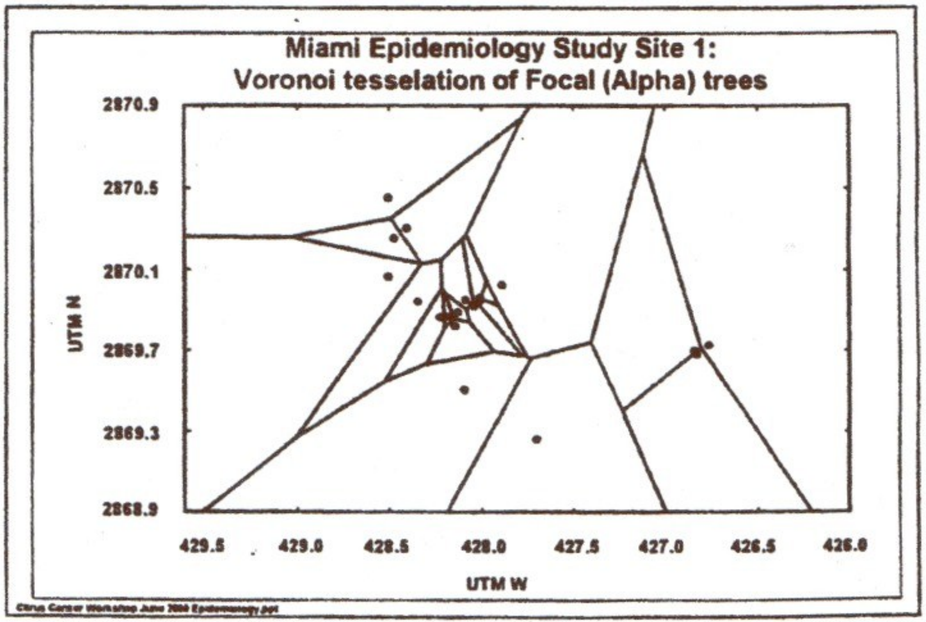
APPENDIX F

SPATIAL POINT PATTERN ANALYSIS



1. INTRODUCTION

In the 2002 published article, the authors claim to introduce an extension to Ripley-K analysis of a spatial point pattern (SPP) which does not require special consideration of edge effects. The intent of the analyses was to identify distances where the incidences of citrus canker were spatially dependent. This method depends of the DNC method, step 2, the parsing of infected trees into time periods, as discussed in Appendix A.

The review concluded the method was not an extension of the Ripley-K method, but rather consistent with inter-point distance (IPD) analysis as discussed in Appendix D. This appendix is further theoretical extension and application of the IPD analysis as reviewed in Appendix D.

The application spatial point patterns to plant disease epidemiology is discussed in Madden, al., 2006 and Campbell, 1990. An important difference between the applications in these references and in this review- all applications in these textbooks concern commercial crops grown on farms or orchards and the Florida field study was exclusively conducted in urban residential areas of Florida. As will be discussed, the underlying arrangement of host plants can not be overlooked in a spatial point pattern analysis.

A short summary on spatial point pattern (SPP) is presented here in terms of concepts, terminology and sources of information. The focus is on general concepts rather than equations and rigorous mathematics, but some mathematics is included. Also, the scope was limited to methods given in the Gottwald, 2002. For example, sampling methodologies are not reviewed here because in the Florida field study, all sites were repeatedly surveyed.

Ideally, the appendices should be read in order following Chapter 5. The most relevant Appendices for this discussion would be Appendices A, B and D.

TERMINOLOGY USED IN THIS APPENDIX

The collection of points in a defined area is called a “pattern” which is synonymous with arrangement. The arrangement of points may be regular, random or clustered. Example of regular patterns would be the arrangement of trees in an orchard. Since the Florida field study did not include orchards, spatial methods for regular patterns need not be reviewed here. Random points may be intuitively thought of as evenly scattered, or not displaying any aggregation. A clustered set of points displays areas of higher concentration of points (intensity) in certain locations and lower in others.

Random and cluster points patterns can be generated by pseudo random number generation synonymous Monte Carlo simulation or Monte Carlo modeling. An excellent summary is available from Wikipedia as cited in the references. To keep our terminology unambiguous, the word “distribution” has purposely been avoided in relation to the arrangement of points, and used only in relationship of a probability distribution.

A cumulative distribution function (c*df)* or *F(x)* provides the cumulative probability of values equal to or less than *x.* A *cdf* can be calculated using quantiles, which assigns probabilities by ranking the data from lowest to highest, and assigning a probability to each data value. These *cdf* are often referred to as empirical *cdf’s*. Alternatively, one can assume a random variable is represented by a certain theoretical *cdf,* such as a uniform distribution, normal (Gaussian), hypergeometric, Poisson and binomial; the latter three distributions are used with discrete random variables.

COMPLETE SPATIAL RANDOMNESS

Campbell, 1990 description or properties of a random pattern includes:

1. Every point on a surface has an equal probability of being occupied.
2. Providing the location of any one individual on a surface provides no information as to the location of other individuals.

Cressie provides an equivalent interpretation of a random pattern (page 586), which follows after being stated mathematically,

> Intuitively, this says [referring to equations involving probabilities] that events are equally likely to occur anywhere within A and that events do not interact with each other.

Complete spatial randomness (CSR) is a conceptual standard by which other patterns can be judged.

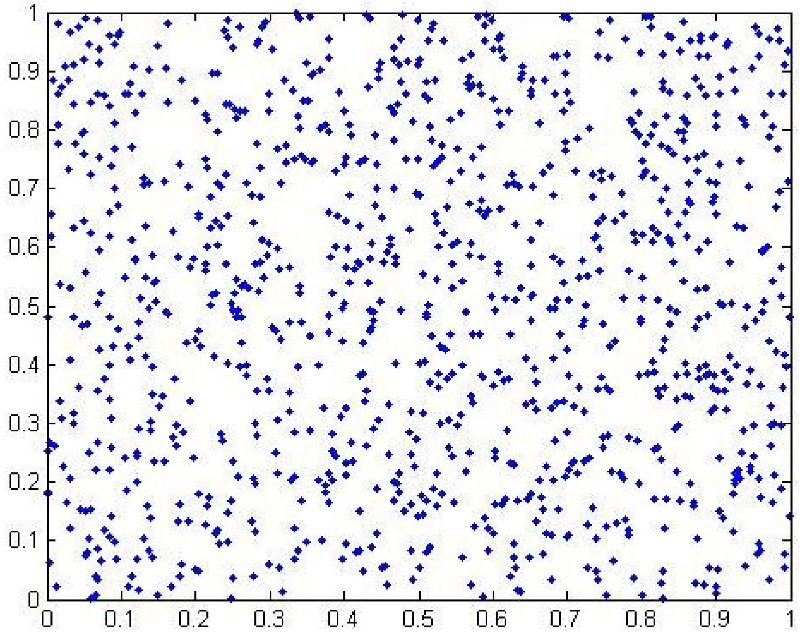
Random patterns can be defined mathematically. There are also interpretations and example which extend the conceptual understanding. According to Diggle, 2003, a hypothesis for a complete spatial randomness for a spatial point pattern assets that:

> (i) the number of events in any planar region with area |A| follows a Poisson distribution with mean $\lambda|A|$
>
> (ii) given n events x_i in a region A, the x_i are an independent random sample from the uniform distribution on *A.*

Random pattern of points are generated on the computer by calling random functions, which produce equal size vectors of X and Y random coordinates using random number generators. Diggle explain that intuitive ideas of what constitutes a 'random pattern' can be misleading. Using computer generated random patterns, the points in one realization may appear to be aggregated, but the pattern is in fact, random. A realization of a computer generated random pattern is in Figure 1.

.

Figure 1: Computer generated complete spatial randomness pattern



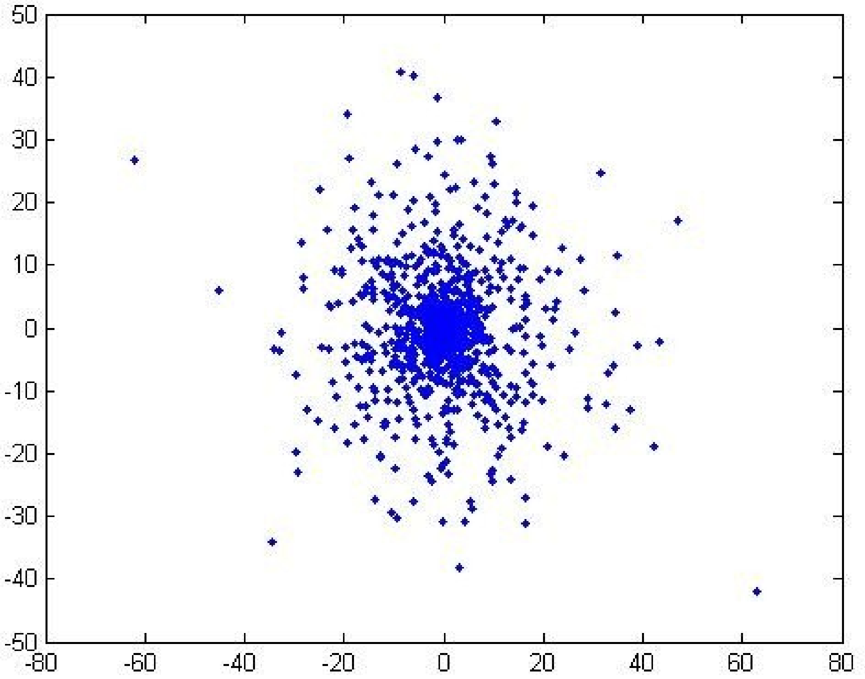
Matlab Code:

```
clear;
x= rand(1000,1)
y = rand(1000,1)
plot(x,y,'.')
```

AGGREGATION IN POINT PATTERNS

A computer generated aggregate pattern is shown below, showing a single clustering of points. There can be multiple clusters, each one showing a central concentration.

Figure 2: Clustered Pattern



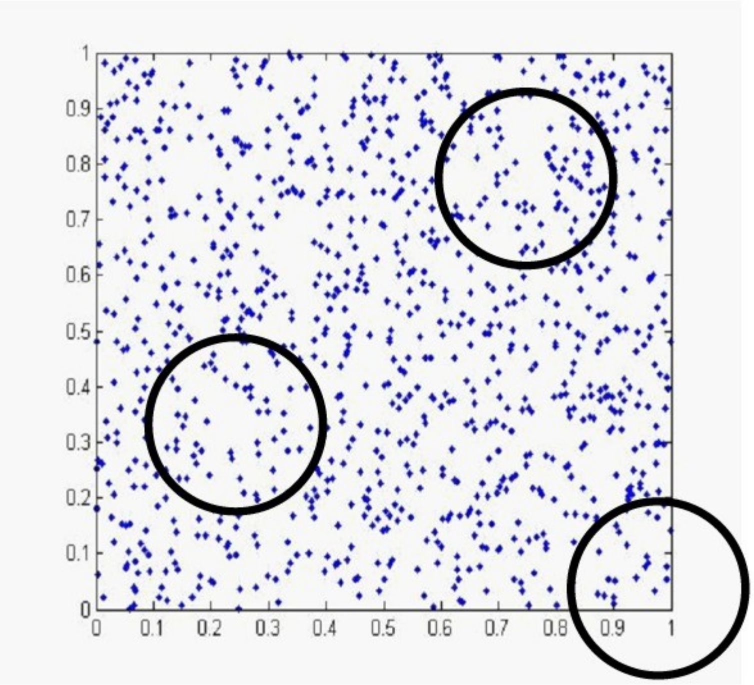
Visually, it is very easy to distinguish between this pattern from the random pattern shown in Figure 1. The clustering patterns can become much more complex. If several clusters are present, of varying degrees of aggregation, then attempting to interpret the pattern likely requires statistical analyses. Further, if the observation window is limited to only a part of the pattern, particularly if it is on the fringes of the cluster, the pattern may appear to be random. Finally, a pattern which originally was clustered may be thinned over time, so the appearance is more consistent with a random pattern.

RIPLEY-K ANALYSIS

It is stated at the onset that this discussion will be limited to the analyses within the paper. Ripley-K analyses is certainly discussed in the paper, but it is our contention that Ripley-K or some modified version of Ripley-K analysis was never done. But to explain this, it is necessary to describe the Ripley-K method. It is our contention that the method used in the spatial point pattern analysis is properly termed “inter-point analysis” or “inter-event analysis”, which is presented in the next section

Consider a pattern is sampled by a circle with an origin x,y, radius d, and area A ($= \pi d^2$). The the number of points within the circle can be determined and intensity calculated. Now, we can randomly select origins, and continue to count the points within the circle. The same can be repeated with different radii. The mean counts at distances can be compared with what should be the result had the points been an evenly scattering (CSR) pattern sampled as the number of points tends to infinity. Using Monte-Carlo methods, confidence limits can be determined for the CSR pattern with limited points.

Figure 4: Ripley-K Sampling



For our pattern, the overall intensity λ is 1000 since we have 1000 points distributed over a unit area. Let X represent the number of points within each circle. The number of points within circles completely within the is circle Let X represent the number of points within each circle.

The probability P{X = x} where x may be any value from 0 to 1000 points, as in our example, can be determined with the Poisson distribution, with mean λ A.

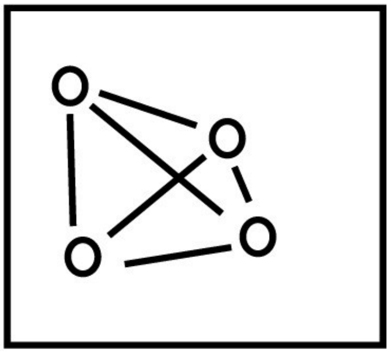
This will work well unless our circular sample extends over a boundary resulting in an incomplete sample. There are known “fixes” to this problem, one is to adjust upwards the number of counted points in our sample. The second is to employ “guard rails” so the sampler does not extend beyond the boundaries of our points. The article by Ward and Ferrandino (5) proposed a new upward adjustment for samples which were not entirely within the study area.

INTER-POINT DISTANCE (IPD) ANALYSIS

In inter-point distance analysis, every point is connected to every other point excluding all duplicate distances. The generated statistics and statistical analysis will be referred to as IP statistics and IPD statistical analysis. This is synonymous with inter event distances, as referred to by Diggle, 2003. In Appendix C, the method to generate distances was referred to as the inter-tree distance analysis, in review of Figure 3 of the Gottwald, 2002.

Inter-point distances are shown below. Given n points, the number of calculated distances is *(n)(n-1)/2.* An empirical cumulative distribution can be determined for random variable, X, representing the distance between points.

Figure 4: Inter-point distance analyses example with 4 points and 6 distances



The random variable, *X,* is the distance between randomly selected points. The cumulative distribution function is:

$F(d) = P\{X \leq d\} = p$ or the probability of an event is less than *p.*

The event is defined as a distance as calculated from two randomly selected point is less or equal to d.

For square study area, the cumulative distribution function (*cdf)* can be calculated using an Bartlett's equation (Diggle, 2003).

$$\begin{aligned} F(x) = {} & \pi x^2 - 8x^3/3 + x^4/2 && 0 \le x \le 1 \\ & 1/3 - 2x^2 - x^4/2 + 4(x^4 - 1)^{1/2}(2x^2 + 1)/3) && 1 < x \le \sqrt{2} \\ & \quad + 2x^2 sin^{-1}(2x^{-2} - 1) && \end{aligned}$$

An empirical *cdf* for a CSR pattern for any geometry can be generated using simulation. For hypothesis testing with limited data, appropriate CSR pattern envelope, typically 5 and 95% limits of the *cdf* can be determined using simulation as discussed in Diggle, 2003 on pages 14-17. ..

The primary field study article (Gottwald, 2002) does not specifically identify IPD method, but it is easily recognizable in the methodology section of the paper. The excerpt in the paper, as presented below, indicates that "direct enumeration of tree pairs" (entire sentence highlighted)

> An alternative method for calculating the *K*-function has recently been presented that directly incorporates the edge effect into CSR model predictions, yielding more accurate estimates of the variance (32). In what follows, this approach is extended to a sample space consisting of an irregular lattice of points (trees). For this case, the prediction of infected point pairs based on CSR is obtained by direct enumeration of tree pairs. Thus, no edge correction is necessary and expectation values and confidence limits can be obtained by direct statistical inference.

2. SELECTED EXCERPTS FROM THE PUBLISHED ARTICLE

Selected excepts from Gottwald, T.R., X. Sun, Riley, T. Graham, J.H., Ferrandino, F. and Taylor, E., 2002, Geo-Referenced Spatiotemporal Analysis of the Urban Citrus Canker Epidemic in Florida, Phytopathology, Vol 92, No. 4.

Every effort has been taken to transcribe the excerpted passages related to spatial point pattern exactly as published. Figures and table numbers used in this section are based on the published article. These selected sections may exclude important details, and it is recommended that the full article be reviewed. The full article may be downloaded free of charge from a number of websites, www.citruscankerdocs.com.

--- Page 362, left hand side, begins middle of third paragraph, beginning with the third paragraph, beginning with the sentence "Semivariance analysis... ."

More robust approaches to measures of spatial randomness of discrete (i.e., binary) spatial point pattern data are the stochastic modeling approaches proposed by Ripley (5,26) and have found application predominantly in ecological studies. For Ripley's methods, first order properties of a spatial point process describe how the mean number of points per unit area (intensity) varies through space. Point-to-point nearest neighbor and origin-to-point nearest neighbor statistics provide objective methods to examine small-scale interactions between points (first order intensity) by calculation of their respective empirical distribution functions (EDF). If the respective point-to-point and origin-to-point nearest neighbor EDFs differ significantly, then a departure from randomness is assumed. The second-order properties (second order intensity) of spatial point processes describe the interaction or spatial dependence between points through space and can be described by Ripley's *K*-function. In this case, if *y* is the mean number of diseased plans per unit area (a density), then yK(d) is the number of diseased plants within the distance (d) of an arbitrary (or randomly selected) diseased plant.

By plotting estimates of $\widehat{K}$ as a function of distance, and comparing it to a plot of estimates of complete spatial randomness (CSR), the range of spatial dependency (RSD), i.e., the range of distance over which there is a departure from CSR, and the maximum spatial dependency (MSD), i.e., the distance at which the maximum departure from randomness occurs, can be estimated. Ripley's original method ignores the finite size of the plot in calculating the CSR prediction and uses a weighting that inflates the pairs counts to compensate for the reduced numbers of pairs near the plot boundaries, i.e., edge effects. At low incidence, this weighting tends to overestimate the number of pairs at the boundaries of the plot. An alternative method for calculating the *K*-function has recently been presented that directly incorporates the edge effect into CSR model predictions, yielding more accurate estimates of the variance (32). In what follows, this approach is extended to a sample space consisting of an irregular lattice of points (trees). For this case, the prediction of infected point pairs based on CSR is obtained by direct enumeration of tree pairs. Thus, no edge correction is necessary and expectation values and confidence limits can be obtained by direct statistical inference.

--- Page 364, left hand side :

Evaluation of the spatial point pattern by modified Ripley's *K*-function methods. Evaluation of the spatial point pattern (SPP) of citrus canker from a regional perspective in Dade and Broward counties entailed an examination of differences between the trees in general within the respective study sites.

This was accomplished by the comparison of two cumulative distribution functions (*cdf*s): one representing the fraction of infected tree pairs less than or exactly a distance (*d*) apart [$cdf_t(d)$] and the second distribution for the total population of trees [$cdf_T(d)$]. Assuming that out of a total of *N* trees in the study site, *I* are citrus canker- infected. These *cdf*s can be expressed as

$$cdf_T(d) = \frac{2\sum_{j=1}^{N}\sum_{k>j}^{N} m_{jk}}{N\cdot(N-1)} \qquad (1)$$

$$cdf_I(d) = \frac{2\sum_{j=1}^{I}\sum_{k>j}^{I} s_{jk}}{I\cdot(I-1)} \qquad (2)$$

---

where *m* = 1 for tree pairs that are ≤*di* apart and *m* = 0 otherwise, and likewise, s = 1 for infected tree pairs that are ≤*di* apart and s = 0 otherwise. The CSR assumption translates to the expected equivalence of equations 1 and 2 so a constant fraction {*I*(*I* – 1)/ [*N*(*N* – 1)] } of tree pairs are infected, irrespective of distance (*d*). The factor of '2' in the numerators of equations 1 and 2 accounts for the fact that the pairs are unordered (i.e., *k* > *j* in second summations for equations 1 and 2). For a particular distance (*d*), the probability of selecting infected pairs in a sample of *N*(*N* – 1)$/2\ cdf_T(d)$ tree pairs chosen randomly from a population of size *N*(*N* – 1)/2 of which *I*(*I* – 1)/2 are infected is given by the hyper- geometric distribution:

$$H\left(i, \frac{N(N-1)\cdot cdf_T(d)}{2}, \frac{I(I-1)}{2}, \frac{N(N-1)}{2}\right) \qquad (3)$$

with mean value

$$i_{\exp} = \frac{I(I-1)\cdot cdf_T(d)}{2} \qquad (4)$$

and variance

$$\sigma^2_{\exp} = i^2_{\exp}\cdot[1-cdf_T(d)]\cdot\left(1-\frac{I(I-1)}{N(N-1)}\right) \qquad (5)$$

An alternative method for calculating the *K*-function has recently been presented that directly incorporates the edge effect into CSR model predictions, yielding more accurate estimates of the variance (32). In what follows, this approach is extended to a sample space consisting of an irregular lattice of points (trees). For this case, the prediction of infected point pairs based on CSR is obtained by direct enumeration of tree pairs. Thus, no edge correction is necessary and expectation values and confidence limits can be obtained by direct statistical inference.

For the range of incidence and sample size examined in this study, the hypergeometric function is well approximated by a normal distribution with the same mean and variance. This fact can be used to estimate confidence intervals.

The citrus canker-infected trees are tested for spatial dependency by applying the one-sample single-tailed Kolmogorov-Smirnov test to the maximum distance $\{D = D = max[cdf_t(d) - cdf_T(d)]\}$ between the above two cumulative distributions (3). The above describes a new analytical approach to SPP analysis that is an outgrowth of the modification to Ripley's *K*-function presented by Ward and Ferrandino (32). Note this analysis is equivalent to Ripley's *K*-function in the limit of an infinite number of trees uniformly covering the study site.

The above analysis was accomplished through the use of a VBA written by F. Ferrandino and compares the infected SPP to the total SPP. This generalized Ripley's *K*-function was used to examine SPP of the five research sites regardless of time period and subsequently to examine the cumulative disease incidence for each site by time period T1 through T25. Via this method, analytical results were obtained for the range of spatial dependency (RSD), i.e., the distance at which the estimated and observed *cdf*s in equation 1 intersect; the effective range of spatial dependency (RSD_{eff}), the distance over which the *cdf*s in equation 1 were significantly ($P < 0.05$) different; and the distance at which maximal spatial difference (MSD) occurred, and the cdf (K-value) associated with the maximum spatial difference (Max_{diff}).

Page 370. second paragraph, left hand side

Ripley's *K*-function. Modified Ripley's *K* analysis of the citrus canker SPP for each of the study sites was performed encompassing the entire study period. That is, the SPP evident in each time period, T1 to T25, was calculated and examined in order to better understand and explain the change among successive SPP associated with the spread of ACC in the study sites. The SPP examined in each time period consisted of the new infections occurring in that time period, as well as all diseased citrus trees from previous time periods. The dynamics of the range of spatial dependency (RSD = the distance at which the estimated and observed modified Ripley's *K cdf* values intersect), the effective range of spatial dependency (RSD_{eff}) at $\alpha = 0.05$, and the maximum departure from randomness (Max_{diff}) and its associated distance (MSD) were calculated for each 30-day period (Figs. 4 and 5A to D). Aggregation of the SPP, exhibited by a significant difference between the cumulative distribution functions in equations 1 and 2 above, was exhibited in all of the respective sites through the entire range of time. By applying this method across the 25 temporal periods of the study, the regional spatiotemporal relationships of ACC were examined for each study site and among study sites. In general, aggregation increased concomitantly with disease incidence. This aggregation was expressed by the RSDeff that increased across all study sites and approached a maximum during the first few temporal periods. The $_{RSDeff}$ associated with sites D2, D3, and B2 increased through time until it reached a maximum plateau, whereas for D1 and B1, the $_{RSDeff}$ occurred during T3 and T7, respectively, and decreased over the next several temporal periods before reaching a lower plateau. The maximum RSD_{eff} for sites D1, D2, D3, B1A, and B1B was 1.53, 2.13, 0.85, 3.78, and 1.61 km and corresponded to 30-day periods T3, T2, T18, T6, and T9, respectively. The greatest departures from randomness, $Max_{diff,}$ for D1, D2, D3, B1A, and B1B were *K*-values of 0.45, 0.69, 1.00, 0.73, and 0.71 and corresponded to distances of 1.03, 0.69, 0.01, 0.95, and 0.31 km, respectively. Examination of the SPP maps for each plot revealed that for D1, D2, and D3, the distribution of citrus canker-infected trees over the extents of each plot was first seen at T3, T2, and T17, respectively, and related to RSD_{eff} spatiotemporal distance relationships of 1.5, 2.2, and 1.6 km, respectively.

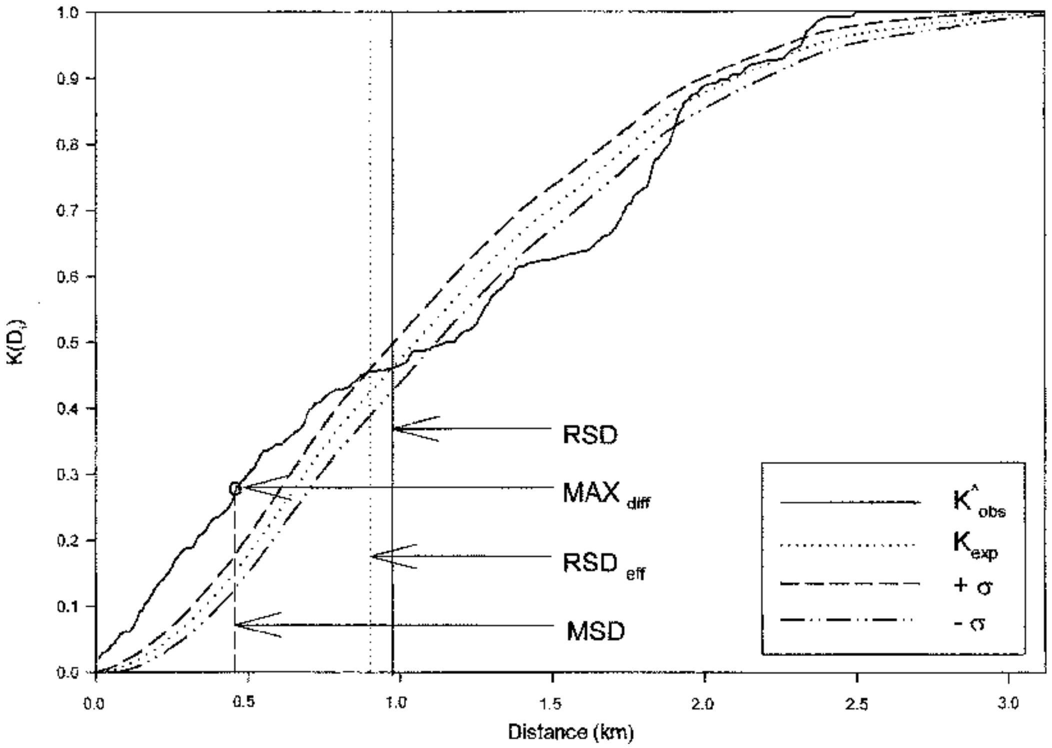


Fig. 4. Example of the calculation of the modified Ripley's *K*-function for the citrus canker spatial point pattern over distance for site D1, for a single time period T4. Dotted line represents the estimated *K* with surrounding á = 0.05 confidence limits (dashed lines). Solid vertical line crosses the *x* = distance axes at the point relative to the intersection of estimated and observed *K*-value, and represents the range of spatial dependency (RSD) in kilometers. Dotted vertical line crosses the *x* = distance axes at the point relative to the intersection of the observed *K*-value where it enters into the upper confidence limit, and represents the effective range of spatial dependency $_{(}$RSD$_{eff)}$ in kilometers. Open circle and dashed vertical line show the maximum departure from randomness (Max$_{diff}$) and the associated maximum spatial difference (MSD) in kilometers. Calculation are described in equations 1 to 5.

Note: Figures 5A-D as presented on page 373 of the published article, are a part of the spatial point pattern analysis. These figures were not considered important for this review. The reader can refer to the published article on the website for these figures.

3. REVIEW OF THE SPATIAL POINT PATTERN ANALYSIS

The above excerpts do not include discussion of the "range of spatial dependency" (RSTD) as given on page 365 and 370. This analysis is reviewed in Appendix G, Semi-Variance Analysis.

Equation 1, on page 364 is an empirical cumulative distribution related to the inter-tree distances in a particular site. The probability of the distance between two randomly selected trees being al distance "d" or less apart is equal to $cdf_T(d)$. As a statement of probability, $P\{D \leq d\} = cdf_T(d)$, where D is the random variable representing inter-point distances.

The empirical distribution for the infected trees, $cdf_t(d)$, is calculated in a similar manner as given in equation 2. Both equations 1 and 2 are correct means of estimating these probabilities, consistent with quantile methods. As stated in the article, the number of inter-point distances for infected trees is $(I-1)/2$ and for the total population is $N(N-1)/2$. The infected trees are a subset of the total population (full set), with members equal *I* and *N,* respectively.

The spatial point analysis reviews two analyses:

- A hypothesis test as whether the pattern of all citrus trees is significantly different from the pattern of diseased trees
- A determination of spatial dependency by a new approach identified as "modified Ripley-K" method.

It is noted after considerable literature review, we could find no application of either the DNC parsing method, or the modified Ripley-K method as presented in the published article. The authors do not offer any previous applications.

As discussed previously, these equations are completely consistent with inter-point distance analysis, and not the Ripley-K method. For this reason, the statistics and approach will be denoted as "IPD" statistics or IPD statistical analysis", respectfully.

Prior to defining the empirical distributions (equations 1 and 2), it is stated, "Assuming out of a total of *N* trees in the study site, *I* are citrus canker infected." This seems a strange statement, as the number of trees infected with citrus canker would be known based on survey data.

HYPOTHESIS TEST

The discussion beginning on page 364, beginning with "The CSR assumption ..." seems out of place or details missing, as no "CSR assumption" has been discussed prior to this statement. The paragraph ends with the statement, "This fact can be used to estimate confidence interval", yet is far from clear what "this fact" refers to. These statements and analyses in this paragraph is discussed in detailed in Appendix F1. It is considered a theoretical discussion.

The next sentence provides the details for a Kolmogorov-Smirnov test, a well known means of comparing two distributions.

> The citrus canker-infected trees are tested for spatial dependency by applying the one-sample single-tailed Kolmogorov-Smirnov test to the maximum distance $\{D = max[cdf_t(d) - cdf_T(d)]\}$ between the above two cumulative distributions (3).

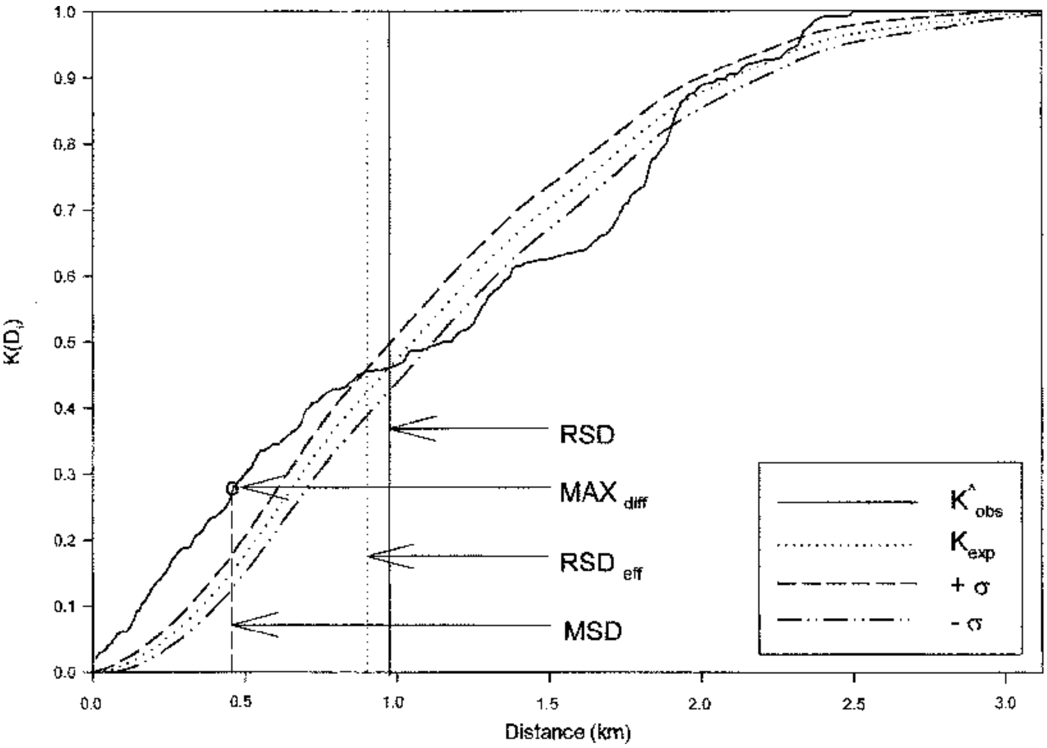
Following this statement, no test results are given. It is very odd. The authors have presented the means by which the empirical distributions can be calculated, and proposed the K-S tests, then fail to present any results.

It should be recognized that the citrus trees (healthy and infected) are likely to be aggregated. The aggregation of citrus trees is an intuitive conclusion based on where citrus trees are typically located in the backyards and these backyards are typically adjoining. Citrus trees are rarely planted in front yards, school yards and public parks. Large non-citrus areas exist in the study sites including parking lots, lakes, canals, industrial parks and commercial centers.

SPATIAL DEPENDENCY USING THE INTER-POINT DISTANCE (IPD) STATISTICS

The figure below is presented on page 372. We will refer to the solid line as the "observed" cumulative distribution, while the dotted line is the "expected" cumulative distribution with confidence intervals.

Figure 5: Spatial Statistical Analysis results



The notation with Figure 4 is provided below:

> **Fig. 4.** Example of the calculation of the modified Ripley's *K*-function for the citrus canker spatial point pattern over distance for site D1, for a single time period T4. Dotted line represents the estimated *K* with surrounding α = 0.05 confidence limits (dashed lines). Solid vertical line crosses the *x* = distance axes at the point relative to the intersection of estimated and observed *K*-value, and represents the range of spatial dependency (RSD) in kilometers. Dotted vertical line crosses the *x* = distance axes at the point relative to the intersection of the observed *K*-value where it enters into the upper confidence limit, and represents the effective range of spatial dependency $_{(}$RSD$_{\text{eff})}$ in kilometers. Open circle and dashed vertical line show the maximum departure from randomness $_{(\text{Maxdiff})}$ and the associated maximum spatial difference (MSD) in kilometers. Calculation are described in equations 1 to 5.

The functions $K\^_{obs}(D_i)$ and K_{exp} are not specifically defined in equations 1 to 5. From the overall discussion, it is concluded that $K\^_{obs}(d)$ is the inter-tree point distance distribution for canker infected trees as calculated by Equation 2. A calculation example is given on page 8 using 4 points.

The introductory statements on page 362, would indicate the expected curve is the cumulative distribution of a theoretical CSR pattern.

> By plotting estimates of $\widehat{K}$ as a function of distance, and comparing it to a plot of estimates of complete spatial randomness (CSR), the range of spatial dependency (RSD), i.e., the range of distance over which there is a departure from CSR, and the maximum spatial dependency (MSD), i.e., the distance at which the maximum departure from randomness occurs, can be estimated.

But, there is not one sentence within the published article which reveals how the theoretical CSR distribution was generated. For a square or circular area, it may be calculated numerically, as given in Diggle (see references at the end of this appendix). Most certainly, the curve was generated using Monte-Carlo simulation.

Why would the authors fail to mention this? Monte-Carlo simulation has been used for over 40 years, and there would be no question as to the validity of a distribution from this pattern. The CSR pattern have to be consistent with the boundaries of the site, which seem difficult to pin down (see Appendix A).

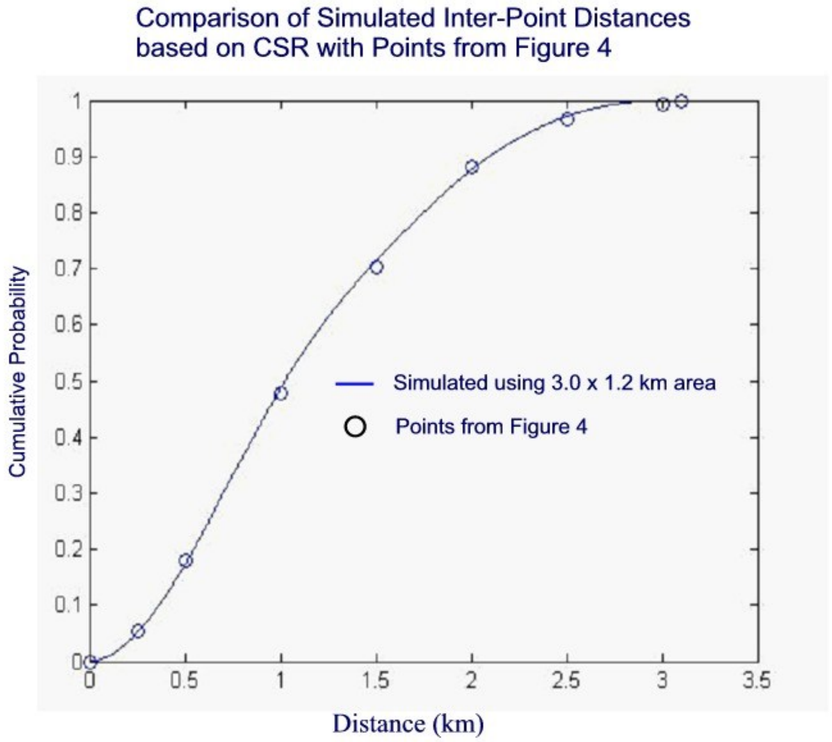
However, the many realizations of the CSR pattern itself, would be absurd if any one realization was overlain on a street map. Points would placed randomly, everywhere, on top of houses, in lakes, canals, roads, etc. It follows that the departure from a pattern that could not exist, is not a meaningful analyses. There is no evidence within the published paper that any spatial analysis on the healthy trees was analyzed.

FURTHER INVESTIGATION OF THE CSR CURVE IN FIGURE 4

It is believed the expected curve in Figure 4 was generated by simulation using the Matlab software. To test this theory, a CSR pattern was generated using a 3.22 x 3.22 km (10.4 km^2, equivalent to 4 square miles, per the area given in the 2002 article. The simulation was run with 300 points. The cumulative distribution from simulation was a poor match to Figure 4. The model was run a second time with the dimensions 3.22 x 1.61 km (2 x 1 mi rectangle) and again, the distribution did not match.

After some experimentation, a good match was made with 3.0 x 1.2 km as shown in below in Figure 5. As stated in the caption in the published article, this distribution is for the Site D1. It is concluded that the authors chose to use a rectangle with an area is 3.6 km^2 , instead of the 10.3 km^2 as given on page 363 of the published article. While it not in the intent to determine methodology based on presented results, the incomplete description of procedures made this approach necessary. If this result is correct, then the authors have been been deceptive in the method, and parameters used (size and geometry of study site).

Figure 5: Comparison between simulation and field study results.



RESULTS

Results are shown in Figure 5A- D of the article. The data used are from the “distance necessary to circumscribe” method where the time periods are not real time observations, but calculated or synthetic time periods based on lesion ages. The method is discussed in Appendix B.

Each successive time period adds to the infected tree population, so changes to the empirical distribution should diminish with time. Therefore, range of spatial dependency in Figures 5A (RSD) and 5B (RSD_{eff}) are similar.

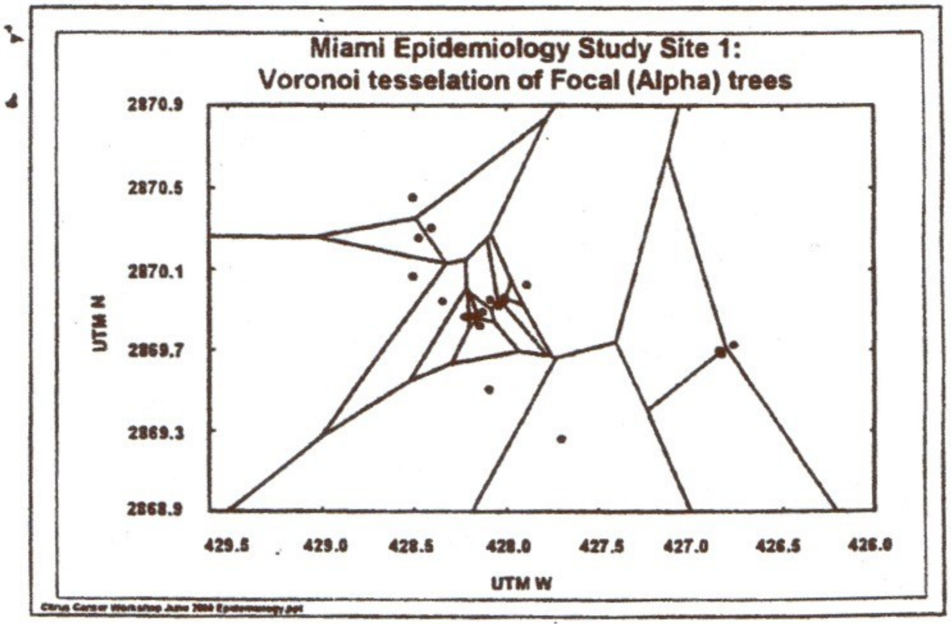
The results are not reviewed in detail because the statistical method employed is improper. Any evaluation of aggregation of citrus canker infected trees, must first show the aggregation is not due to the general population of citrus in the site being non-randomly distributed. It would be intuitively obvious that the trees should be aggregated because of the extensive non-citrus areas in the sites and the normal practice of planting citrus in back yards, close to adjoining back yards.

Further, a detailed review of the distances from spatial dependences, are not reviewed, because they are not related to any “distances of spread” nor is there any mention of these distances should be considered as guide for increased eradication radius.

4. OTHER PRESENTATIONS

VORONOI TESSELLATIONS

Figure 6: Voronoi Tesssellations from Gottwald's presentation (Viewgraphs submitted into evidence, November 2000)



This viewgraph was presented in the Broward County court house and it is believed to be the same as presented at the International Citrus Canker Research Workshop in June 2000, because the footnote at the bottom states 2000 Citrus Canker Workshop June 2000.ppt.

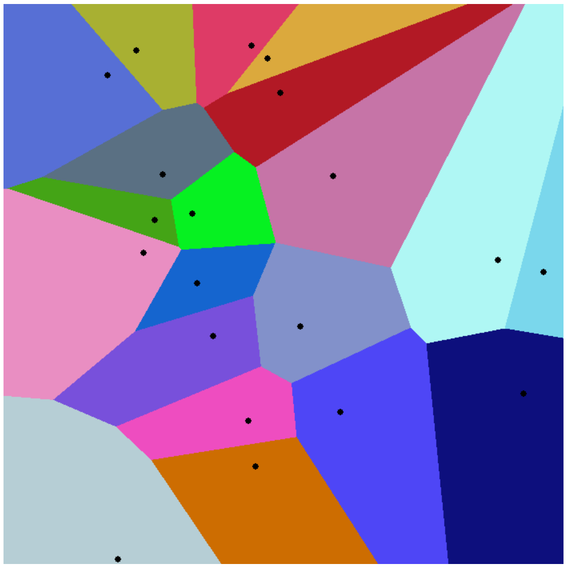
> To start to analyze this data at the suggestion of Dr. Gareth Hughes, a friend of mine and colleague, we decided to try to look at the distribution from known focal trees to surrounding trees. To do that, we attempted to use what is known as Voronoi tessellation. This allowed us to carve up this data set such that if we used a focal tree and examined that, we knew anything that fell within this area (indicating) given to us by the tessellation was closer to this focal tree than any other focal tree in the area.
>
> That worked very well for a few number of focal trees, but as the number of focal trees increased, that particular analytical technique no longer worked for us. This would be what it would look like if you overlaid it on top of, in this case, Site 1. We weren't

> able to then use this analysis, so we had to switch away from the Voronoi tessellation technique and developed our own program for estimating distances, and I'll talk about that, through a series of Visual Basic routines which analytically do the same thing as that technique which is a spatial analysis to try to determine how far citrus canker will spread from individual points of infection.

Voronoi tessellations is a means of subdividing a given area into subareas based on boundaries drawn equidistant from each point in the area.

An example is shown below. In mining, Voronoi polygons are used to estimate the reserves of valuable materials, minerals, or other resources. Exploratory drillholes are used as the set of points in the Voronoi polygons.

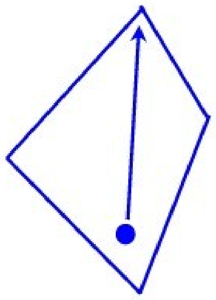
Figure 7: Example of Voronoi Tessellations



It is theorized that Voronoi tessellations were used perhaps early in the planning process to help define eradication radius. The area of each bounded polygon can be calculated using commercial software programs. The areas could then be ranked in order, lowest to highest, and using percentiles, a large area equal to the 95% quantile could be identified. Then a radius equal to this large area would be calculated, to assure clear-cutting of 95% of all citrus trees in an area.

Alternatively, the distance between infected trees and the furthest corner of the polygon could be calculated. A similar analysis of these distances, using percentiles, could be used to identify a radius to clear-cut a high percentage of all citrus trees. This is illustrated in Figure 8.

Figure 8: Use of furthest corner to define eradication radius



This conclusion is speculative. but it appears to be the most obvious reason to use this method.

CONCLUSIONS

1. All results are based on the highly subjective and an unscientific means of parsing the trees into time periods using oldest lesion ages. There are no observed data, but inspectors discovered trees many months after infection dates as described in Appendix A.

2. While the authors claim their method is a new extension of Ripley-K analysis, it is actually an application of inter-point distance analysis. This form of analysis has been in the literature for approximately 40 years.

3. Hypothesis testing to determine whether the distribution of infected trees is significantly different from the total population of trees was proposed in the article but never done. There is no explanation of why hypothesis testing was not done.

4. The published article compares an observed distribution with an "expected" distribution. The means by which the expected distribution was calculated is not given in the article. It is suggested that the Monte-Carlo simulation of inter-point distances was used to generate this distribution.

5. Any presentation of the CSR point pattern overlain on a street map, would show points (trees) in lakes, canals, on roofs of house, in parks and other "non-citrus" areas. Thus, the departures from spatial randomness are invalid, because the CSR pattern used to generate the the expected distribution would contain impossible point locations. The definitions and criteria for CSR patterns requires there must be equal probability of an event (a tree being present) at every location.

6. Monte-Carlo methods were used to replicate the “expected curve” in Figure 4 . A match was only possible with dimensions of 3.0 x 1.2 km, inconsistent with the given area of site D1 of 4.0 square miles (10.36 km^2). Thus, it would appear that a subset of the collected data in Site D1 was used in calculation of measures of spatial dependencies.

7, The Voronoi tessellations analysis is another form of spatial point pattern analysis. It was presented in Broward Court in November 2000. It is also believed to have been presented in the June 2000 at the International Citrus Canker Research Workshop. The analysis may have been part of early efforts to define a eradication radius. This is reviewed in Chapter 7.

REFERENCES

Gottwald, T.R., X. Sun, Riley, T. Graham, J.H., Ferrandino, F. and Taylor, E., 2002, Geo-Referenced Spatiotemporal Analysis of the Urban Citrus Canker Epidemic in Florida, Phytopathology, Vol 92, No. 4.

Madden, L.V., Hughes, G. and van den Bosch, F., 2006, The Study of Plant Disease Epidemics, 2006, The American Phytopathology Society, St. Paul, Minnesota (ISBN 978-0-89054-354-2).

Campbell, C. Lee and Madden, L.V. 1990, Introduction to Plant Disease Epidemiology, John Wiley and Sons, New York. (ISBN 0-471.83236-7)

Diggle, P.J., 2003, Statistical Analysis of Spatial Point Patterns, Oxford University Press, New York. ISBN 0-340 74070.