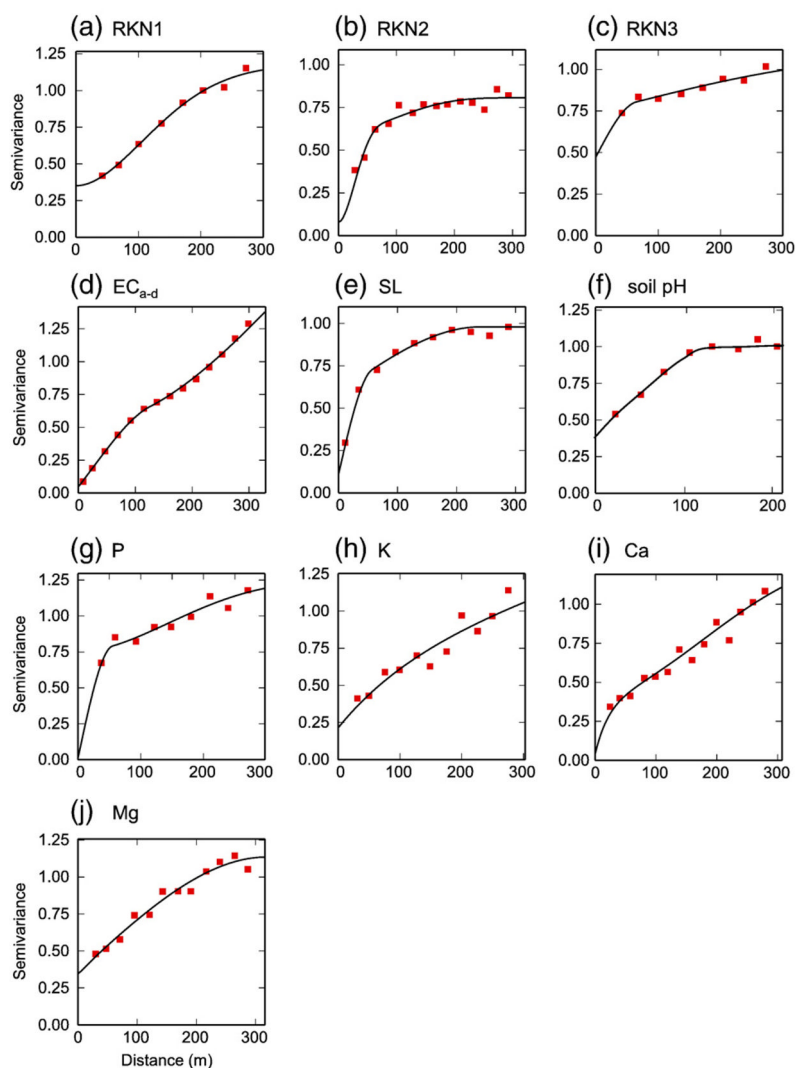
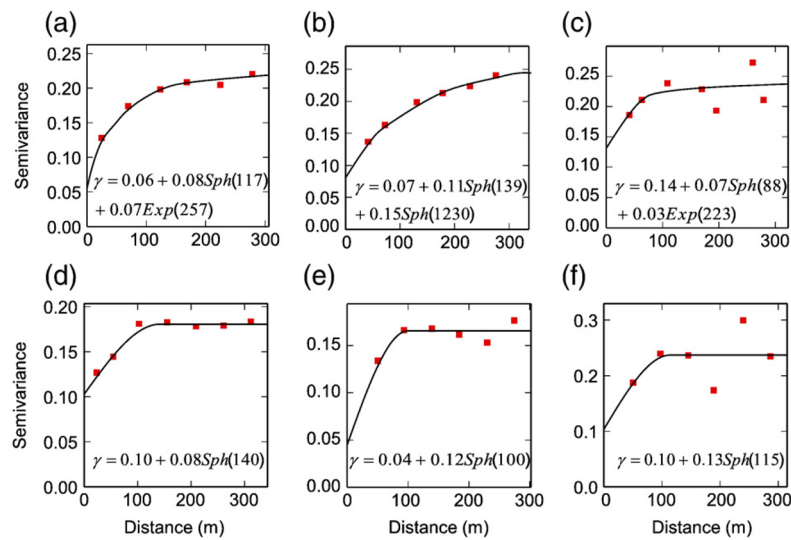
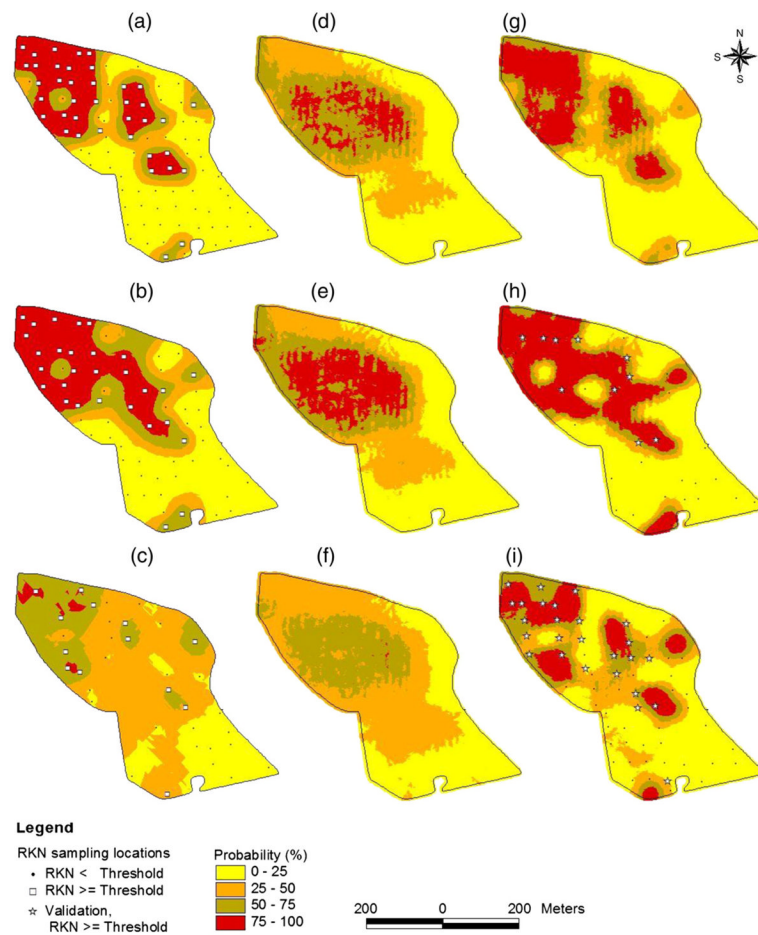


Fig. 1. Normalized semivariograms for cotton RKN samples and soil physical and chemical properties. Squares indicate the empirical semivariogram and the solid line is the fitted model.

**Fig. 2.**

Indicator semivariograms of the RKN2 population above a threshold of 100 RKN second stage juveniles per 100 cm³ soil calculated using 99 (a), 64 (b), and 35 RKN observations (c). Residual semivariograms using soft data and 99 (d), 64 (e), and 35 RKN observations (f).

**Fig. 3.**

Maps of the risk that the population exceeds 100 RKN second stage juveniles per 100 cm³ soil produced using three different algorithms and hard data densities: IK (hard data alone) based on 99 (a), 64 (b), and 35 (c) RKN2 observations, logistic regression using EC_{a-d} as secondary information based on 99 (d), 64 (e), and 35 (f) RKN2 observations, and soft indicator kriging based on 99 (g), 64 (h), and 35 (i) RKN2 observations.

Descriptive statistics of the cotton RKN population density and the soil physical and chemical properties.

Table 1

Variable	Descriptive statistics				
	Mean	Min.–Max.	SD ^a	CV (%) ^b	Skewness
RKN1 ^c	43.9	1–1281	139.4	319.7	7.5
RKN2 ^c	132.1	1–1629	226.4	171.4	4.3
RKN3 ^c	116.2	1–729	145.8	125.5	2.0
EL (m)	78.2	74–81	1.7	2.1	–0.5
SL (%)	1.2	0.13–3.54	0.6	50.4	0.6
EC _{e,d} (mS/m)	1.2	0.45–7.95	1.1	89.5	4.0
soil pH	6.4	5.34–7.49	0.4	6.5	0.1
P (kg/ha)	83.9	45–141	20.1	24	0.7
K (kg/ha)	80.2	41–163	22.5	28	1.2
Ca (kg/ha)	897.0	276–1924	324.9	36.2	0.6
Mg (kg/ha)	122.1	30–345	55.0	45	1.3

^aStandard deviation.

^bCoefficient of variation, percentage.

^cCounts of second stage juveniles per 100 cm³ soil.

Table 2

Scale-dependent correlation of cotton RKN population density with soil physical and chemical properties.

Variables	Pearson correlation coefficient	Scale-dependent correlation coefficients		
		Spatial components		
		Short-range	Regional ^a	
<i>RKN1^b with</i>				
EC _{a-d} ^b	−0.39	0.06	−0.54	
SL	−0.35	−0.19	−0.61	
Soil pH	−0.19	0.08		
P	−0.09	−0.04	−0.07	
K§	−0.07	−0.23	0.00	
Ca	−0.30	−0.08	−0.46	
Mg ^b	−0.21	−0.06	−0.31	
<i>RKN2^b with</i>				
EC _{a-d} ^b	−0.50	0.17	−0.77	
SL	−0.37	0.05	−0.59	
soil pH	−0.26	−0.04		
P	−0.09	−0.03	−0.18	
K ^b	−0.28	−0.05	−0.46	
Ca	−0.45	−0.08	−0.80	
Mg ^b	−0.35	−0.06	−0.63	
<i>RKN3^b with</i>				
EC _{a-d} ^b	−0.26	0.01	−0.61	
SL	−0.19	−0.14	−0.42	
Soil pH	−0.23	−0.11		
P	0.04	0.00	0.19	
K ^b	−0.15	0.04	−0.55	
Ca	−0.22	−0.04	−0.69	
Mg ^b	−0.20	−0.05	−0.53	

^a Regional component corresponds to the long-range component plus local mean or trend component.^b Spatial correlation was performed on normal score transformed data.

Occurrence, in number and percentage, of RKN observations above (Risk) or below (No Risk) the management threshold on various estimated ranges of risk probabilities.^a

Table 3

Estimated probability of RKN above threshold (%)	Hard data		Soft data		Hard and soft data	
	Risk		No Risk		Risk	
	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
<i>99 Observations</i>						
0–25	0 (0)	58 (97)	3 (8)	35 (58)	0 (0)	52 (86)
25–50	0 (0)	2 (3)	12 (31)	14 (23)	0 (0)	7 (12)
50–75	5 (13)	0 (0)	17 (43)	10 (16)	9 (33)	1 (2)
75–100	34 (87)	0 (0)	7 (18)	3 (5)	30 (59)	0 (0)
<i>64 Observations</i>						
0–25	0 (0)	30 (86)	3 (10)	18 (52)	0 (0)	35 (100)
25–50	0 (0)	5 (14)	6 (21)	11 (31)	0 (0)	0 (0)
50–75	6 (21)	1 (3)	11 (38)	4 (11)	0 (0)	0 (0)
75–100	22 (76)	0 (0)	9 (31)	2 (6)	29 (100)	0 (0)
<i>35 Observations</i>						
0–25	0 (0)	15 (68)	0 (0)	9 (41)	0 (0)	22 (100)
25–50	0 (0)	7 (32)	9 (69)	9 (41)	0 (0)	0 (0)
50–75	11 (85)	0 (0)	4 (31)	4 (18)	0 (0)	0 (0)
75–100	2 (15)	0 (0)	0 (0)	0 (0)	13 (100)	0 (0)

^aThe management threshold value in Georgia is 100 RKN second stage juveniles per 100 cm³ of soil.

Results of jackknife validation comparing three indicator kriging mapping strategies – hard, soft, and hard plus soft data – for identifying RKN population areas above the management threshold.^a

Table 4

Estimated probability of RKN above threshold (%)	Hard data		Soft data		Hard and soft data	
	Risk	No Risk	Risk	No Risk	Risk	No Risk
	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
<i>64 observations for mapping and 35 observations for validation</i>						
0–25	0 (0)	12 (48)	0 (0)	14 (56)	0 (0)	15 (60)
25–50	0 (20)	5 (20)	2 (20)	5 (20)	1 (10)	2 (8)
50–75	5 (50)	5 (20)	5 (50)	4 (16)	1 (10)	4 (16)
75–100	5 (50)	3 (12)	3 (30)	2 (8)	8 (80)	4 (16)
<i>35 observations for mapping and 64 observations for validation</i>						
0–25	1 (4)	11 (29)	1 (4)	20 (52)	1 (4)	20 (52)
25–50	11 (42)	25 (66)	13 (50)	17 (45)	6 (23)	14 (37)
50–75	13 (50)	2 (5)	10 (38)	1 (3)	9 (35)	4 (11)
75–100	1 (4)	0 (0)	2 (8)	0 (0)	10 (38)	0 (6)

^aData in parentheses correspond to the percentage of observations above or below the threshold for each probability class.