Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713428038

Spatio-temporal analysis of a plant disease in a non-uniform crop: a Monte Carlo approach

Bin Li[a]; R. S. Sanderlin[b]; Rebecca A. Melanson[c]; Qingzhao Yu[d]
[a] Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA, USA [b] LSU Agricultural Center Pecan Research-Extension Station, Shreveport, LA, USA [c] Department of Plant Pathology and Crop Physiology, Louisiana State University, Baton Rouge, LA, USA [d] School of Public Health, Louisiana State University, New Orleans, LA, USA

First published on: 16 August 2010

To cite this Article Li, Bin , Sanderlin, R. S. , Melanson, Rebecca A. and Yu, Qingzhao(2010) 'Spatio-temporal analysis of a plant disease in a non-uniform crop: a Monte Carlo approach', Journal of Applied Statistics,, First published on: 16 August 2010 (iFirst)
To link to this Article: DOI: 10.1080/02664760903301150
URL: http://dx.doi.org/10.1080/02664760903301150

Journal of Applied Statistics
2010, iFirst article

Taylor & Francis
Taylor & Francis Group

Spatio-temporal analysis of a plant disease in a non-uniform crop: a Monte Carlo approach

Bin Li[a]*, R.S. Sanderlin[b], Rebecca A. Melanson[c] and Qingzhao Yu[d]

[a]*Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, USA;* [b]*LSU Agricultural Center Pecan Research-Extension Station, P.O. Box 5519, Shreveport, LA 71135, USA;* [c]*Department of Plant Pathology and Crop Physiology, Louisiana State University, Baton Rouge, LA 70803, USA;* [d]*School of Public Health, Louisiana State University, New Orleans, LA 70112, USA*

(*Received 18 December 2008; final version received 1 September 2009*)

Identification of the type of disease pattern and spread in a field is critical in epidemiological investigations of plant diseases. For example, an aggregation pattern of infected plants suggests that, at the time of observation, the pathogen is spreading from a proximal source. Conversely, a random pattern suggests a lack of spread from a proximal source. Most of the existing methods of spatial pattern analysis work with only one variety of plant at each location and with uniform genetic disease susceptibility across the field. Pecan orchards, used in this study, and other orchard crops are usually composed of different varieties with different levels of susceptibility to disease. A new measure is suggested to characterize the spatio-temporal transmission patterns of disease; a Monte Carlo test procedure is proposed to test whether the transmission of disease is random or aggregated. In addition, we propose a mixed-transmission model, which allows us to quantify the degree of aggregation effect.

Keywords: hypothesis testing; lattice system; Monte Carlo; spatial; spatio-temporal analysis

1. Introduction

Identification of the type of disease pattern and spread in a field is critical in epidemiological investigations [3]. Information about spatial attributes of plant pathogens and disease in plant populations provides insight into disease progress and the determinants of disease spread. For example, a random pattern of infected plants suggests that, at the time of observation, pathogen movement is not limited to plants near already infected plants. Conversely, aggregations (clusters) of infected plants suggest that the pathogen is spreading largely from infected plants to spatially close plants within a field.

*Corresponding author. Email: bli@lsu.edu

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760903301150
http://www.informaworld.com

Spatio-temporal analysis of disease spread in agricultural crops is, for the most part, performed in fields in which the variety of plant in question is consistent throughout the planting (monoculture). A consistency in variety, and thus, a consistency in disease susceptibility to a particular pathogen, allows for the use of point-pattern or geostatistical approaches for the analysis of disease spread within the field. In some crops, however, different varieties of a plant may be grown in the same planting intermixed within rows and across rows, as is common with certain orchard crops. This type of arrangement, for example, is typical of pecan (*Carya illinoinensis*) orchards.

The objective of this study was to develop a method that can be used to accurately determine the pattern of disease movement within a planting of mixed varieties with differential disease susceptibilities. The remainder of the article is organized as follows. Section 2 describes the pecan data set and the primary question in this study. Section 3 briefly describes the current methods on analysis of disease patterns and describes a major limitation which prevents us from using them. Section 4 presents a Monte Carlo test based on a new test statistic. Section 5 extends the method to a mixed-transmission model, which allows us to quantify the degree of aggregation effect. Section 6 concludes the article.

2. Data description

For this paper, the term *variety* is used to refer to pecan selections of a specific genetic type. Because pecan is a dichogamous, cross-pollinated tree in order to maintain the genetic character of a variety, it is necessary to asexually propagate (clone) pecan trees through grafting. Grafting involves vegetatively attaching tissue in the form of a scion, selected from the desired variety, onto a pecan root source (rootstock). The tree that develops from the scion will thus be of the same genetic type (variety) as the scion. Because they are clonally produced, trees grown by grafting will produce nuts that are identical to the source of the scion, which is a commercially valuable trait.

The orchard where the data were collected was planted in 1987. It was designed to include 20 variety/rootstock combinations (three varieties and five rootstocks) in 16 (4 by 4) blocks. Each block had 20 trees (4 by 5) with one in each variety/rootstock combination and one non-grafted tree of each rootstock. Each of the three pecan varieties varied in susceptibility to *pecan bacterial leaf scorch* (PBLS) disease [14]. The trees were spaced 9.14 m within rows and 12.19 m across rows. The data were collected from 2004 to 2006. Infection of each tree was evaluated visually for disease symptoms. When symptoms were identified, infection was verified by serological assay for the pathogen. Only trees that gave a positive serological test for the plant pathogen were recorded as infected. For this study, one variety of high disease susceptibility was evaluated and is hence referred to as susceptible variety "A" (SVA). Even though spacing within rows and across rows was constant, the distance between trees of SVA was variable because of the intermixing of the two other varieties, plus the non-grafted rootstock trees, and the randomized complete block design of the orchard. Because rootstock type did not have any effect on disease susceptibility of the varieties grafted to them, this paper considers all SVA trees in the orchard regardless of rootstock.

PBLS is caused by the plant pathogenic bacterium *Xylella fastidiosa* [15], which causes a wide range of economically important plant diseases, including diseases of grape, peach, plum, almond, pecan, oleander, oak, maple, elm and coffee [9]. PBLS causes defoliation in pecan throughout the summer and fall months. Severely infected trees can have about 60% more defoliation at the end of the year than non-infected trees, and a kernel weight loss of near 16% [15]. It is known that *X. fastidiosa* can be transmitted in two ways in pecan: (1) graft-transmission through scions and rootstock and (2) insect vectoring [16]. The pathogen transmission method involved in disease spread in the work discussed here was apparently through insect vectoring, although the specific species of insects involved were not identified. *Primarily, we would like to know if there is an*

indication of aggregation of infected SVA trees near the previously infected trees, or if disease spread in this mixed variety orchard largely fits a random pattern within the SVA trees.

3. A brief review of existing methods

Various methods of spatial pattern analysis are used to characterize the spatial position of diseased plants within fields. Most can be categorized as either point-pattern or geostatistical approaches.

The point-pattern approach aims to quantify the pattern of diseased individuals within sampling units by describing each observation as a discrete 'point' in the landscape [17]. It includes doublet analysis [18,19], ordinary runs analysis [6,7], binomial dispersion index [10], Ripley's K function [13], 2-D distance class model [8], etc. In these methods, the null hypothesis is usually that the infected plants are randomly distributed. A key assumption under the null hypothesis is that each observation should have the same probability of being infected.

A major limitation of point-pattern methods is that they fail to recognize the degree of dependency among neighboring observations. On the other hand, geostatistical methods take into consideration both the random and systematic characteristics of spatially distributed variables and quantify spatial dependence by measuring the variation among samples separated by the same distance. Among these are the semi-variogram approach [4,5] and the spatio-temporal autocorrelation analysis [11,12], etc. One of the fundamental assumptions of geo-statistical methods is that the mean measure of interest is constant with respect to location, such as the probability of being infected.

The methods listed above are concerned with only one variety of plant at each location. Thus, there is no variation in disease susceptibility at a given location. The pecan orchard used in this study, however, is composed of different varieties with different levels of susceptibility to the disease.

The auto-logistic models proposed by [1,2] can be applied to test the null hypothesis that the infected trees are randomly distributed with more than one variety of plants at each location and with non-uniform disease susceptibility across the field.

4. A Monte Carlo test for randomness

Based on the primary question posed in Section 2, we consider the following null and alternative hypotheses.

Null hypothesis: Infected SVA trees are randomly distributed.

Alternative hypothesis: Infected SVA trees exhibit an aggregation pattern.

First, we propose a new test statistic that characterizes the spatial pattern of PBLS. Then, a Monte Carlo test procedure is described and applied to the data set together with the test result and conclusions. Finally, we examine the properties of the test through a simulation study with a known underlying data generation mechanism.

4.1 *Test statistic*

A *link* is defined as two adjacent diseased trees in any direction. It starts from a previously infected tree (of any variety), and ends at a newly infected SVA tree. As an illustration, the distribution of PBLS in 2005 is shown in Figure 1. The links created in 2005 are shown as arrows. For example, there are two links that start from a previously infected non-SVA (2, 12) and end at two newly infected SVAs (2, 11; 3, 12). The empty spots in Figure 1 are the trees which died before 2005.

Let $T_s(P, N)$ be the number of links generated in year s with two arguments: P is the index set of previously infected trees, and N is the index set of newly infected SVAs in year s. For example in 2005, $T_{2005}(P, N) = 9$ (Figure 1). Notice that P includes all the previously infected non-SVAs

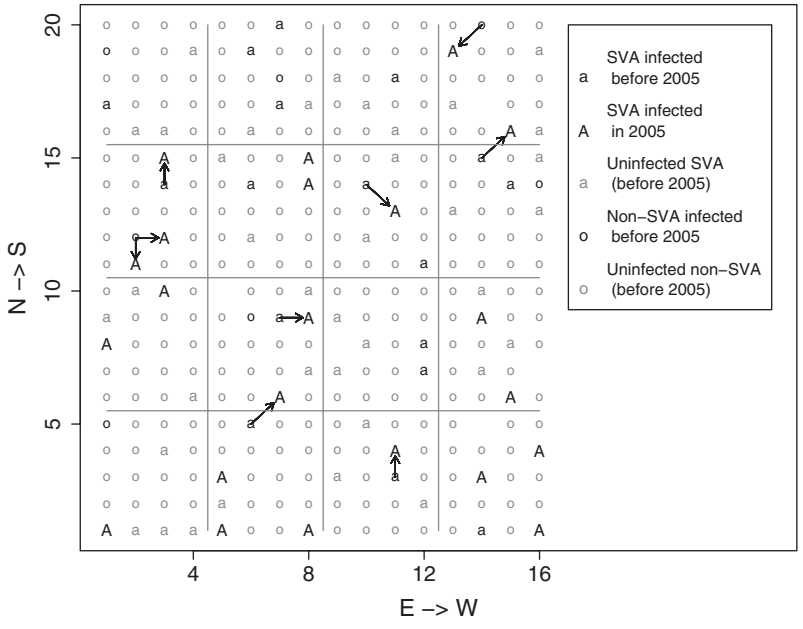

Figure 1. PBLS distribution plot in 2005. Black lower (upper) case 'a' ('A') represents SVA infected before (in) 2005. Grey lower case 'a' is SVA not infected in 2005. Black (grey) 'o' represents non-SVA (not) infected before 2005. Each arrow represents a link generated in 2005.

(black 'o's in Figure 1) and SVAs (black and lower case 'a's), where N includes all the SVAs infected in 2005 (black and upper case 'A's).

The total number of links generated from 2004 to 2006, i.e. $T^* = T_{2004} + T_{2005} + T_{2006}$, is used as the test statistic for the test of randomness (against aggregation). Note that the value of T_s will be a large number if there is an aggregation pattern of infected SVAs.

4.2 *Test procedure*

The data consist of different varieties of trees with different susceptibility to PBLS. Therefore, the exact distribution of the test statistic T^* under the null hypothesis is difficult to determine. Hence, we propose to use a Monte Carlo technique to generate a large number of artificial data sets under the null hypothesis. For each generated data set, the corresponding T^* is calculated. The histogram of T^*s from all generated data sets can be used as an empirical approximation of the underlying distribution of the test statistic T^* under the null hypothesis. Let C be the index sets of all SVAs and D_s be the index set of non-SVAs infected in year s. Let P_0 be the index set of all trees (both SVAs and non-SVAs) infected before 2004. The details of the simulation procedure are as follows.

ALGORITHM 1 Monte Carlo procedure

(1) *Initialize the set* $P = P_0$.
(2) (*a*) *Randomly select eight non-infected SVAs* (*without replacement*) *from* C *to* P *as* N*;* (*b*) *calculate* $T_{2004}(P, N)$; (*c*) *update* $P \leftarrow P \cup D_{2004} \cup N$.
(3) (*a*) *Randomly select* 22 *non-infected SVAs* (*without replacement*) *from* C *to* P *as* N*;* (*b*) *calculate* $T_{2005}(P, N)$; (*c*) *update* $P \leftarrow P \cup D_{2005} \cup N$.

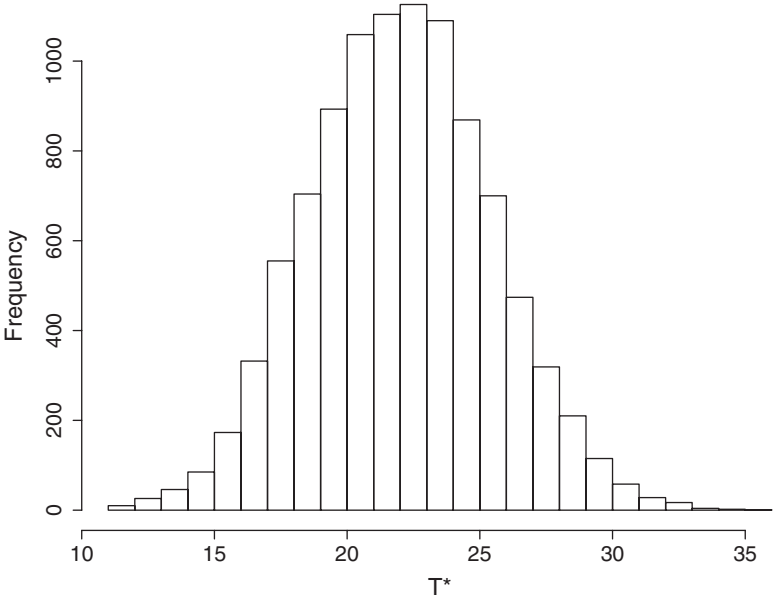

Figure 2. The empirical distribution (histogram) of T^* based on 10,000 runs of simulations.

(4) (*a*) *Randomly select* 10 *non-infected SVAs* (*without replacement*) *from C to P as N;* (*b*) *calculate* $T_{2006}(P, N)$.
(5) $T^* = T_{2004} + T_{2005} + T_{2006}$.

The minus operator is the (asymmetric) set difference operator, i.e. $X - Y = \{x \in X | x \notin X \cap Y\}$. Namely, $X - Y$ is the set of all elements from X but not in Y. For example, if $X = \{1, 2, 3\}$ and $Y = \{2\}$, then $X - Y = \{1, 3\}$. Notice that 8/22/10 are the observed numbers of SVAs infected in 2004/2005/2006, respectively. The empirical distribution of the test statistic T^*, based on 10,000 runs of simulations described above, is shown in Figure 2. The observed total number of links created in 2004–2006 from the original data set T^*_{obs} is equal to 26.

Note that in this study, we assume that non-randomness can only be caused by an excessive aggregation. Hence, the test is one-sided. The *p*-value, which is the proportion of T^*s that are greater than or equal to 26, is equal to 0.1928. If we fix the significance level at 0.05, we fail to reject the null hypothesis that the infected SVAs are randomly distributed. Based on the empirical distribution shown in Figure 2, the rejection region (at significance level 0.05) is $\{T^* \geq 29\}$. This is because that proportions of $T^* \geq 29$ and $T^* \geq 28$ are 0.0435 and 0.0754, respectively.

4.3 *Simulation studies*

The purpose of the simulation study is to investigate the type I error and the power of the proposed test under various situations. The simulated data are based on the 50 by 50 lattice system. Each point can be viewed as a tree (i.e. there are $50 \times 50 = 2500$ trees in the simulated orchard). Among all 2500 trees, 25% of them are SVA while the rest are non-SVA trees. The mixing of SVA and non-SVA trees is random. Like the pecan study, the simulated data have 3 years infection data after the initialization.

The initial infected SVAs and non-SVAs are randomly selected from the binomial process with $p = 0.09$ and 0.01, respectively. These probabilities are approximately equal to the proportion of infected SVA and non-SVA trees in the pecan study.

Table 1. Results for the simulation study.

	p_1^S	p_2^S	p_1^N	p_2^N	Proposed	Auto-logistic
Case 1	0.20	0.00	0.04	0.00	1.00	1.00
Case 2	0.20	0.05	0.04	0.01	1.00	1.00
Case 3	0.20	0.10	0.04	0.02	0.60	0.46
Case 4	0.20	0.15	0.04	0.03	0.16	0.12
Case 5	0.20	0.20	0.04	0.04	0.04	0.06

After initialization, the uninfected SVA trees adjacent to or not adjacent to infected trees have a probability of p_1^S or p_2^S of getting infected, respectively. For the uninfected non-SVA trees, trees adjacent to or not adjacent to infected trees have a probability of p_1^N or p_2^N of getting infected, respectively. Note that if $p_1^S = p_2^S$ and $p_1^N = p_2^N$, the disease transmission pattern is random, while if $p_2^S = 0$ and $p_2^N = 0$, we have an aggregated disease transmission pattern.

Five different transmission patterns (cases 1–5) are considered. For each case, 50 simulated data sets were generated. Table 1 shows the proportion of rejection of the null hypothesis at $\alpha = 0.05$, for the proposed method and the auto-logistic model [1,2] under various situations based on 50 replications. Note that in case 1, the disease can only be transmitted from adjacent previously infected trees, while in case 5 the disease is randomly transmitted. In cases 2–4, the disease transmission pattern is mixed while the disease transmission pattern becomes less aggregated and more random. We see that in both cases 1 and 2, the powers of both the proposed test and the auto-logistic model to detect the aggregation pattern are 100%. In case 5, the type I error of the proposed test is 0.04, which is close to the pre-defined 5% level. In cases 3 and 4, the power of the proposed test is 0.60 and 0.16, respectively, which is slightly higher than the auto-logistic approach.

5. Mixed-transmission model

Spread of PBLS may be random within an orchard or through proximal spread from point sources creating aggregates. Hence, it is useful to develop a mixed-transmission model, which is able to account for both aggregation and random disease patterns simultaneously. Let p_{agg} be the proportion of infected SVAs in an aggregation pattern. Then, to carry out the simulation based on the mixed-transmission model, we only need to modify the part a's in Steps 2–4 of the simulation procedure in Section 4.2 as follows. Given p_{agg}, a Bernoulli trial (with probability p_{agg} of being a 'success', otherwise a 'failure') is executed before selecting a non-infected SVA. If 'success', we randomly pick a non-infected SVA which is adjacent to a previously infected tree. Otherwise, we randomly pick a non-infected SVA (without the location constraint). Within a step, if a non-infected SVA is selected twice, then we will repeat the selection until a new non-infected SVA is selected.

Based on the mixed-transmission model, the null and alternative hypotheses in Section 4 can be rewritten as $H_0 : p_{agg} = 0$ versus $H_1 : p_{agg} = 1$. Since p_{agg} can be any real number between 0 and 1, the simple alternative hypothesis is changed to a composite hypothesis, i.e. $H_1 : p_{agg} > 0$. Note that for a fixed significance level, for instance $\alpha = 0.05$, the conclusion of the test (with composite H_1) is the same as the one with the simple H_1, because the null hypothesis does not change. However, by using the mixed-transmission model, we are able to (1) quantify the strength of the aggregation effect or local dependence where other point-pattern methods fail to estimate p_{agg}; and (2) visualize the gradual change of power as p_{agg} changes.

We simulate the mixed-transmission model at p_{agg} ranges from 0 to 1 with a step length 0.05 ($p_{agg} \in \{0, 0.05, \ldots, 0.95, 1\}$). For each value of p_{agg}, 10,000 runs of simulations were carried out. The power is the proportion of T^*s that are greater than or equal to 29 (the rejection region is

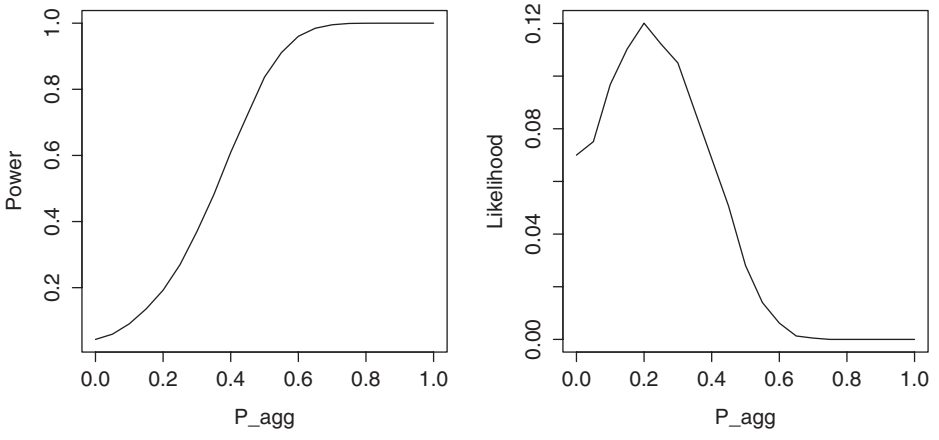

Figure 3. The power curve (left) and likelihood function (right).

based on the empirical distribution of T^* at $p_{\text{agg}} = 0$). To estimate p_{agg}, the maximum-likelihood estimate (MLE) is used. The likelihood function is defined as the conditional probability of having 26 links ($T^*_{\text{obs}} = 26$) given the value of p_{agg}. The left panel of Figure 3 shows the power curve that is monotonically increasing. We see that the power is over 0.8 (0.99) when p_{agg} is above 0.5 (0.7). The right panel of Figure 3 shows the likelihood function. We see that the likelihood function reaches its global maximum when p_{agg} is 0.2, i.e. the MLE of p_{agg}, with approximately 12% of the T^*s being equal to 26.

Sensitivity analysis is the study of how model output varies with changes in model inputs. An important parameter in the study is the number of runs in the Monte Carlo simulation. To check whether 10,000 runs is sufficient, we repeat the experiments three times. Figure 4 shows the empirical distributions of T^* under the null hypothesis (left) and the likelihood functions (right) in the three additional (grey) and original (black) experiments. We see that the empirical distributions are close to each other in all four experiments. The *p*-values in the three repeated experiments are 0.1952, 0.1790 and 0.1846. Similarly, the likelihood functions are also clustered

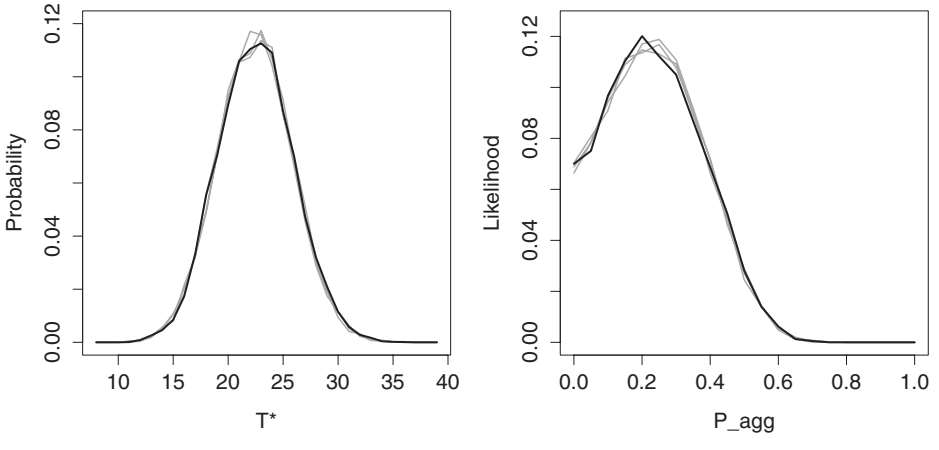

Figure 4. The empirical distributions of T^* under the null hypothesis (left) and the likelihood functions (right) in three repeated (grey) and original (black) experiments.

together. However, we have a split vote in MLE (two are 0.20 and two are 0.25). This is probably due to the coarse grid on p_{agg}; and the underlying p_{agg} may fall between 0.20 and 0.25.

6. Conclusion

Characterization of the spatial pattern of diseased plants can facilitate the determination of the influence of biological and environmental factors on disease dispersal processes. In this paper, we proposed a Monte Carlo method that allows us to (1) test for randomness of infected plants against aggregation patterns; and (2) quantify the degree of dependence among neighboring plants. Unlike most of the existing methods, it can be applied to data with mixed varieties with different disease susceptibility, non-uniform plant spacing and even missing values. In this study, since we are only interested in testing whether the PBLS disease transmission pattern among SVA is random or aggregated, we did not jointly model and test all the varieties. However, the proposed test procedure can be directly extended to jointly test both SVA and non-SVA trees. The proposed method potentially can be applied to other studies in a broad range of epidemiology and sociology problems. Sometimes, we are interested in the disease pattern in one or more particular directions, e.g. diagonal direction. Then, in our approach, we only need to modify the definition of *link*. Namely, the *link* can only end at adjacent newly infected SVAs in the diagonal direction of a previously infected tree.

References

[1] J.E. Besag, *Nearest-neighbour systems and the auto-logistic model for binary data*, J. R. Stat. Soc. Ser. B 34 (1972), pp. 75–83.
[2] J.E. Besag, *Spatial interaction and the statistical analysis of lattice systems (with discussion)*, J. R. Stat. Soc. Ser. B 36 (1974), pp. 192–236.
[3] C.L. Campbell and L.V. Madden, *Introduction to Plant Disease Epidemiology*, John Wiley and Sons, New York, 1990.
[4] N. Cressie, *Fitting variogram models by weighted least squares*, Math. Geol. 17 (1985), pp. 563–586.
[5] N. Cressie, *Spatial prediction and ordinary kriging*, Math. Geol. 20 (1988), pp. 405–421.
[6] J.D. Gibbons, *Nonparametric Statistical Inference*, McGraw-Hill, New York, 1971.
[7] J.D. Gibbons, *Nonparametric Methods for Quantitative Analysis*, Holt, Rinehart, and Winston, New York, 1976.
[8] S.M. Gray, J W Moyer, and P. Bloomfield, *Two-dimensional distance class model for quantitative description of virus-infected plant distribution lattices*, Phytopathology 76 (1986), pp. 243–248.
[9] D.C. Hopkins and A.H. Purcell, *Xylella fastidiosa: Cause of Pierce's disease of grapevine and other emergent diseases*, Plant Dis. 86 (2002), pp. 1056–1066.
[10] L.V. Madden and G. Hughes, *Plant disease incidence: Distributions, heterogeneity, and temporal analysis*, Annu. Rev. Phytopathol. 33 (1995), pp. 529–564.
[11] L.V. Madden, T.P. Pirone, and B. Raccah, *Analysis of spatial patterns of virus-diseased tobacco plants*, Phytopathology 77 (1987), pp. 1409–1417.
[12] K.M. Reynolds and L.V. Madden, *Analysis of epidemics using spatio-temporal autocorrelation*, Phytopathology 78 (1988), pp. 240–246.
[13] B.D. Ripley, *Spatial Statistics*, John Wiley & Sons, New York, 1981.
[14] R.S. Sanderlin, *Cultivar and seedling susceptibility to pecan bacterial leaf scorch caused by Xylella fastidiosa and graft transmission of the pathogen*, Plant Dis. 89 (2005), pp. 446–449.
[15] R.S. Sanderlin and K.I. Heyderich-Alger, *Evidence that Xylella fastidiosa can cause leaf scorch disease of pecan*, Plant Dis. 84 (2000), pp. 1282–1286.
[16] R.S. Sanderlin and R.A. Melanson, *Transmission of Xylella fastidiosa through pecan rootstock*, HortScience 41 (2006), pp. 1455–1456.
[17] W.W. Turechek and L.V. Madden, *Spatial pattern analysis of strawberry leaf blight in perennial production systems*, Phytopathology 89 (1999), pp. 421–433.
[18] J.E. Vanderplank, *A method for estimating the number of random groups of adjacent diseased plants in a homogeneous field*, Trans. R. Soc. South Afr. 31 (1946), pp. 269–278.
[19] J.E. Vanderplank, *Analysis of epidemics*, in *Plant Pathology*, Vol. 3, J.G. Horsfall and A.E. Dimond, eds., Academic Press, New York, 1960, pp. 229–289.