SPATIAL PATTERN DETECTION MODELING OF ONION THRIPS (THRIPS TABACI) ON ONION FIELDS

Paulo Justiniano Ribeiro Jr[1]; Denise Nunes Viola[2]; Clarice Garcia Borges Demétrio[3*]; Bryan F. Manly[3]; Odair Aparecido Fernandes[4]

[1]*UFPR - Lab. de Estatística e Geoinformação - C.P. 19.081 - 81531-990 - Curitiba, PR - Brasil.*
2UFBA/DES - Depto. de Estatística, Av. Adhemar de Barros, s/n - Campus de Ondina - 40170-110 - Salvador, BA - Brasil.
[3]*USP/ESALQ - Depto. de Ciências Exatas, Caixa Postal 9 - 13418-900 - Piracicaba, SP – Brasil.*
[4]*UNESP/FCAV - Depto. de Fitossanidade, Via de Acesso Prof. Paulo Donato Castellane s/n, 14884-900, Jaboticabal, SP- Brasil.*
[*]*Corresponding author <clarice@esalq.usp.br>*

SPATIAL PATTERN DETECTION MODELING OF ONION THRIPS (THRIPS TABACI) ON ONION FIELDS

ABSTRACT: Onion (*Allium caepa*) is one of the most cultivated and consumed vegetables in Brazil and its importance is due to the large workforce involved. One of the main pest that affect this crop is the onion thrips *(Thrips tabaci)*, but the spatial distribution of the insect, although important, has not been considered in crop management recommendations, experiment planning or sampling plans. Our purpose here is to consider statistical tools to detect and model spatial patterns in the occurrence of onion thrips. In order to characterize the spatial distribution pattern of the onion thrips a survey was carried out to record the number of insects in each development phase on onion plant leaves, on different dates and sample locations, in four rural properties with neighboring farms with different infestation levels and planting methods. The Mantel randomization test proved to be a useful tool to test for spatial correlation which when detected was described by a mixed spatial Poisson model with a geostatistical random component and parameters allowing for a characterization of the spatial pattern as well as the production of prediction maps of susceptibility to levels of infestation throughout the area.

Key words: spatial statistics, randomization tests, geostatistics, Poisson distribution

DETECÇÃO DE PADRÕES ESPACIAIS NA OCORRÊNCIA DO TRIPES DO PRATEAMENTO *THRIPS TABACI* NA CULTURA DA CEBOLA

RESUMO: A cebola é uma das hortaliças mais cultivadas e consumidas no Brasil e sua importância social se deve à grande demanda por mão-de-obra. Uma das principais pragas que afeta essa cultura é o tripes do prateamento (*Thrips tabaci*) e sua distribuição espacial, embora importante, não tem sido considerada nas recomendações de manejo da cultura, planejamento de experimentos ou estudos amostrais. O objetivo desse artigo foi considerar métodos estatísticos para detectar e modelar padrões espaciais na ocorrência do tripes do prateamento da cebola. Para caracterizar o padrão espacial da dispersão do tripes do prateamento da cebola foi feito um levantamento anotando-se o número de insetos por fase de desenvolvimento em folhas de plantas de cebola, em diferentes datas e pontos amostrais dentro de quatro propriedades com fazendas vizinhas apresentando diferentes níveis de infestação e métodos de plantio. O teste de aleatorização de Mantel mostrou-se útil para testar a presença de padrão espacial, que quando detectado foi descrito por um modelo de Poisson misto espacial com componente aleatório geoestatístico com parâmetros que possibilititam a caracterização do padrão espacial bem como a obtenção de mapas de predição dos níveis de susceptibilidade à infestação na área.

Palavras-chave: estatística espacial, testes de aleatorização, geoestatística, distribuição de Poisson

INTRODUCTION

Onion (Allium caepa) is one of the most cultivated and consumed vegetables in Brazil. The social importance of the crop is due to the large workforce involved. It is estimated that 70% of the production is small scale, because it is typically grown on small and medium sized properties. It is an annual plant for bulb production, biannual for seed production, and propagated by direct sowing, bulbs or seedlings planted in beds and transplanted to the field.

One of the main pest that affects onion crops is the onion thrips *(Thrips tabaci),* which in high infestation levels can damage the harvest (Workman & Martin, 2002), with reduction in the production of up to 80% during hot and dry periods (Sato, 1989). The insect is typically found at the base of leaves. It feeds from the sap and the leaves parenchyma causing gray spots, which gradually change to silver as a result of the external tissue damage of the leaves. Massive attacks on the aerial part of the plant cause loss in bulb production, which reduces the size and quality, damaging the commercial value and creating obstacles to exports. When an attack is very intense, the leaves get yellowish, dry and with wrenched tips, causing the wilting and the death of the plant (Sato, 1989), and also allowing for the entrance of water to the bulb, which gets rotten. The insect is also considered a vector of a phytopathological agent with the capacity to transmit a virus to the plant.

The insect development occurs in the four phases of egg, nymph, pupa and adult, with the nymph and adult stages damaging the production, because the pupa phase is restricted to the soil. The nymph has low mobility, while the adult, although winged, has restricted movement. The development cycle varies typically from 14 to 30 days, changing to 10 and 11 days when the temperature is over 30^{o}C.

The spatial distribution of thrips in commercial fields is important for the efficient application of insecticides. However, this has not been considered in crop management recommendations, experiment planning and sampling plans. Considering the low mobility of nymphs and adults it is reasonable to assume that the wind is the main dispersion factor for the thrips that potentially determines the spatial pattern.

A spatial pattern can be classified as random, aggregate or uniform. The random pattern occurs when there is a constant and independent probability of infestation for all the plants, while the aggregate pattern is associated with low insect mobility. The uniform pattern rarely occurs naturally, but can be induced, for instance, by alternated planting of resistant and susceptible plants. In order to study whether infant leukemia cases tend to be close in space and time, Mantel (1967) proposed a randomization test, based on matrices of time and space distances between observations. This test can be used to test for spatial correlation in an insect distribution, but its usage has not being considered in practical applications, and in particular, in studies of the spatial distribution of the onion thrips.

It is common in insect distribution studies, to find the use of indices based on the relationship between the variance and the mean, such as the David & More index, the Taylor power law, and the aggregate indices of Lloyd and Iwao, among others (Ruiz et al., 2003). However, these indices ignore the spatial location of the samples, have limited capacity to describe spatial patterns, and strongly depend on the size of the sample unit.

Geostatistical methods (Isaaks & Srisvastava, 1989; Goovaerts, 1997) have been used to describe insect spatial patterns as, for instance, in Grego et. al. (2006). Such methods were originally developed for continuous response variables, with several computational implementations available for data analysis. The insect counts are discrete and typically distributed in clusters, with many zero counts. Therefore, the data cannot have a covariance structure of the type assumed by traditional methods of geostatistical analysis, with a stationary spatial covariance structure in the study area (Ruiz, 2002). For this reason it is appropriate to use models that incorporate explicitly a data generating mechanism such as the Poisson distribution, combined with structures that describe the spatial pattern of the counts. These kinds of models have been proposed in the statistical literature (Diggle et al., 1998) but have had few practical applications.

This paper describes a study of the spatial distribution of onion thrips with data from surveys of four different properties with different infestation levels and planting methods. We aimed to detect spatial patterns in the occurrence of onion thrips at different production fields and propose an statistical model for such patterns. We adopt the Mantel randomization test (Manly, 2006) to decide for the presence of spatial autocorrelation which when detected was modeled by a mixed spatial Poisson model with a random term given by a geostatistical component. This model allows the characterization of the spatial pattern as well as the production of maps of levels of susceptibility to infestation in different areas.

MATERIAL AND METHODS

Data description

This work is motivated by a set of data originated from a study involving sampling onion thrips in onion crop in four different farms, located in the municipality of São José do Rio Pardo, São Paulo State, Brazil (21°36'S, 43°53'W; altitude 705m), from June to September, 1996. The aim is to study the spatial and temporal distribution of thrips. The four chosen properties used the onion hybrid Granex 33 and the seedling planting method. The trial areas were chosen with neighbors who adopted different kinds of planting and had different infestation levels.

Details referring to the kind of planting in the neighborhood and collection dates and numbers of samples collected in the different farms are shown in the Table 1. The São Paulo farm is located at a high elevation of the region and the nearest neighboring onion crop is situated over one kilometer away. The neighborhood of Estância Bela Vista had already had some crops attacked by onion thrips pest.

Table 1 - Characteristics about the data precedence, types of neighbors, sample times and number of samples.

Farm	Neighborhood	Sampling dates	Number of samples
Fazenda São Paulo	Isolated from other plantings	07/10, 07/24, 07/31, 08/07, 08/14, 08/21, 08/28, 09/04	100, 100, 100, 98 100, 100, 100, 100
Estância Bela Vista	Bulbs	07/11, 08/01, 08/08, 08/14, 09/09	100, 100, 84, 99, 99
Sítio Rosário	Seedlings	06/21, 06/29, 07/07, 07/14, 07/21, 07/28, 08/04, 08/11, 08/18, 08/25, 09/03	50, 50, 48, 50, 50, 50, 50, 50, 50, 50, 50
Sítio Novo II	Seedlings	06/04, 06/19, 06/27, 06/28, 07/04, 07/11, 07/24, 07/31, 08/07	100, 100, 100, 100, 100, 100, 100, 100, 100

The sampling unit was a 1m radius circle with a center stake. One plant was then randomly selected from within the circle. The position of the stakes in the four studied farms, in general with a 10x10m grid, but with some variations at Fazenda São Paulo is shown in Figure 1. The measured variables were the stake location on the coordinate axes, the number of nymphs, the number of adult insects and the number of leaves per plant. The number of samples and sampling times varied from farm to farm as shown in Table 1. The response variables are discrete because of result of counting. In some cases, the counts are multiples of 5 or 10 and some values over 100 were truncated to 100.

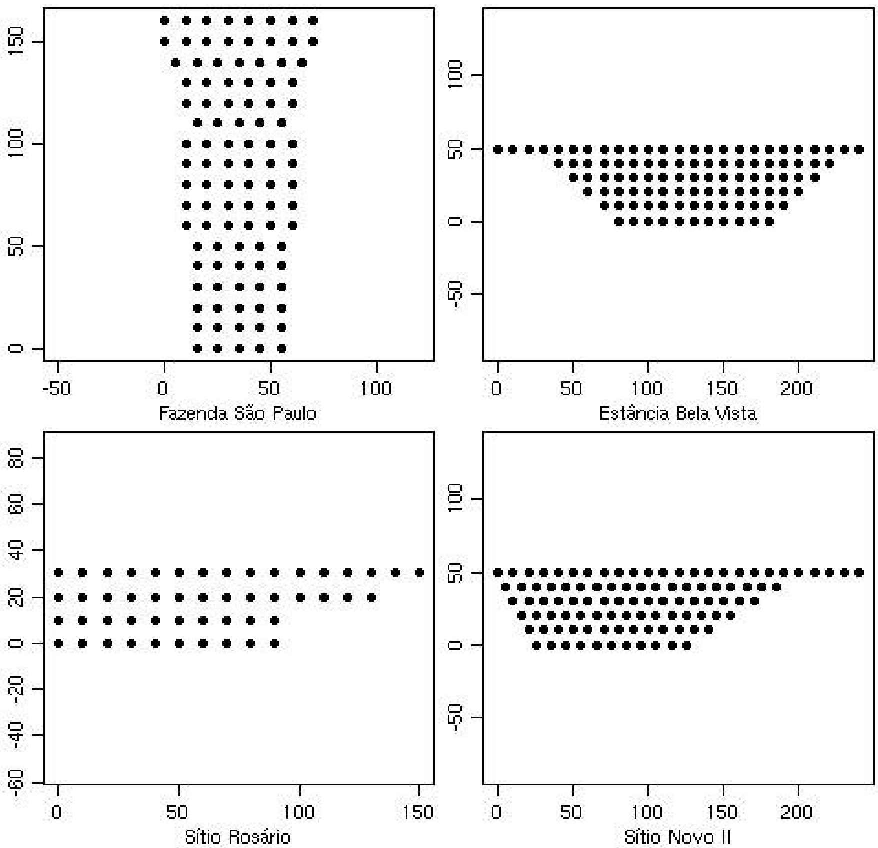

Figure 1- Localization of the stakes in each farm.

Figure 2 shows box-plots for the average number of insects per leaf, at the four farms, for the various sample times. There is great variability in the counts and also some outliers, not all of them being influential on the model fitting. At the São Paulo farm the average number of insects and the variability increased with time, while at the other farms, the average increased and then decreased. In all cases the observations above the median are more variable showing positive asymmetry, with some extreme values.

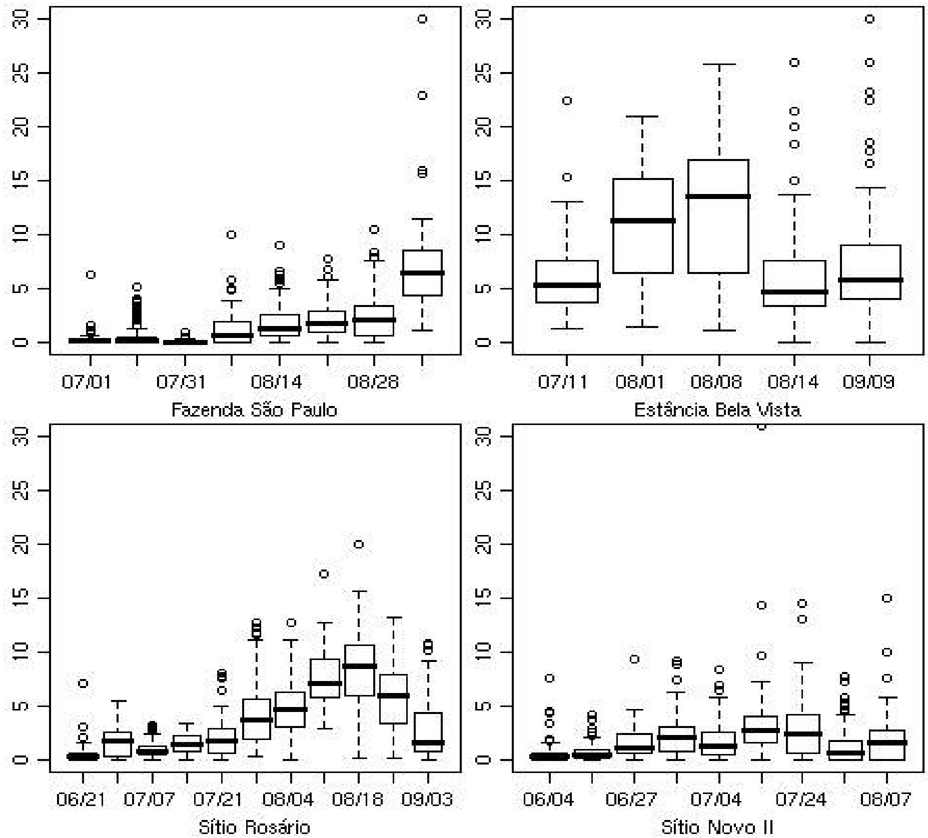

Figure 2 - Box-plots for the average number of insects per leaf.

At the São Paulo farm the lowest average number of insects per leaf and also the lowest variance wer found at 07/31, with one insect per leaf as the maximum value. In contrast on 09/04 this farm had a much larger average number of insects per leaf and much greater variability. The percentage of infested plants ranged from 35% to 100%. For the Estância Bela Vista, the lowest average number of insects per leaf occurred on 07/11 and 08/14 with 89% to 100% plants infested. The Rosário farm had only 50 plants sampled and the highest average number of insects per leaf on 08/11 and 08/18. Sítio Novo II had the least average for the number of insects per leaf with low variability except for one outlier count of 30.

Mantel's test for the detection of spatial pattern

The non existence of spatial pattern in the dispersion of insects may be considered a randomization hypothesis, and the existence of a spatial pattern can be tested through the randomization of the order of the observed values (Manly, 2006).

Randomization tests are based on the fact that, if the null hypothesis is true, then all of the possible orders of the data have the same chance of occurrence. Therefore, the value e_o of a statistic E is calculated for a set of observations, and then a large number of randomizations are made. For spatial data these randomizations are made by randomly reordering the data. For each randomization a value e_a is calculated and the set of the e_a values generate an approximation of the randomization distribution of E. Just as for classic statistical tests, the decision is guided by a *p-value*, which in the case of randomized tests is given by the proportion of the e_a values that are larger than or equal to e_o, for a one-sided test. For instance, if $p < 0.05$, it's concluded that there is evidence that the null hypothesis is not true (Manly, 2006).

Randomized tests have some advantage in comparison to classic statistical tests. For example, the statistics are usually easy to calculate, relatively to the classic statistical tests. They are based on non standard statistics and they do not need previous information about the population from which the samples were taken. Also, they can be applied with non-random samples which can consist only of the data that need to be analyzed (Manly, 2006). However, the randomization tests are easier to justify when the analyzed samples are random or the experimental design suggests a randomization test.

Usually, when considering spatial data, it is desired to test the null hypothesis of a random spatial pattern *versus* the alternative of a non-random spatial pattern. A test for this hypothesis was proposed by Mantel (1967). The test is implemented as follows. Let a variable be observed in n locations. Two symmetric matrices A and B are obtained, each with n x n dimensions. The elements represent distances between the observations. These matrices can be denoted as

$$A = \begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}.$$

The matrix A is the Euclidian distances between the stakes with locations given by $(x_{1\mathrm{i}}, x_{2i})$ and $(x_{1\mathrm{j}}, x_{2j})$, i.e., with elements of the form $a_{ij} = \sqrt{(x_{1\mathrm{i}} - x_{1\mathrm{j}})^2 + (x_{2\mathrm{i}} - x_{2\mathrm{j}})^2}$ and B is the matrix with elements $b_{ij} = \sqrt{(z_i - z_j)^2}$, where Z is the mean of the number of insects per leaf. The test statistic is given by the Pearson correlation coefficient between the correspondent elements of A and B, i.e.,

$$r = \frac{m\sum_{i<j} a_{ij} b_{ij} - \sum_{i<j} a_{ij} \sum_{i<j} b_{ij}}{\sqrt{\left[m\sum_{i<j} a_{ij}^2 - (\sum_{i<j} a_{ij})^2\right]\left[m\sum_{i<j} b_{ij}^2 - (\sum_{i<j} b_{ij})^2\right]}}, \quad (1)$$

which produces the r_0 value when calculated for the observed values. For the randomization test the rows and columns of one of the matrices are permutated a large number (N) of times, and the values r_{ak} are obtained, for $k = 1, 2, \dots, N$. The proportion p of values $r_{ak} > r_0$ is then compared with a pre-fixed significance level α (for example, 0.05) and the null hypothesis is rejected if $p < \alpha$ (Manly, 2006).

As the matrices A and B are symmetric, the correlation amongst all the elements outside the main diagonal is the same as the correlation of the $m = \frac{n(n-1)}{2}$ elements in the upper or lower triangular part of the matrix. Note that the only term of (1) that is altered by changing the order of the elements in one of the two matrices is the sum of products $Z = \sum a_{ij} b_{ij}$.

Other possible metrics used for the calculation of the distances are *Euclidian with standardized data, Euclidian squared, Euclidian squared with standardized data, proportional distance* and *sample difference*. The alternative is given by Snäll et. al. (2003) who built a randomized test using flexible forms for the relation between the distance measurements, given by the structure of additive generalized models.

When the Mantel test rejects the null hypothesis there may be interest in knowing the kind of association amongst the variables. This can be shown by the graph of b_{ij} *versus* a_{ij}. One of the possible models of association is the simple linear regression, in which the elements of the *A* matrix give an explanatory variable and the elements of the *B* matrix a response variable, so that,

$$b_{ij} = \beta_0 + \beta_1 a_{ij} + \varepsilon_{ij}$$

where β_0 e β_1 are parameters to be estimated and ε_{ij} is the error associated with the response assumed to be Gaussian, independently and identically distributed. This assumption is a pragmatic approach avoiding more complex structures for the error term which would require further modeling assumptions we wish to avoid at this exploratory stage. Also, more complex forms of spatial dependence than given by the linear relation can, in principle, also occur. Our approach is to rely on simple assumptions for the randomization tests and leaving more complex structures to be considered by the model discussed in the next Section.

In this study, the randomization test for spatial pattern was carried out on the observations for each sampling date. The test can be extended for the detection of time patterns. However, this raises the question of how to combine the information from several units of observation. Although such alternative has been studied for the data on thrips occurrences, it was decided not to include the results here because of the small number of observations in time and the lack of a specific interest in testing for time patterns.

Modeling the spatial pattern

Having detected a spatial pattern, it may be of interest to describe the pattern by means of a stochastic model. Modeling allows not only the characterization of the dependence pattern but also for the prediction of quantities of interest such as a map of expected levels of infestation over the area, the proportion of the area with infestation above or below a certain threshold, and areas with high and low infestation, among others possible quantities of interest.

One possible way of modeling the spatial distribution is by adopting the geostatistical framework, which associates the level of spatial dependency with distances between sampled plots. Usually the description of the spatial dependence assumes that the closest sampled plots are more alike than those farthest apart (Montagna, 2001). Diggle et al. (2003) uses the term geostatistics to identify a part of the spatial statistical methods in which the used model describes a continuous variation of the observations over the space.

The basic geostatistical data format is (x_i, y_i), $i=1, 2, ..., n$, in which $x_i = (x_{1i}, x_{2i})$ identifies the spatial location, generally in the two dimensions and y_i is the measure of interest at the x_i position of the *i*th observation. The response variable can be potentially measured at any point within the studied region (Diggle & Ribeiro Jr., 2007).

The geostatistical model is specified by assuming two processes over the study region (Diggle et al., 1998; Diggle & Ribeiro Jr., 2007) described as follows. $Y(x): x \in A$ is a measure process within the study region *A* which is observed at a set of locations x to obtain the y_i's, the observed data. This first process is related to a underlying Gaussian process $S = \{S(x): x \in R^2\}$ with mean μ, variance σ^2 and correlation function $\varphi(u)$, where u is the distance between pairs of observations. The values of $S(x)$ are usually not directly observed. Conditional independence is assumed in the sense that the $Y(x)$ are independent, conditionally on the values of $S(x)$ meaning all the spatial dependency is modeled through $S(x)$. The exact form of the relation between the two processes may vary according to the type of variable being measured. For instance, when Y follows the Gaussian distribution, the model can be written as $Y_i = S(x_i) + Z_i$, in which the Z_i values are mutually independent and follow the normal distribution, with mean 0 and variance τ^2. In this case the observations y_i can be seen as a *noisy* version of $S(x_i)$ at the location x_i, and, for a finite set of plots, the random vector Y follows a multivariate Gaussian distribution. More generally, Y may follow other distributions and Diggle et al., (1998) specify a model within the class of the generalized linear model (McCullagh & Nelder, 1989) in which the *S* process defines random effects with spatial dependence structure. Diggle and Ribeiro Jr. (2007) call this a generalized linear geostatistical model (GLGM). This model allows the explicit specification of a Poisson distribution for the observations, which is compatible with the insect counting structure of the data considered here.

The GLGM is a special case of a mixed generalized linear model, in which the Y_i,

i=1, 2, …, n are conditionally independent given $S(x)$, with expected values given by $E[Y_i \mid S(x)] = \lambda_i$ and linear predictor $h(\lambda_i) = S(x_i)$, i=1, 2, …, n with a known link function $h(.)$, which, for the Poison model considered here is typically given by the logarithm function. The model can extended allowing for covariates considering $S(x_i) = S(x_i) + d(x_i)^T \beta$, in which $d(x_i)$ is the observed covariate values and β is the regression parameter vector. (Diggle et al., 1998, Diggle et al., 2003).

Let $Y(x_i) \mid S(x_i)$ be the observed total number of insects with a Poisson distribution with mean $t_i \exp[S(x_i)]$, i=1, 2, …, n in which t_{i} represents the number of leaves. Then the probability function is given by

$$P[Y(x) \mid S(x) = s(x_i)] = \frac{e^{-t_i e^{S(x_i)}} \left(t_i e^{S(x_i)}\right)^{y(x_i)}}{y(x_i)!}.$$

The likelihood function is often considered for inference about the model parameters within the context of generalised linear models. However, in this case the likelihood function does not have a closed form and is given by

$$L = \int \prod_{i=1}^{n} \frac{e^{-t_i e^{S(x_i)}} \left(t_i e^{S(x_i)}\right)^{y(x_i)}}{y(x_i)!} \frac{1}{\sqrt{2\pi \mid \sigma^2 R \mid}} e^{-\frac{1}{2\sigma^2}[S(x_i)-\mu]^T R [S(x_i)-\mu]} ds$$

where R is the correlation matrix for S and with dimension equal to the number of observations which cannot be solved by analytical or numerical methods. Each element of R is given by the corresponding value of the correlation function of the S process and therefore having model parameters within non-linear functions which explains the lack of such solutions. A possible solution is to use Monte Carlo Markov Chain (MCMC) methods and a computational implementation is available through the package **geoRglm** (Christensen & Ribeiro Jr., 2002) for the **R** statistical environment (R Development Core

Team, 2007).

For discrete random variables, the variogram is not a natural summary of the data, but it may be useful as a diagnosis tool, after fitting the mixed generalized linear model (Diggle & Ribeiro Jr., 2007). In this case, the variogram obtained from the estimated parameters can be compared to the experimental variogram, obtained through the residuals from a GLM model fit. The variogram if given, respectively, by

$$\gamma_Y(h) \quad = \quad \frac{1}{2}Var[Y(x)] + \frac{1}{2}Var[Y(x+h)] - Cov[Y(x), Y(x+h)]$$

which can be written as

$$\gamma_Y(h) \quad = \quad \exp(\beta + \frac{\sigma^2}{2}) + \exp(2\beta + \sigma^2)\{\exp(\sigma^2) - \exp[\sigma^2 \rho(u)]\}.$$

However, this approach must be used with caution because the variogram is even more erratic then the one usually obtained for data with a symmetric and continuous distribution, because of the asymmetric data.

After the choice of a specific model, a map that describes the behaviour of the study variable over the region can be obtained. Supposing that the parameters are known and that the interest is in the expected insects number given by $\lambda(x_0) = \exp[\beta + S(x_0)]$, for the location $x_0 = (x_{10}, x_{20})$, from the S marginal distribution and the $Y \mid S$ conditional distribution, it is possible to simulate the conditional distribution of $[S \mid y]$, using the MCMC method. The predicted surface is given (Diggle et al., 1998) by

$$\hat{\beta} + \hat{S}(x) + \frac{Var(x)}{2},$$

where $\hat{\beta}$ is the process mean in this case because there are no explanatory variables or trend, and $\hat{S}(x)$ is the linear kriging predictor and $Var(x)$ is the prediction variance.

RESULTS AND DISCUSSION

Spatial pattern detection through Mantel's randomization test

The Mantel's test was applied separately for each sampling date and each farm and the obtained *p*-values contrasted with the adopted 5% significance level. For the Fazenda São Paulo, there was evidence of spatial pattern in the number of insects per leaf for the first three data collections on 10^{th} , 24^{th} and 31^{th} of July. These patterns can be observed in the dispersion plots shown in Figure 3 where symbols sizes are proportional to the number of insects per leaf. In general, considering all the farms and dates, the distribution of the mean number of insects per leaf is asymmetric and, does not show trends against the spatial coordinates. Also, the linear regression between the number of insects by leaf distances and the stakes location distances shows that, for the above mentioned dates, there is evidence of positive association in conformity with Table 2 which shows, as well, analogous results for the dates that detected spatial pattern at the other farms.

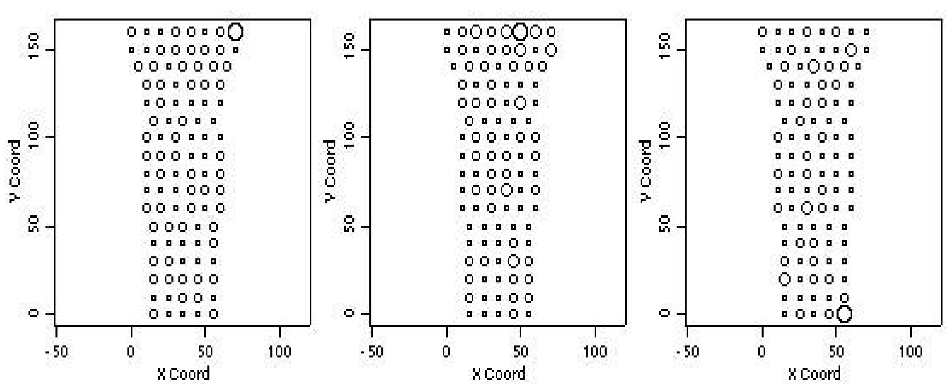

Figure 3 – Dispersion graphs for the mean number of insects, Fazenda São Paulo (symbol sizes are proportional to the number of insects per leaf).

Table 2 – Regression models for the distance matrices from the randomization test.

Farm	Data	Model	*p*-value
Fazenda São Paulo	10/07	Insects/leaf=0.2102+0.002325loc	0.0205
	24/07	Insects/leaf=0.6024+0.004216loc	0.0022
	31/07	Insects/leaf=0.0932+0.000417loc	0.0264
Estância Bela Vista	08/08	Insects/leaf=6.2180+0.009037loc	0.0334
Sítio Novo II	04/06	Insects/leaf=0.3035+0.007206loc	0.0012
	27/06	Insects/leaf=1.1810+0.004034loc	0.0258
	04/07	Insects/leaf=1.5240+0.003371loc	0.0455

For Estância Bela Vista, the spatial pattern was detected only for the third data collection on 8th of August. The dispersion plot for this date are shown in Figure 4. For Sítio Rosário, evidence of spatial patterns was not found for any of the dates. At least, analysis for Sítio Novo II, suggests presence of spatial pattern for the 2nd, 4th and 6th data collections on 4th and 27th of June and for 4th of July, with data shown in Figure 5.

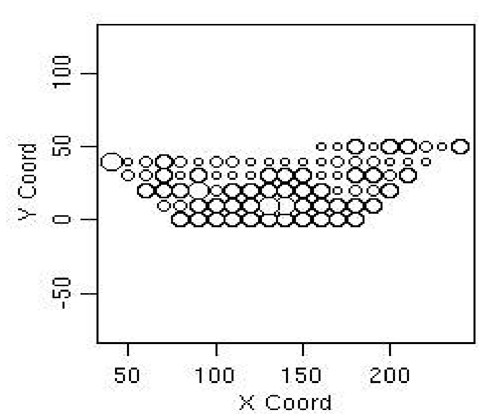

Figure 4 – Dispersion graphs for the mean number of insects, Estância Bela Vista (symbol sizes are proportional to the number of insects per leaf).

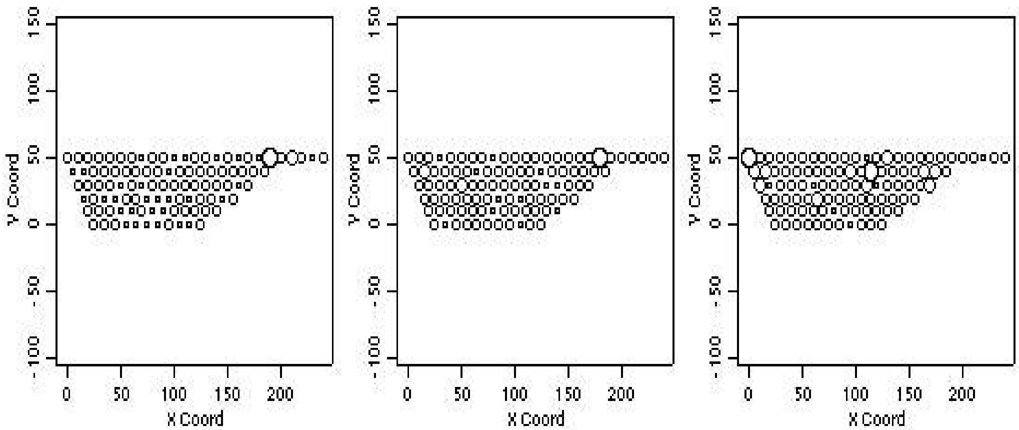

Figure 5 – Dispersion graphs for the mean number of insects, Sítio Novo II (symbol sizes are proportional to the number of insects per leaf).

Geostatistical generalized linear models with Poisson distributions and logarithmic link functions were used for the modeling for the data for farms and dates that showed some evidence of spatial pattern. Maximum likelihood parameter estimates were obtained by the MCMC algorithm and results are summarized in Table 3. A total of 120,000 iterations chains were obtained, with a burn in cycle of 20,000, keeping the first of every 100 generated samples, amounting to a total of 1,000 samples. The obtained chain for each parameter was analyzed to verify the convergence of the MCMC algorithm. The estimates for the ϕ parameter reflects the spatial correlation, and for the case of an exponential correlation model the *practical range* of spatial dependence corresponds to three times the parameter value. The interpretation of the extent of the correlation also depends on the distances between points within the area, which vary from 10 to 170 meters at the Fazenda São Paulo, 10 to 200 at the Estância Bela Vista and 10 to 204 meters at the Sítio Novo II. There were cases in which the estimate is smaller than the minimum distance between sampled points, reflecting short range correlation which would be better captured with sampling points at closer locations.

Table 3 - Point estimates and confidence intervals for the parameters of the geoestatistical model.

Farm	Date	β	σ^2	ϕ	τ^2
Fazenda São Paulo	10/07	-1.55	0.15	30.33	0.47
	24/07	-1.11	1.25	18.00	0.00
	31/07	-1.50	0.15	3.25	0.00
Estância Bela Vista	08/08	2.35	0.19	18.15	0.95
Sítio Novo II	04/06	-0.73	0.62	50.00	1.11
	27/06	0.26	0.37	19.08	0.14
	04/07	0.34	0.53	22.35	0.14

The parameter β is associated with the link function and σ^2, ϕ and τ^2 are parameters associated with the surface $S(x)$. Outliers values at a location on the top right corner of the area were removed for Fazenda São Paulo since this local feature was highly influential on the global model. The negative values for the estimates of β parameter at the Fazenda São Paulo reflect the fact that this farm was isolated from other onion plantations, which resulted in low means of infestation. High values of the estimates were observed at the Estância Bela Vista, which was surrounded by onion plantations infested by thrips. At the Sítio Novo II estimates near zero were the result of the low mean for the number of insects per leaf.

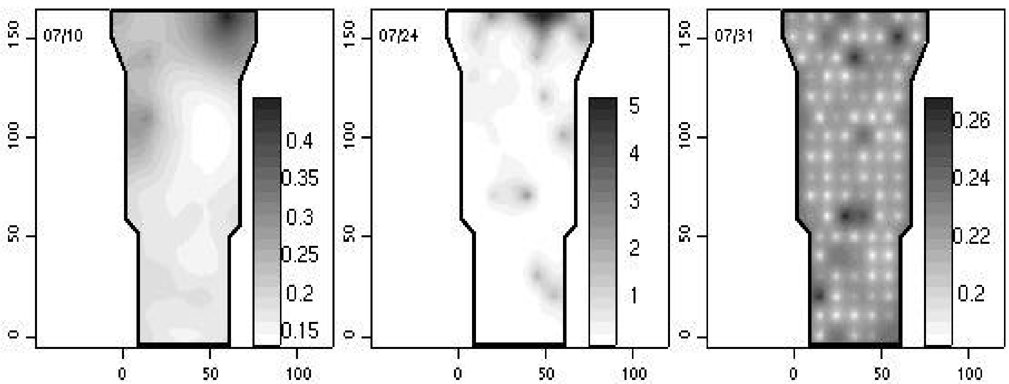

Figure 6 - Prediction maps for Fazenda São Paulo.

From the fitted models prediction maps of the susceptibility infestation in the area were produced. Comparing the prediction maps showed in Figures 6, 7 and 8 where the lighter colours indicate low infestation and the dark colours indicate high infestation with the dispersion plot in Figure 3, Figure 4 and Figure 5 it is possible to see a pattern in the second, as the low and high infestation areas are the same. The white points on the prediction map shown on the right hand panel of the Figure 6 are centered on the sampling points as an artefact of the fitted model.

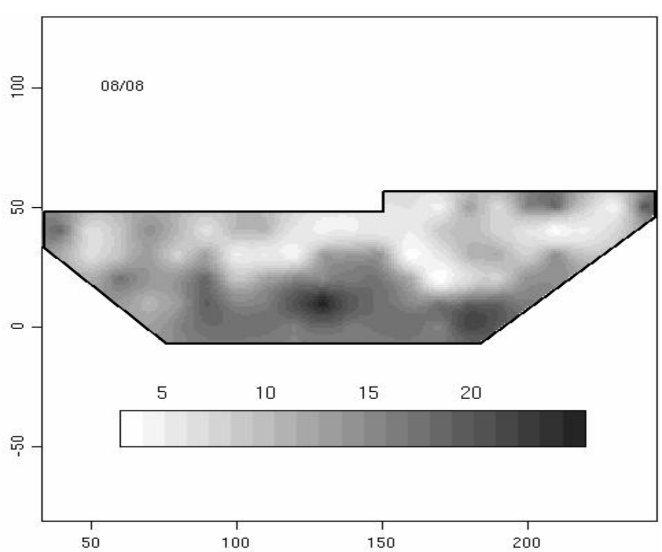

Figure 7 - Prediction maps for Estância Bela Vista.

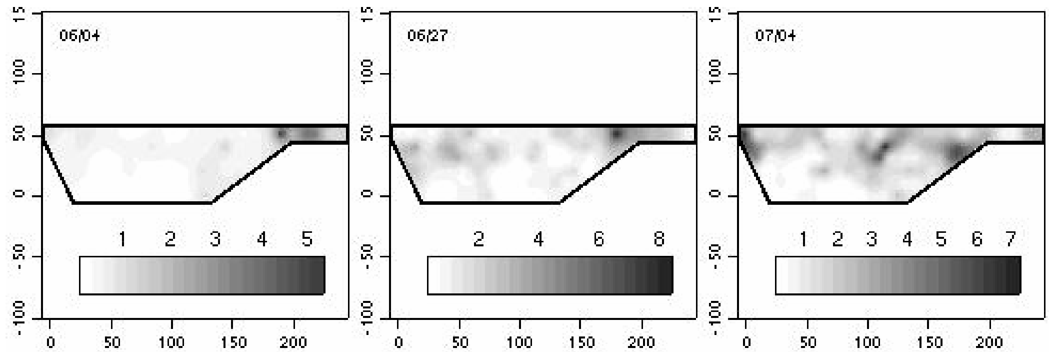

Figure 8 - Prediction maps for Sítio Novo II

Apparently there is some influence of the kind of the plantation in the neighborhood on the number of insects per leaf on plants. Estância Bela Vista had as the neighborhood an area already infested with thrips and showed the highest means for the number of insects per leaf and the greatest proportions of infested plants, whereas Fazenda São Paulo, isolated from other plantations of onion, was the one with the smallest proportion of infested plants, however increasing with times. This conjecture cannot be tested statistically with the available data, but can be considered for future studies.

CONCLUSIONS

The adopted methods allow for testing for the presence of spatial patterns in the distributions of onion thrips using Mantel's randomization test, as well as suggest mechanisms for describing the processes by means of the geostatistical generalized linear model which provides a possible model for the data which also allows for covariates that could affect the insect distribution. The usage of such methods is new in the application context and they should be considered for the detection and description of the spatial patterns of pests in field conditions.

Overall, the data analysis using Mantel´s test supports the conjecture of the presence of a spatial patterns, although not consistently detects for all dates which may be influenced by the high variability of the observations, with a possible effect of the imprecise recording of high values. Also, the effects of non-measured covariates may have generated heterogeneous conditions of sampling, hiding spatial patterns.

It is recommended that future sampling should be carried out including some pairs of observations with smaller spaces between them to allow a better description of the spatial patterns. This is especially relevant considering the limited mobility of the insect.

ACKNOWLEDGEMENTS

This work was partially supported by grants from Coordenadoria para o Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazilian science funding agencies.

REFERENCES

CHRISTENSEN, O.F.; RIBEIRO JR., geoRglm - a package for generalised linear spatial models. **R-News**, v.2, p.26-28, 2002.

DIGGLE, P.J.; TAWN, J.A.; MOYEED, R.A. Model-Based geostatistics. **Applied Statistics**, v.47, p.299-350, 1998.

DIGGLE, P.J.; RIBEIRO JR, P.J.; CHRISTENSEN, O.F. An introduction to Model-Based Geostatistics. In: MϕLLER, J. (Ed.). **Spatial statistics and computational methods**. Springer-Verlag, 2003. Chapter 2, p.43-82.

DIGGLE, P.J.; RIBEIRO JR., P.J. **Model-Based geostatistics**. New York: Springer,

2007. 228p.

GREGO, C.R.; VIEIRA, S.R.; LOURENÇÃO, A.L. Spatial distribution of *Pseudaletia sequax* franclemlont in triticale under no-till management, **Scientia Agricola**, v.63, p. 321-327, 2006.

GOOVAERTS, P. **Geostatistics for natural resources evaluation**. New York: Oxford University Press, 1997. 483p.

ISAAKS, E.H.; SRIVASTAVA, R.H. **Applied Geostatistics**. Oxford University Press, 1989. 561p.

MANLY, B.F.J. **Randomization, Bootstrap and Monte Carlo Methods in Biology**, Chapman & Hall, 2006. 460p.

MANTEL, B.F.J. The detection of disease clustering and a generalised regression approach. **Cancer Research**, v.27, p.209-220. 1967.

McCULLAGH, P.; NELDER, J.A. **Generalized linear model**s. 2nd ed. London: Chapman & Hall, 1989. 511p.

MONTAGNA, M.A. Distribuição espacial e amostragem seqüencial da mosca-branca Bemisia Tabaci Raça B (Homoptera: Aleyrodidae) no agroecossistema do melão. Ribeirão Preto: USP/RP, 2001, 112 p. Dissertação (Mestrado em Entomologia).

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2008. URL http://www.R-project.org.

RUIZ, R. Modelagem da Distribuição Espaço-Temporal da Broca do Café (Hypothenemus hampei Ferrari) em uma Cultura de Região Central Colombiana. Piracicaba: USP/ESALQ, 2002. 119p. Dissertação (Mestrado em Estatística e Experimentação Agronômica).

RUIZ, R.C.; DEMÉTRIO, C.G.B.; ASSUNÇÃO, R.M.; LEANDRO, R.A. Modelos hierárquicos Bayesianos para estudar a distribuição espacial da infestação da broca do café em nível local. **Revista Colombiana de Estadística**, v.26, p.1-24, 2003.

SATO, M.E. Avaliação do Dano e Controle do *Thrips tabaci* Lindeman, 1988, na Cultura da Cebola (*Allium cepaL.*). Piracicaba: USP/ESALQ, 1989. 93p. Dissertação (Mestrado em Estatística e Experimentação Agronômica).

SNÄLL, T.; RIBEIRO JR, P.J.; RYDIN, H. Spatial occurrence and colonisations in patch-tracking metapopulations: local conditions versus dispersal. **Oikos**, v.103, n.3, p.566-578, 2003.

WORKMAN, P.J.; MARTIN, N.A. Towards Integrated Pest Management of *Thrips Tabaci* in onions. **Plant Protection**, v.55, p.188-192, 2002.