Spatial patterns detection of onion thrips (Thrips tabaci) on onion fields

Denise N. Viola (LCE-ESALQ/USP – DES/IM-UFBa), Clarice G. B. Demetrio (LCE-ESALQ/USP), Paulo J. Ribeiro Júnior (LEG-UFPR), Bryan J. F. Manly (Western EcoSystems Technology), Odair A. Fernandes (FCAV-UNESP)

Summary

Onion 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. 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 was used to test for spatial correlation, and when detected a mixed spatial Poisson model with a geostatistical random component was fitted to the data, allowing for a characterization of the spatial pattern as well as the ~~obtention~~ production of prediction maps of susceptibility to levels of infestation throughout the area.

Key-words: Thrips, Onion, Spatial Patterns, Randomization tests, Geostatistics, Poisson distribution.

1. Introduction

Onion 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 ~~from~~ small scale farms ~~production~~, because it is typically ~~typical~~ grown on small and medium sized properties. It is an annual plant for bulb production, ~~and~~ 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 (Costa & Medeiros, 1950 and 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 (Costa & Medeiros, 1950).

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 ~~given~~ the pupa phase is ~~is~~restricted to the soil. The nymph has low mobility, while the adult, although winged, has restricted movement. The ~~total~~ development cycle varies typically from 14 to 30 days, changing to 10 and 11 days when the temperature is over 30° 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~~n~~ 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 [redacted] 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, Morre? 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 & Srisvartava, 1989; Goovaerts, 1997) have been used to describe insect spatial patterns as, for instance, in Greco, Vieira and Lourenção, (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 ~~count~~ data cannot have a covariance structure of the type assumed by traditional methods of geostatistical analysis, with a stationary spatial covariance structure ~~stationary~~ 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 statistic literature but have ~~still~~ had few ~~applications algorithms available as well as usage in~~ practical applications (e.g. Diggle et al., 1998).

This paper describes a study of the spatial distribution of onion thrips ~~is described~~, with data from surveys of four different properties with different infestation levels and planting methods. The Mantel randomization test (Manly, 2006), was used to test for the presence of spatial autocorrelation and when detected, modelled by a mixed spatial Poisson model with a random term given ~~givel~~ by a geostatistical component. This model allows the ~~for~~ characterizing of the spatial pattern as well as the production of ~~producing prediction~~ maps of levels of susceptibility to infestation in different areas.

The remainder of this paper is organized as follows. Section 2 describes the data to be used, Section 3 reviews the Mantel randomization test for the detection of spatial patterns, and Section 4 presents the Poisson geostatistical model. The results of analyses are presented and discussed in Section 5, and some final considerations are presented in Section 6.

2. 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, 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	Sample times	Number of samples
Fazenda São Paulo	isolated from other plantings	10/7, 24/7, 31/7, 7/8, 14/8, 21/8, 28/8, 04/9	100, 100, 100, 98 100, 100, 100, 100
Estância Bela Vista	Bulbs	11/7, 1/8, 8/8, 14/8, 9/9	100, 100, 84, 99, 99
Sítio Rosário	Seedlings	21/6, 29/6, 7/7, 14/7, 21/7, 28/7, 4/8, 11/8, 18/8, 25/8, 3/9	50, 50, 48, 50, 50, 50, 50, 50, 50, 50, 50
Sítio Novo II	Seedlings	4/6, 19/6, 27/6, 28/6, 4/7, 11/7, 24/7, 31/7, 7/8	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. Figure 1 shows the position of the stakes in the four studied farms, in general with a 10x10m grid, but with some variations at the São Paulo farm. 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. From Table 1 ~~we can see~~ it can be seen that the number of samples and sampling times varied from farm to farm. 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.

much too small letters

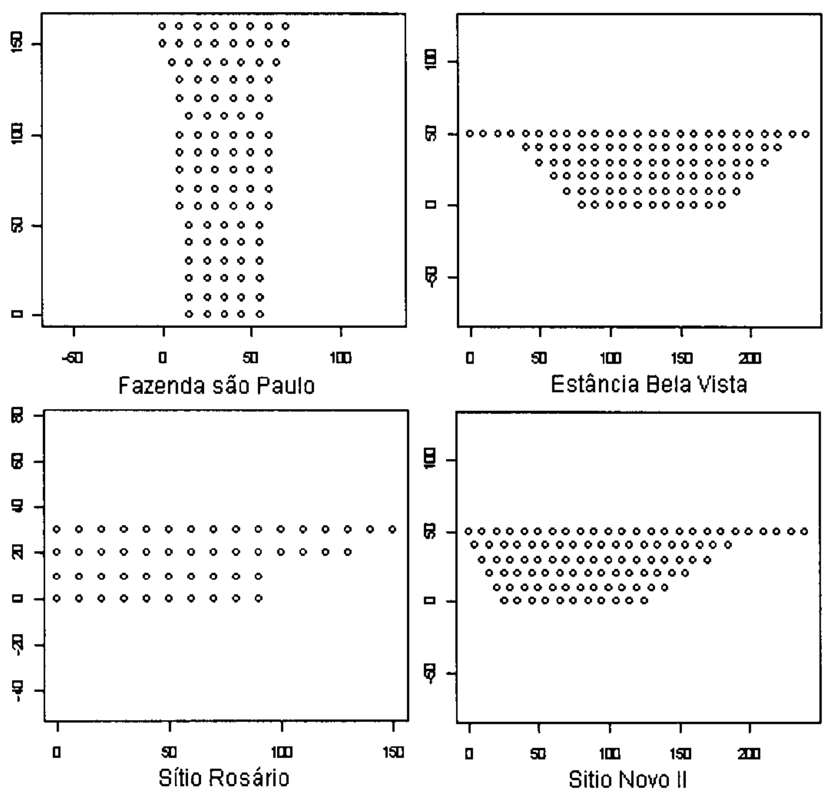

Figure 1: Localization of the stakes in each farm.

The 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. 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.

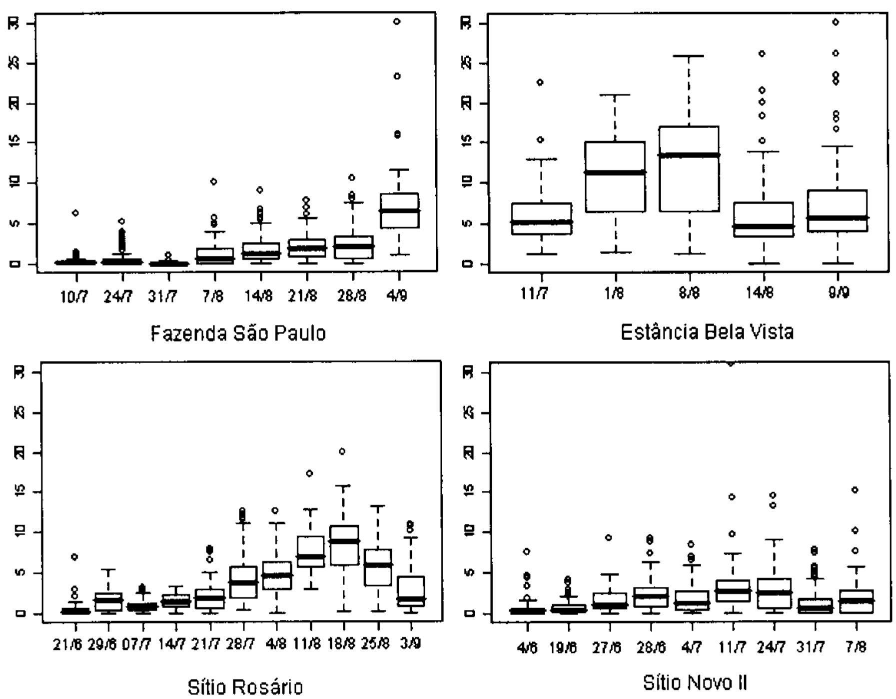

Figure 2: Box-plot onion trips per leaf.

At the São Paulo farm, on 31/07 there was the lowest average number of insects per leaf and also the lowest variance, with one insect per leaf as the maximum value. In contrast on 04/09 this farm had a much larger average number of insects per leaf and much greater variability. This farm had 35% to 100% of infested plants.

For the Estância Bela Vista, the lowest average number of insects per leaf occurred on 11/07 and 14/08. This farm had from 89% to 100%, of plants infested.

The Rosário farm had the highest average number of insects per leaf on 11/08 and 18/08. This farm only had 50 plants samples.

The Sítio Novo ll farm had the least average for the number of insects per leaf with low variability except for one very high count of 30.

3. 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 has 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 or equal 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 ~~don't~~ 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. ~~symmetric,~~ The ~~which~~ elements represent distances ~~, in some metric,~~ between the observations. These matrices can be denoted as

$$A = \begin{pmatrix} 0 & a_{21} & \dots & a_{n1} \\ a_{21} & 0 & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & b_{21} & \dots & b_{n1} \\ b_{21} & 0 & \dots & b_{n2} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & 0 \end{pmatrix}.$$

The matrix A is the Euclidian distances between the stakes with locations given by (x_{1i}, x_{2i}) and (x_{1j}, x_{2j}), i.e., with elements of the form $a_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}$ 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.,

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

which produces the r_o 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 $r_{ak} > r_o$, is then compared with a pre-fixed level of significance value 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.* ~~This~~ The test assumes that the correlation ~~in the adopted metric is linear and the alternatives include the proposal of Besag-Diggle,~~ Edginton, Monte Carlo, amongst ~~others (Manly, 2006)~~. An~~other proposal~~ alternative is given by Snäll, Ribeiro Jr. and Rydin (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,

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

where β_0 e β_1 are parameters to be estimated and ε_{ij} is the error associated with the response. However, more complex forms of spatial dependence can also occur.

In this work, 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 ~~verified~~ 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.

4. Modeling the spatial pattern

Having detected a spatial pattern, it may be of interest to describe the pattern by ~~the~~ 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 ~~values~~ 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, ~~maximum levels,~~ among others posible 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 two dimension ~~the bidimensional space~~ and y_i is the measure of interest at the x_i position of the ith observation. The response variable can be potentially measured at any point within the studied region (Diggle & Ribeiro Jr., 2007).

This will not make much sense to the reader. It should be explained more simply for an agricultural journal

The geostatistical model is specified by $[S,Y]=[S][Y \mid S]$, in which $Y(x): x \in A$ is a measure process within the study region A and y_i are the observed data; $S=\{S(x): x \in R^2\}$ is a Gaussian process with mean μ, variance σ^2 and correlation function $\phi(u)$, where u is the distance between pairs of observations. The values of the signal $S(x)$ are usually not directly observed, (Diggle et al., 1998; Diggle & Ribeiro Jr., 2007). The y_i observations may be seen as a *noisy* version of $S(x_i)$ ~~the~~ at the location x_i. Usually, a sample design assumes the locations x_i are either fixed or stochastically independent of the process that generates the y_i measurements. 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 (Diggle & Ribeiro Jr., 2007) and, for a finite set of plots, the random vector Y follows a multivariate Gaussian distribution. More generally, Y may follow other distributions. 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, $i=1, 2, \ldots, n$ are conditionally independent given $S(x)$, with expected values given

A lot of the words used here need to be defined

by $E[Y_i \mid S(x)] = \lambda_i$ and the predictor linear $h(\lambda_i) = S(x_i)$, i=1, 2, …, n with a known link function $h(.)$, where $\{h^{-1}(S(x)) : x \in A\}$ is the signal process and, in the presence of covariance, its extended to $S(x_i) = S(x) + d(x)^T \beta$, in which 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

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

The likelihood function of this model does not have a closed form, being described by the integral

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

with dimension equal to the number of observations, which cannot be solved by analytical or numerical methods. A possible solution for inference is to use Monte Carlo Markov Chain (MCMC) methods. A computional implementation is available through the package **geoRglm** for the **R** statistical environment (http://www.R-project.org).

For discrete random variables, the variogram is ~~a less~~ 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 ~~which can be explained by the presence o~~f 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, ~~or any other quantity quantity of interest~~. Supposing that the parameters are known and that the interest is in the expected insects ~~intensity~~ numbers 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 $Y|S$ conditional distribution, it is possible to simulate the conditional distribution of $[S|y]$, using the MCMC method. The predicted surface is given ~~by~~ (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 predictor of ? krigagem and $Var(x)$ is the prediction variance.

---

5. Results and discussion

5.1 Spatial pattern detection through ~~the~~ Mantel's randomization test

Through ~~the~~ the Mantel's test ~~it were obtained~~ p values for each sampling date, were obtained. For ~~considering~~ the Fazenda São Paulo, these were 0.0228, 0.0022, 0.0235, 0.0588, 0.1005, 0.0749, 0.1297 and 0.5540, suggesting ~~therefore detecting~~ a spatial pattern for the number of insects

~~by~~ per leaf for the 10th, 24th and 31st ~~of~~ July. ~~Such~~ There patterns can be observed in the dispersion plots (Figure 3) in which different symbols indicate the data distribution quartiles. In general, considering all the farms and dates, the distribution of the mean number of insects ~~by~~ per leaf is asymmetric and, ~~apparently,~~ does not ~~present~~ show trends ~~with the region coordinates~~ in space (?). 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 X which shows, as well, analogous results for the dates that detected spatial pattern at the other farms.

These graphs are far too small

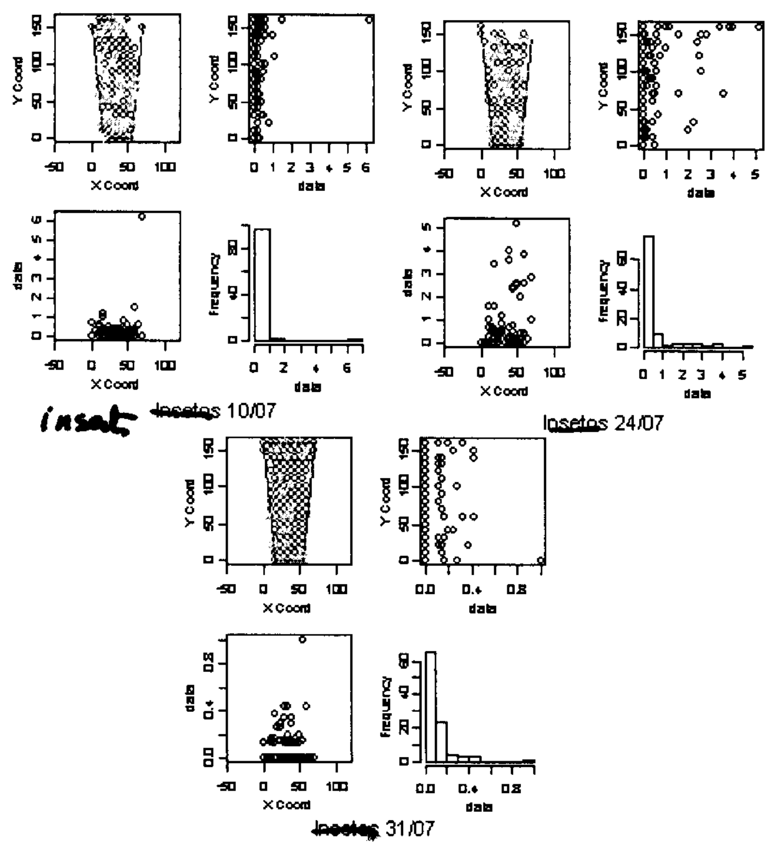

Figure 3 – Dispersion graphs and histograms for the mean number of insects at the Fazenda São Paulo.

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

loc = ?

insects/leaf

Propriedade	Data	Modelo	P
Fazenda São Paulo	10/07	Insetos/folha=0,2102+0,002325loc	0,0205
	24/07	Insetos/folha=0,6024+0,004216loc	0,0022
Estância Bela Vista	31/07	Insetos/folha=0,0932+0,000417loc	0,0264
	08/08	Insetos/folha=6,2180+0,009037loc	0,0334
Sítio Novo II	04/06	Insetos/folha=0,3035+0,007206loc	0,0012
	27/06	Insetos/folha=1,1810+0,004034loc	0,0258
	04/07	Insetos/folha=1,5240+0,003371loc	0,0455

For the Estância Bela Vista, the p values, obtained by the Mantel's test, for each sampling date were 0.8986, 0.0902, 0.0338, 0.7880 and 0.6224 ~~therefore detecting~~ suggesting a spatial pattern for the 8th ~~of~~ August. Data summary plots for this date are shown in the Figure 2. Considering the Sítio Rosário, the p values obtained, using the Mantel's test, were 0.5309, 0.5961, 0.9236, 0.9512, 0.9412, 0.7297, 0.3223, 0.1897, 0.2771, 0.1771 and 0.7020 ~~therefore, no~~ showing no evidence ~~of~~ spatial patterns ~~was observed~~. Finally, considering the Sítio Novo II, the p values obtained by the Mantel's test were 0.0006, 0.6127, 0.0251, 0.7326, 0.0478, 0.4781, 0.0651, 0.6084 and 0.4264, ~~therefore detecting~~ suggesting spatial patterns for the 4th and 27th ~~of~~ June and for ~~the~~ 4th ~~of~~ July, ~~whose~~ with data ~~are represented~~ shown in ~~the~~ Figure 5.

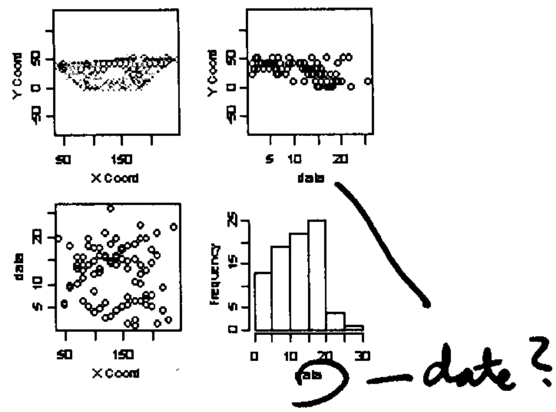

Figure 4 – Dispersion graphs and histograms

for the mean number of insects – Estância Bela Vista.

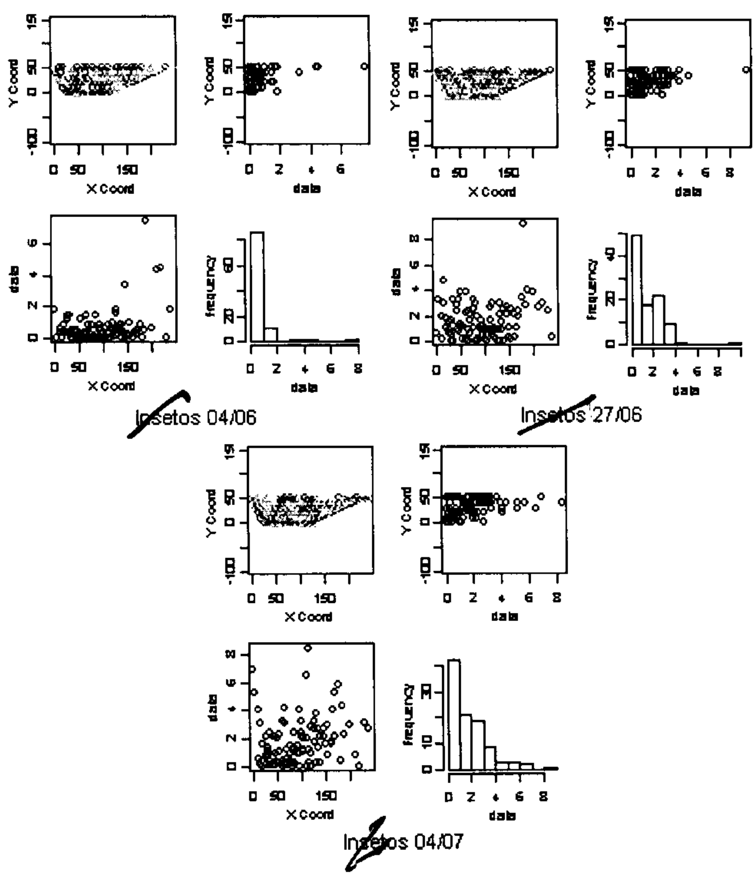

Figure 5 – Dispersion graphs and histograms for the

mean number of insects – Sítio Novo II.

Geostatistic generalized linear models with Poisson distribution s and logarithmic link function s were used ~~on~~ for the ~~spatial~~ modeling for the data ~~of~~ for farms and dates that showed some evidence of spatial pattern. Parameter confidence intervals were obtained using maximum variance estimates obtained by the MCMC algorithm. A total of 50,000 iterations chains were obtained, with a burn in cycle of 10,000, ~~and~~ keeping the first ~~at~~ of every 20 generated samples, amounting to a total of 2,500 samples. This process was repeated 1,000 times and, through the 1,000 obtained estimates, confidence intervals for the 2.5% and 97.5% quantiles were obtained. The results are summarized in the Table X. The obtained chain for each parameter was analyzed to verify the MCMC algorithm convergence. The confidence intervals for the ϕ parameter, which reflects the spatial dependence ~~extension~~, ~~presented~~ showed a large range, reflecting the difficulty in estimating this parameter with precision using scattered data. ~~Considering~~ For the case of an exponential correlation ~~function~~ model the practical range of correlation corresponds to three times the parameter value. This ~~parameter value~~ interpretation depends on the ~~interpretation of the~~ distances ~~amongst~~ between points ~~values~~, 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. Also, ~~it is observed that~~ there are three cases in which the estimate is smaller then the minimum distance.

? what are they?

? how scattered?

Table 3: ~~Pontual~~ Point estimates and confidence intervals for the parameters to geoestatistic model.

Prorpiedade	Data	β	σ^2	φ	τ^2	LogMV
Fazenda	10/07	-0,6302	0,2763	3,7852	0,0000	231,2

		(-0,57; -0,46)	(0,03; 0,34)	(0,57; 36,86)	(0,00; 7,04)	
	24/07	-0,5123 (-0,81; -0,62)	0,3773 (0,16; 1,01)	13,1430 (5,67; 82,86)	0,5781 (0,00; 5,40)	186,5
	31/07	-0,9424 (-0,97; -0,88)	0,2550 (0,04; 0,30)	1,0000 (0,56; 16,04)	0,0000 (0,00; 5,31)	320,6
Estância Bela Vista	08/08	2,3522 (2,34; 2,38)	0,1994 (0,18; 0,28)	14,2512 (10,72; 21,19)	0,9220 (0,39; 1,28)	127,5
Sítio Novo II	04/06	-0,4212 (-0,45; -0,15)	0,5535 (0,07; 0,58)	1,0002 (0,57; 110,18)	0,0000 (0,00; 5,94)	160,7
	27/06	0,2341 (0,14; 0,25)	0,2233 (0,14; 0,73)	23,6466 (5,60; 40,50)	2,0880 (0,00; 4,28)	55,53
	04/07	0,2697 (0,20; 0,30)	0,2660 (0,22; 0,46)	23,5905 (13,68; 37,71)	2,3366 (0,90; 3,08)	51,63

It ~~is observed that~~ The interval range's for β and σ^2 are small. Also, β is the parameter associated with ~~to~~ the link function, and σ^2, ϕ and τ^2 are parameters associated ~~to~~ with the surface S(x). The negative values for the 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 ~~reflex~~ of the low mean for the number of insects per ~~by~~ leaf.

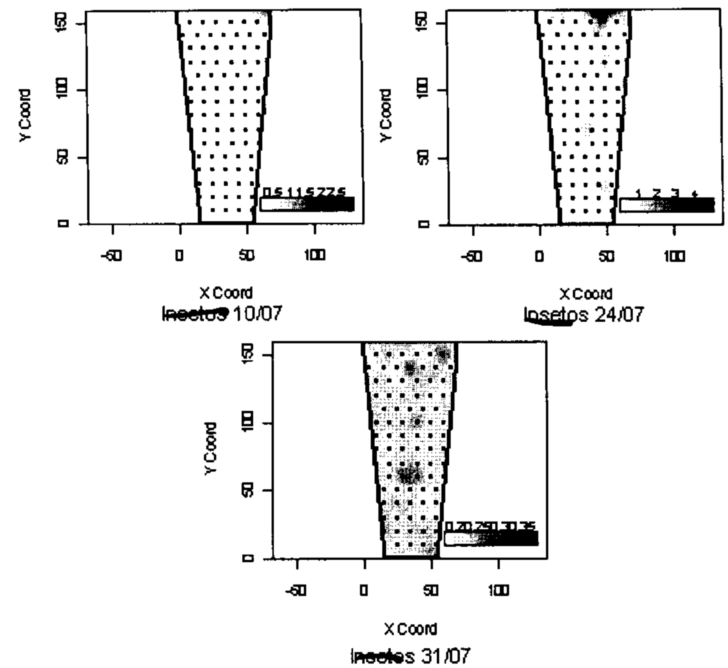

Figure 6: Prediction maps – Fazenda São Paulo.

From the fitted models prediction maps of the susceptibility ~~to levels o~~f infestation in the area were produced. Comparing the prediction maps showed in Figure 4, Figure 5 ane Figure 6 where the lighter colours indicate low infestation ~~while~~ 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 t~~hat the first details the~~ pattern ~~shown~~ in the second, as the low and high infestation areas are the same.

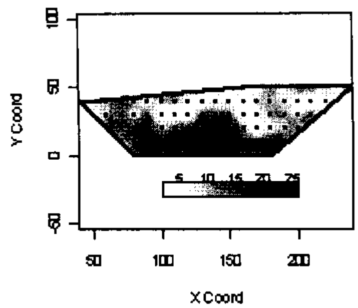

Figure 7: Prediction maps – Estância Bela Vista.

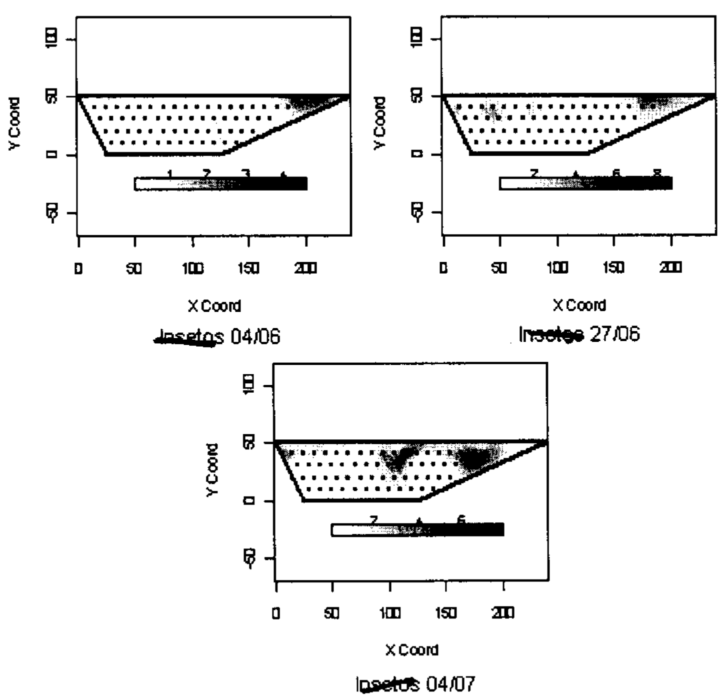

Figure 8: Prediction maps – Sítio Novo II

6. Conclusions

The methods presented above allowed the testing of spatial patterns in the distribution, of onion thrips, as well as suggesting mechanisms that can be used as possible descriptions of the processes. In the first case ~~the~~ Mantel's randomization test was very satisfactory for the ~~verification~~ detection of ~~the~~ random spatial patterns, while the geostatistic generalized linear model gave a possible model for the data, and a natural way to determine the ~~structure relations~~ relationships with covariates that could affect the insect distribution. The employment of these methods is new in the context of the application and the results obtained suggest that this ~~tool~~ method of analysis should be considered for the detection and description of the spatial patterns of pests in field conditions.

~~The~~ Spatial patterns, although detected for some dates, was not consistently indicated by the randomized tests. The e~~xplanator~~y analysis of the data indicated that the hypothesis of the presence of a spatial pattern is reasonable ~~for the phenomenon~~. However, its non-detection in certain cases may be attributed to the high variability of the observations, with a possible effect of the imprecise recording of high values. ~~Besides~~ Also, the effects of non~~-~~measured covariates may have generated heterogeneous conditions of sampling that ~~may have~~ hid~~den the~~ spatial patterns.

Apparently~~,~~ there is some influence of the kind of the plantation in the neighborhood on the number of insects per leaf on plants, because the Estância Bela Vista, which had in the neighborhood an area, already infested with thrips, was the one that had the highest means for the number of insects per leaf and the greatest proportions of infested plants. The Fazenda São Paulo, isolated from other plantations of onion, was the one that had the smallest proportion of infested plants, although this was increasing as time went by, indicating that there may have been some influence from the neighboring plants. This conjecture cannot be tested statistically with the available data, but it can be considered for future studies.

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