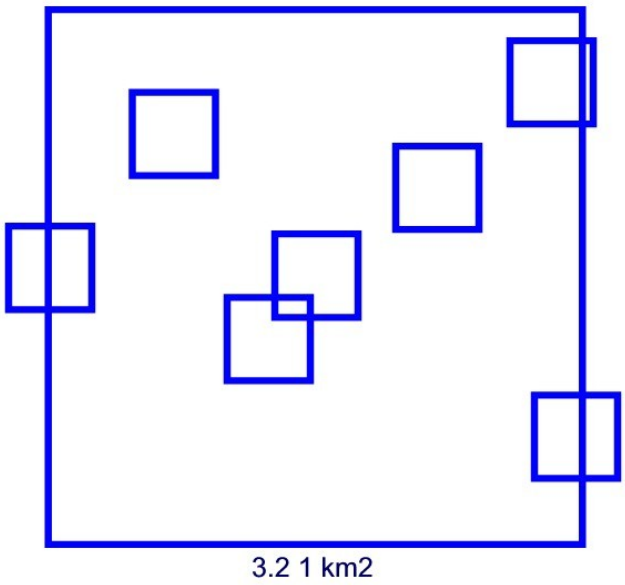
# APPENDIX E

# RANDOM QUADRAT SAMPLING AND RELATED ANALYSES




Seven quadrats are 0.5 $km^2$ on a side in a 10.8 $km^2$ (4 square mile) area.

## INTRODUCTION

The random quadrat sampling was used in the Florida field study in correlation analyses (Figure 6 of published article) and krig mapping (Figure 7, published article). Quadrat sampling as reviewed here, is based on information from the published article of the Florida field study (2).

Random quadrat sampling is a standard method for disease assessments and discussed on page 373 of Campbell's textbook (1). Random sampling is used to describe characteristics of a large area which are impractical to completely survey. The procedure provides evaluation of changes in the population in a timely manner. Random sampling is used epidemiology and other sciences including environmental science and ecology. Typically, sampling is done out of necessity, because the area of interest is too large. Because it is random sampling, the sample should be unbiased, and proper manner to make inferences to the general population.

Large area sampling is not the application here. This is an unusual application of random sampling. According to the article, the study areas were completely surveyed, and quadrat sampling was done after data collection was completed..

The sampling process is termed “stochastic quadratization” by the authors. This may sound complicated, but it is actually very simple in this particular case. Sampling units are randomly selected within the study site. The sampling units or quadrats in this case are squares. All relevant information within the squares is collected. The center of the squares are generated using a uniform random number generator for the X and Y Cartesian coordinates. A procedure which uses random number generation may be called a stochastic procedure or a Monte Carlo simulation.

The illustration on the title page, shows seven randomly located quadrats. Some squares will overlap, and some will cross the boundaries of the site. It is possible to take preventative measures to eliminate quadrat overlap and extension beyond the boundaries of the site, but there is no mention of this in the published paper.

Figure 6 of the published article (2) shows the 10 scatter diagrams. Various attributes such as host density are compared with other attributes, such as host susceptibility. The first impression is one of poor correlation as none of the regressions appear adequate for prediction purposes. So, one of the most obvious questions, is there any value at all to these correlations? But, a second question immediately comes to mind, why use random quadrat sampling at all for the correlation of variables? Figure 7 of the published article (2) shows a series of maps, with an increasing presence of citrus canker. How should these maps be interpreted?

## Approach to the Review

A “top down” review would quickly conclude that all correlations as given in Figure 6 are poor and therefore the linear regression relationships can not be useful as to make inferences beyond the study sites. Also, the published article does not attempt to extend the regression results to the general population (Florida canker epidemic), so the impact of correlation analysis seems minimal on the overall analysis. But, there is an implied conclusion, that the computed indexes show either no or poor correlation.

This review of the quadrat related analyses went further. What did the analyses expected to accomplish? Was the method appropriate? If the data had been different, might the analyses as presented have been successful in developing meaningful relationships? It is remembered by the time the study was published about four years since it was begun, certainly sufficient time to consider various approaches to correlation analyses.

One would expect the susceptibility of the tree to citrus canker based on cultivar/species and the number of trees within a backyard (host density) would be strong factors in determining the likelihood that a tree would be infected. Also mature trees are known to be less susceptible to citrus canker. (5) The age of the tree is sometimes estimated based on the height and girth (measured at a fixed height). The girth was not measured in the study.

As noted in Chapters 3 and 7, and Appendix A, the citrus leafminer (CLM) mines have been cited as greatly increasing the susceptibility of a tree to citrus canker. (5) The presence and intensity of citrus leaf miner would be easily identifiable. It might be correlated with a number of factors including the cultivar of the tree. Yet, this was not part of the data collection on the approximately 19,000 citrus trees in the study.

This review with close scrutiny to the procedures used was motivated by these seemingly odd, but intriguing aspects of both data collection and correlation analyses. The approach in this review is to first examining quadrat sampling process, then how it was applied to the field study data, and following this, to explore in more detail the two resulting analyses- regression and contour mapping.

## 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. (Reference 2) Published by the American Phytopathology Society with no copyright protection.

Every effort has been taken to transcribe the excerpted passages exactly as published. Figures and table numbers used in this section are based on the published article. These selected sections may inadvertently exclude some 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, including www.citruscankerdocs.com.

-- Page 366, left hand side, under **Materials and Methods,** subsection **Spatiotemporal analysis of the spatial point pattern.**

> Effects of plant density and cultivar susceptibility on disease incidence and severity were examined via a repeated stochastic sampling of square quadrats of 0.25, 0.5, 1.0, 1.5, and 2.0 $km^2$ areas, respectively. Sites D1 and D2 were utilized because they represented large study areas where no disease trees were removed during the study. A VBA was used to randomly select a centroid point for each of 500 quadrats within sites, so as not to bias the calculations by location. A criterion was imposed such that only quadrats containing citrus trees and only quadrats with at least two diseased trees were accepted as part of the 500 quadrat population. For each quadrat, density of citrus trees, final disease incidence, and index of citrus species/cultivar susceptibility were calculated. During data collection surveys, citrus species/cultivar were assigned to 13 categories. Each category was given a susceptibility rating (0 to 6) based on a combination of prior published studies of susceptibility/resistance (7,14,18,20,21,23). Subsequently, an alternative susceptibility rating was utilized that was based on the actual disease incidence of each of the 13 species cultivar categories within each respective site. Trees were also assigned four height categories and their canopies divided into 12 sectors (north, east, south, and west and top, middle, and bottom). The following three indices were calculated for each tree and average indices for each quadrat. The index of disease susceptibility was calculated as:

$$I_{sus} = \left( \sum_{i=1}^{t} c_i \right) / t \qquad i = 0, 1, \ldots .t, \qquad (6)$$

where *c* = the normalized cultivar susceptibility rating for each tree and *t* = the number of trees in each quadrat. The index represents a composite estimation of susceptibility based on all trees and accounted for the diversity of cultivar mixture within a given quadrat.

The index of disease severity was calculated as:

$$I_{sev} = \left( \sum_{i=1}^{t} (\mathrm{DSB}_i \times h_i \times q_i) \right) / n \qquad i = 0, 1, \ldots \ldots t, \qquad (7)$$

where *n* = the total number of quadrats in the study area, DSB = the binary disease status of each tree (0,1), *h* = the normalized height of the tree, and *q* = the proportion of sectors infected for each tree. This index represents a composite estimation of disease severity across all trees in a given quadrat and takes into account tree size and the relative volume of the canopy expressing disease.

The index of host density was calculated as:

$$I_{hd} = t/(\mathrm{QS}) \qquad (8)$$

:

where QS = quadrat size in square kilometers. This index represents a calculation of the citrus tree population saturation (relative to the quadrat with the highest density) of a given quadrat. Each of the above indices was normalized, resulting in values from 0 to 1 for each quadrat sampled.

Index of disease severity values were also used to perform a semivariance analysis followed by a kriging of the data by the block method to visualize the occurrence and position of foci and the development and spread of disease through time. Kriging was performed at four time periods selected to best represent periods following significant increases in disease.

-- Page 371, left hand side, under **Results** section:

**The effects of host plant density and cultivar/species susceptibility on disease incidence and disease severity.** Of the various quadrat sizes tested, the 0.25-km$^2$ quadrat size resulted in the clearest relationship among those variables and indices examined, and therefore, was used for all further analyses. The random distribution of the centroids of each of the quadrats was selected by the stochastic process for sites D1 and D2. The appropriateness of the stochastic quadratization method was evaluated against a simple nonoverlapping quadratization method, the latter of which resulted in far fewer quadrats for comparison. The same data trends

were found with both methods, indicating no unique outcomes associated with the use of stochastic overlapping quadratization, and led to its application for all further comparisons.

Linear regression of disease incidence versus host plant density resulted in low coefficients of regression and slightly positive slopes for both D1 and D2, indicating little or no effect of host plant density on disease incidence. However, the variance associated with the relationship of disease incidence to host plant density decreased as density increased. That is, at lower host densities, disease incidence was more variable and variability decreased with increased host density. This decrease was more apparent for site D1 than for D2. Similarly, linear regression of disease severity (the proportion of diseased sectors of individual trees) versus host density indicated a slight positive slope for both sites, indicating little or no effect of host plant density on disease severity (Fig. 6A and F). The associated variance for this relationship also decreased as density increased. Linear regression of disease incidence versus the index of disease severity resulted in a positive slope and high $r^2$ values for both sites, indicating that much of the variation due to regression was accounted for and indicated a direct relationship between increasing disease incidence and increasing disease severity, as expected.

Susceptibility was best expressed as a function of proportion of diseased individuals in each cultivar/species category that became infected relative to each site (data not shown), and this index of susceptibility was used for all further comparisons. For both sites, linear regression resulted in positive slopes for both disease incidence and the index of disease severity versus the index of susceptibility (Fig. 6B and G). The associated $r^2$ of regression values accounted for more of the variation due to regression for site D2 compared with site D1, and for D2 demonstrated a much greater effect of host susceptibility on incidence and severity. Normalized susceptibility was heavily clustered in the low and midrange for sites D1 and D2, respectively, with considerable variation in disease incidence and disease severity. This clustering of values represented a high population of plants with similar susceptibility in both sites that was independent of host density.

The combined effect and interaction of host density and the index of susceptibility of quadrats on incidence and severity was also investigated (Fig. 6C and H). The associated $r^2$ of regression values, although low, still accounted for more of the variation due to regression for site D2 than site D1, and for D2 demonstrated a more positive slope, indicating a greater effect of the susceptibility–density index on incidence and severity. Although more prevalent for D1 than D2, the variance associated with the relationship decreased as incidence and severity increased, indicating a better relationship of higher values of the susceptibility–density index with incidence and severity.

Kriging of the *Isev* demonstrated the occurrence and development of foci of disease in each of the urban areas (Fig. 7). For each of the study sites, foci of infection that established early in the study can be seen. As these foci continued to increase in severity and spread locally, additional foci began to appear. The effect of previously established foci on the establishment and evolution of new foci was seen. It was noted that early in the epidemic foci often became established at considerable distance from each other. These foci continued to enlarge while simultaneously new secondary foci began to fill in the previously uninfected areas between the original foci.

-- Page 379 Left hand side, under **Discussion:**

The kriging of the index of severity through time visually demonstrated the establishment of foci of infection and the spatial evolution of secondary foci that caused the filling in of the uninfected

areas with disease. This is not surprising and has been demonstrated many times on the plot or field scale, but here was demonstrated on a regional spatial scale in large urban areas.

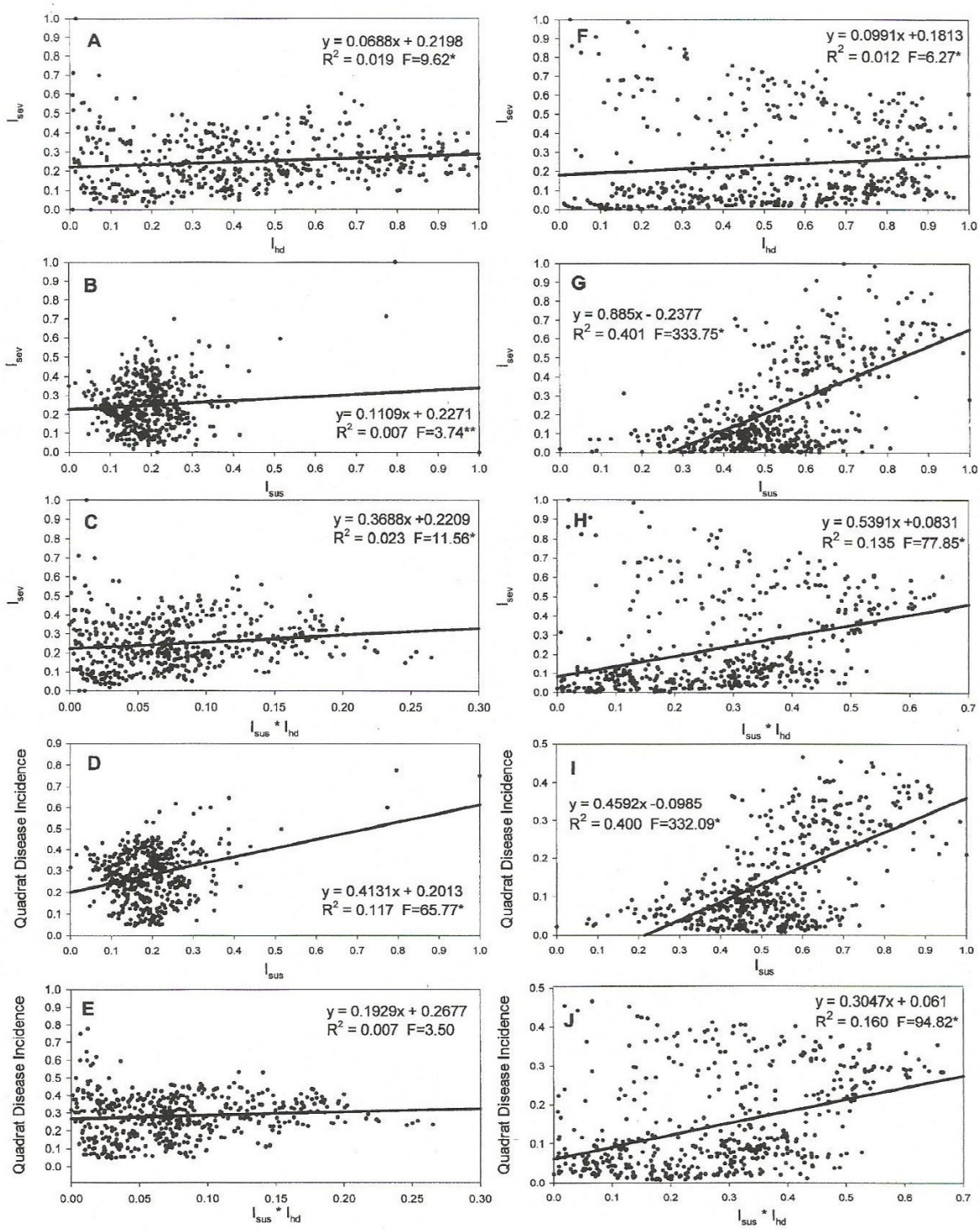



**Fig. 6.** The effect of citrus host plant density and cultivar/species susceptibility on disease incidence and severity for site D1 (**A to E**) and D2 (**F to J**) based on stochastic placement of 500, 0.25-km$^2$ centroids. **A and F,** The effect of host density on disease severity. **B and G,** The relationship of the normalized index of susceptibility and disease severity. **C and H,** The relationship of the normalized index of susceptibility–normalized host density and normalized index of disease severity. **D and I,** The relationship of the normalized index of susceptibility–normalized host density and quadrat disease incidence. **E and J,** The relationship of the normalized index of susceptibility–normalized host density and disease incidence. Calculations are described in equations 6 to 8.

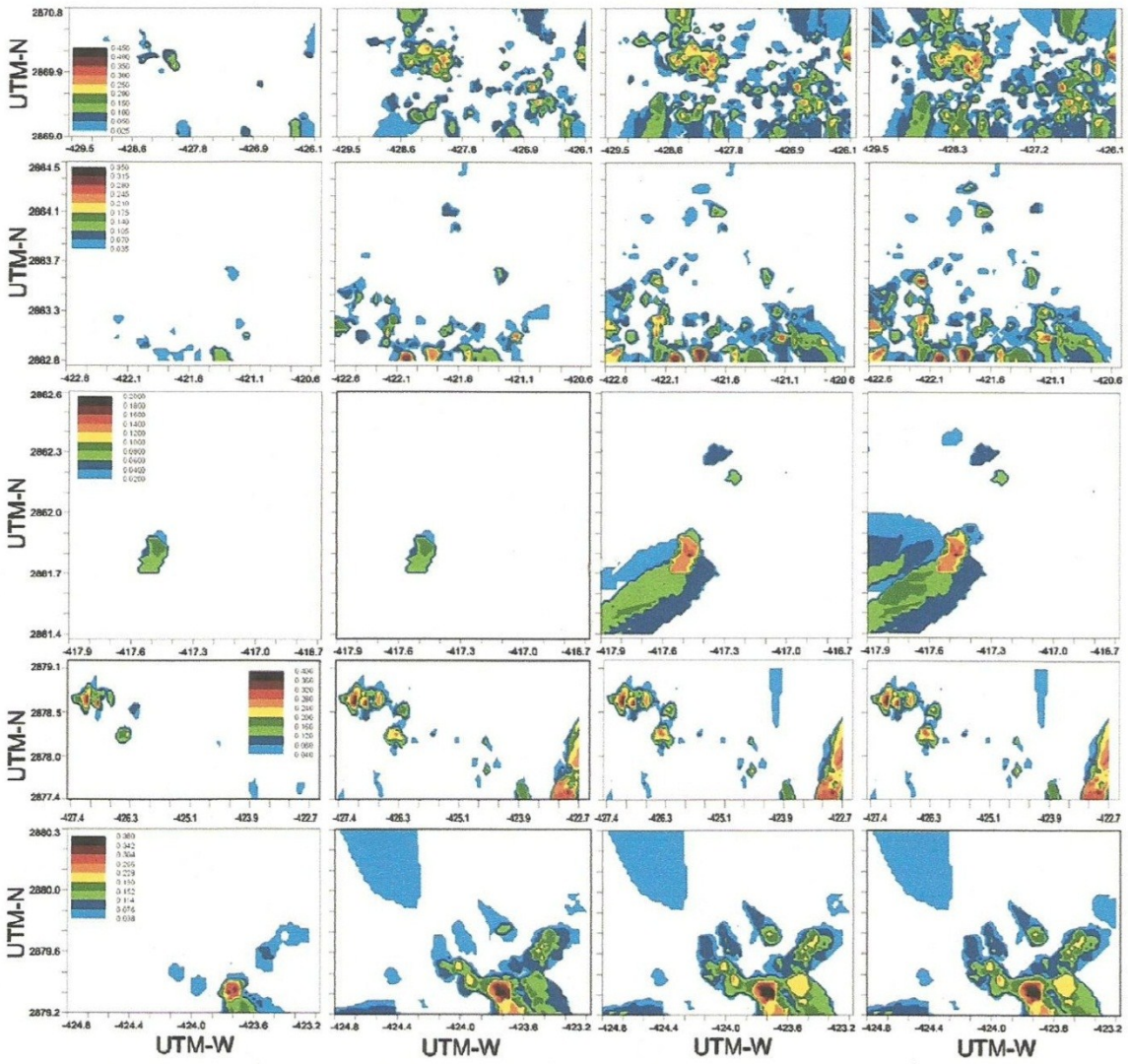



**Fig. 7.** Kriging estimate maps based on the normalized index of disease severity ($I_{sev}$) of citrus canker-infected trees in five study sites in Miami, Dade and Broward counties. Columns of panel graphs from left to right represent 30-day periods T5, T10, T14, and T18, respectively. Rows of panel graphs from top to bottom represent study sites D1, D2, D3, B1, and B2, respectively. UTM-W and UTM-N = universal transverse mercator west and north, respectively. UTM is a geo-referenced measurement system used to locate points on the earth's surface. UTM measurements are in kilometers. Note establishment of primary foci early in the epidemic that enlarge through time combined with spread due to the development of secondary foci that fill in the areas between.

# QUADRAT SAMPLING AND RELATED CORRELATION ANALYSIS

## DESCRIPTION OF REGULAR OR CONTIGUOUS QUADRATS

Regular and randomized quadrats for sampling purposes in plant disease epidemiology are discussed in reference 1, page 291- 295, and references 4, page 280- 281. Some obstacles in using random quadrats in commercial applications is discussed in reference 4, page 281.

An example from reference 3, page 91 related to the distribution of a soil borne fungi, *Sclerotium cepivorum* in a small area of a field. Each number represents the number of sclerotia per 100 g of soil for the plot. Each plot was 1.5 x 30.5 m.

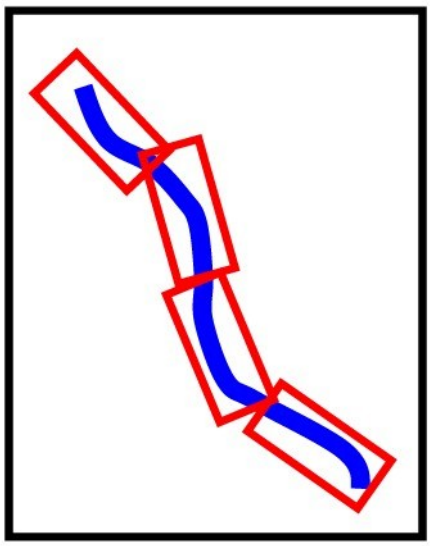
**Figure 1: Example of regular quadrats,** from World Distribution of Soilborne Mycoparasites- An Evaluation, by P.B. Adams from Spatial Components of Plant Disease Epidemics. Jeger, M.J., Editor. (Reference 3).

| | | | |
|---|---|---|---|
| 8.4 | 8 | 6 | 49 |
| 19 | 9 | 50 | 136 |
| 10 | 48 | 81 | 23 |
| 40 | 62 | 11 | 8 |

This is certainly a simple and direct way to visualize the spatial changes of the concentration of fungi in soil samples.

Contiguous quadrats may be irregularly to yield more significant results. Consider the following example of a quadrats oriented along a river. :

**Figure 2: Contiguous Quadrats for River Pollutants**



In the above case, sampling could be done with a line transect which simply follows the river would

Quadrat sampling grids can also be made to conform to geographic, demographic or cultural boundaries. For example, suppose the drinking habits of residents were being analyzed, and quadrat sampling extended beyond the region into areas where alcohol was illegal. Certainly, the quadrats straddling the regional border would have statistics representative of neither region. A distorted visual impression would be given of less alcohol consumption near the border. And yet, nothing had been calculated incorrectly. The distortion occurs in the methodology.

## RANDOM QUADRATS AS USED IN FLORIDA FIELD STUDY

The description of random quadrats begin with:

> Effects of plant density and cultivar susceptibility on disease incidence and severity were examined via a repeated stochastic sampling of square quadrats of 0.25, 0.5, 1.0, 1.5, and 2.0 $km^2$ areas, respectively. Sites D1 and D2 were utilized because they represented large study areas where no disease trees were removed during the study

Infected trees in the Broward sites were cut down as they were discovered. Cutting of healthy trees was temporarily suspended until June 1999. After this period, cutting of healthy trees within the 125 ft circles was re-initiated.

When the results are discussed, the author state that the 0.25 $km^2$ quadrat size was used:

> Of the various quadrat sizes tested, the 0.25-$km^2$ quadrat size resulted in the clearest relationship among those variables and indices examined, and therefore, was used for all further analyses. The same data trends were found with both methods, indicating no unique outcomes associated with the use of stochastic overlapping quadratization, and led to its application for all further comparisons.

This could also be said, that any way the analysis was done, the same poor relationships would have resulted. A quadrat size of 2.0 $km^2$ seems extreme, as the sites D1 and D2 are 10.3 and 5.3 $km^2$ respectively. In Appendix A, it was concluded that there is evidence that only the northern half of Site D1 was used at least for some of the analysis, so D1 site may be 5.15 $km^2$.

The quadrat size of 0.25 $km^2$ would equal 0.0965 $mi^2$ or 61 acres. If an area is strictly residential area and ¼ acre lots appear to be very common, each quadrat would be approximately 240 residential lots. Conversion factors are given at the end of this Appendix. For quadrats lacking in citrus trees or too few trees, the authors state:

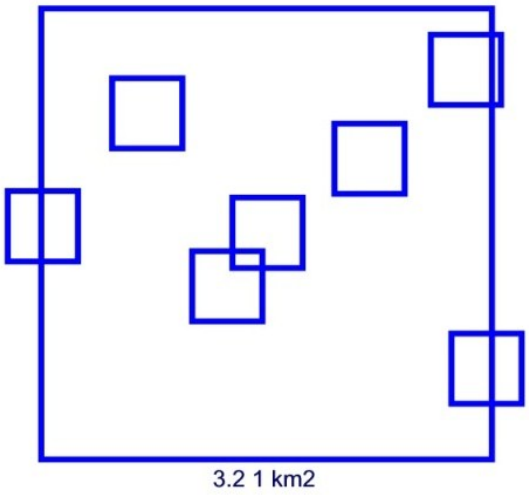
> A criterion was imposed such that only quadrats containing citrus trees and only quadrats with at least two diseased trees were accepted as part of the 500 quadrat population.

This suggests that the quadrat population is thinned, and may be less than 500. The alternative would be to discard and continue sampling until there are 500 acceptable quadrat samples.

As written, both criteria must be satisfied. If there are within a sample, 100 healthy trees, and 1 is infected, then this quadrat sample is rejected, while the same sample but with 2 infected trees is accepted.

The random location of quadrats would be determined by the center of the square. It is very likely that the quadrats would extend beyond the boundary. An example is shown on the following figure for Site D1, with 10.3 $km^2$ (3.21 km on a side).

**Figure 3: Randomized Quadrats, 0.25-$km^2$ in size in a 10.3 $km^2$ area**




The published article states the reason for not using contiguous quadrats, as follows:

> The appropriateness of the stochastic quadratization method was evaluated against a simple nonoverlapping quadratization method, the latter of which resulted in far fewer quadrats for comparison.

This is certainly true, since with randomize quadrats, there is really no limit to the number of values that can be generated. As more quadrat samples are done, and then mapped as the point estimate of an attribute at the center of the square, this becomes a form of spatial interpolation and a surface is created. With more samples, more squares are likely to straddle the boundaries where there are no infected trees and sample data in "non-citrus areas" including houses, lakes, parks and roads.

If a non-overlapping (contiguous) grid were used, a 7 x 7 pattern (49 quadrats) would be sufficient for Site D1 with 0.21 $km^2$ quadrats. For Site D2, a rectangular pattern of 7 x 4 (28 grids) could be used with 0.18 $km^2$ quadrats.

The expected number of citrus trees (healthy and infected) and infected trees likely to be contained in each square if all squares were completely inside of the boundary, is as follow:

**Table 1: Estimated mean citrus trees (healthy and infected) and infected trees**

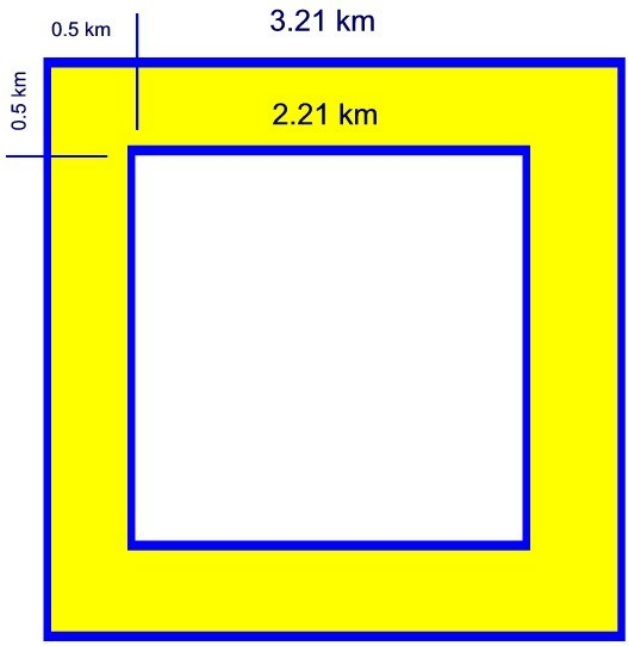
| Sites | Area (sq km) | Number of Citrus trees | Number of Infected Trees | Citrus trees/0.25 $km^2$ quadrat | Infected trees/0.25 $km^2$ quadrat |
|---|---|---|---|---|---|
| D1 | 10.3 | 6056 | 1758 | 146 | 43 |
| D2 | 5.2 | 6072 | 971 | 292 | 47 |

The same trees would be sampled many times in different quadrats. After 500 realizations, it is estimated that 73,000 citrus trees and 24,500 infected trees for Site D1 were sampled. For D2, 146,000 citrus tree and 23,500 citrus trees would be sampled. Thus, the information from the citrus trees would be used in multiple samples.

The space occupied by infected trees is over-estimated as the quadrats would extend outside of the study sites. If non-overlapping (contiguous) quadrats were used, the information from each citrus tree would be used only once.

If the center of the square lands within 0.5 km of the boundary, it would extend outside of the study site, it would extend outside of the study site. The likely occurrences of this event for a 10.3 $km^2$ study site area is the area of the perimeter region, shown in yellow divided by 10.3 $km^2$. This ratio is 5.42/10.3 or 52.6%. Thus, on the average, of the 500 quadrats, 263 quadrats would cross the boundary.

**Figure 4: Perimeter Region**




.

If we consider study site D2, the ratio of the annular region area to the study site area is 3.56/5.2 or 68%. So, on the average, of the 500 quadrats, 342 quadrats would cross the boundary.

## KNOWN INFORMATION

The published articles and in all other presentations, there is no information provided on the density of hosts or infected trees. The following was calculated based on the published article.

**Table 5: Known Information**

| Sites | Area (sq mi) | Area (acres) | Area (sq km) | Number of Citrus trees | Number of Infected Trees | Infected/Total | Density** (trees/ac) |
|---|---|---|---|---|---|---|---|
| D1* | 4 | 2560 | 10.3 | 6056 | 1758 | 0.29 | 2.37 |
| D2 | 2 | 1280 | 5.2 | 6072 | 971 | 0.16 | 4.73 |
| D3* | 1 to 3.14 | 640 to 2010 | 2.59 to 8.13 | 798 | 26 | 0.03 | 0.39 to 1.24 |
| B1 | 6 | 3848 | 2.6 | 4730 | 450 | 0.10 | 1.22 |
| B2 | 1 | 640 | 15.5 | 1113 | 229 | 0.21 | 1.74 |

* Conflicting data given for sites, see Appendix A. Sites D1 and D2 areas are given as approximate areas. Site D3 area not given, but calculated a complete circle, with a radius of 1 mile per published article. Site D1 area is inconsistent with other presentations, and with Figure 7 in published article. Other discrepancies exist as noted in Appendix A.

The differences in host density are never presented. The lowest host density is Site B1 or D3 (ambiguity of site areas is a problem).

## NORMALIZATION

Normalization is a linear transformation of variable, where the domain of the variable changes from [a,b] to [0,1], by use of the equation $X^* = (X - a)/(b - a)$, where $X^*$ is the normalized variable. At the collective level, normalization would have no effect on correlation or regression. If 500 values are transformed from their quadrat average value of X to $X^*$, exactly the same correlation coefficient and F statistic of regression would be calculated. The graphs would appear identical, just scaled differently. The coefficients of regression would be different, but if the original domain of X is known, then then the equations' coefficients could be recalculated.

But, the lower level of quadrat sampling, normalization would generally not make much sense. For example, in one quadrat, 10 trees have heights varying from 2 to 30 ft. On a normalized basis, the heights are now 0 to 1. In the next quadrat, the range of heights might vary from 2 to 6 ft. On a normalized basis, the range of heights are now 0 to 1. So, these two samples can not be compared. It would be the same problem if some heights are measured in meters and some in feet. Of course, if indexation was done, with a uniform minimum and maximum for all samples, then the results are comparable.

Examining Figure 6, there are four variables in total, two dependent and two dependent ones:

**Table 6: Variables used in correlations**

| Dependent variables (y axis) | Independent variables (x axis) |
|---|---|
| **Index of severity** - based on infected trees only, includes height and portion of canopy infected. | **Index of susceptibility** - Calculations based on susceptibility rating for each cultivar. |
| **Quadrat disease Incidence** - not described, but likely based on the ratio of infected trees to all trees. | **Index of host density** - based on number of citrus trees in each quadrat. |

The dependent variables will be reviewed followed by the independent variables.

## DEPENDENT VARIABLES

**--- Index of Severity:**

Per the published article:

> The index of disease severity was calculated as:
>
> $$I_{sev} = \left( \sum_{i=1}^{t} (\mathrm{DSB}_i \times h_i \times q_i) \right) / n \qquad i = 0, 1, \ldots\ldots t, \qquad (7)$$
>
> where *n* = the total number of quadrats in the study area, DSB = the binary disease status of each tree (0,1), *h* = the normalized height of the tree, and *q* = the proportion of sectors infected for each tree. This index represents a composite estimation of disease severity across all trees in a given quadrat and takes into account tree size and the relative volume of the canopy expressing disease.

This equation applies to each quadrat. It is assumed the binary disease status would be "1" for a diseased tree and "0" for a diseased tree. For a healthy trees, the numerator is zero. The number of trees in the quadrat is *t.* The summation sign notation is i = 1 to t (it may be a bit hard to read). The notation, *i = 0, 1, ... t* seems out of place, and may be a typographical error. Due to the random sampling, one quadrat sample may have many trees, while another quadrat may extend over non-citrus area or beyond the limits of the site and contain very few trees.

The problem with this index is the normalization of heights at the quadrat sampling level. This was explained in the prior section. Consider an extreme case, of two diseased trees within a quadrat. The first is the minimum height, so it has an index value of zero. The second one is the maximum height, so it has a index value of one. The minimum and maximum heights will be different in each quadrat sample, so indexation of heights will result in inconsistent comparisons.

The number of *Isev* values calculated would be 500 minus those values discarded because they contained less than one citrus trees or two infected trees. Once all the *Isev* values are calculated, there is apparently a second normalization (see discussion below Figure 6). This does not alter the data analysis as it occurs at the collective level. Also, the division by the number of quadrats seems superfluous, as it would be a constant value.

**--- Quadrat Disease Incidence:**

No calculation procedure is given, however it is assumed that this index is the number of infected trees divided by the total trees in the quadrat. This would be the standard definition with plant disease epidemiology.

**Figure 5: Quadrat disease Incidences (Figures 6D, 6E , 6I, 6H of published article)**

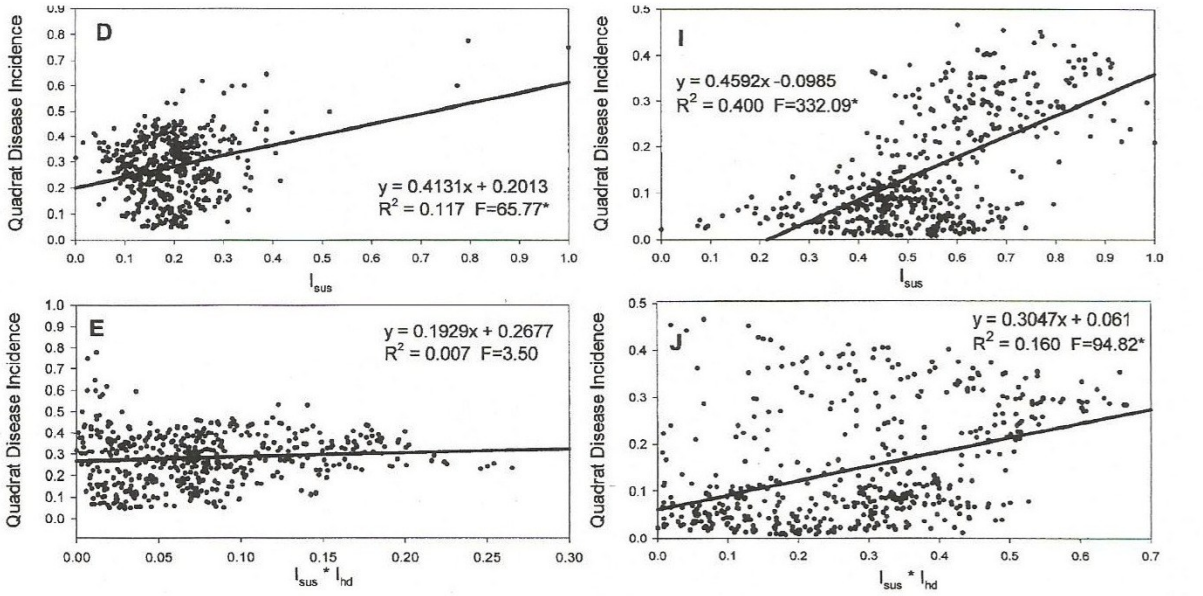



**Fig. 6.** The effect of citrus host plant density and cultivar/species susceptibility on disease incidence and severity for site D1 (**A to E**) and D2 (**F to J**) based on stochastic placement of 500, 0.25-km$^2$ centroids. **A and F,** The effect of host density on disease severity. **B and G,** The relationship of the normalized index of susceptibility and disease severity. **C and H,** The relationship of the normalized index of susceptibility–normalized host density and normalized index of disease severity. **D and I,** The relationship of the normalized index of susceptibility–normalized host density and quadrat disease incidence. **E and J,** The relationship of the normalized index of susceptibility–normalized host density and disease incidence. Calculations are described in equations 6 to 8.

The problem of quadrats extending beyond the border or into non-citrus areas would not affect this calculation as both infected and healthy trees would be excluded. The number of trees in each quadrat would be highly variable due to quadrat sampling across study site limits and into non-citrus areas

The ratio of diseased trees to total trees is 0.29 for Site D1 (Table 2). This is consistent with Figure 6E, which shows a regression line intercept of 0.2677 and a y value of 0.33 at x = 0.33. The ratio of disease tree to total trees is 0.16 for Site D2 (Table 2). This is consistent with the regression analysis as shown in Figure 6J.

## INDEPENDENT VARIABLES

### --- Index of Susceptibility:

Within the methods section, the authors state:

> During data collection surveys, citrus species/cultivar were assigned to 13 categories. Each category was given a susceptibility rating (0 to 6) based on a combination of prior published studies of susceptibility/resistance (7,14,18,20,21,23). Subsequently, an alternative susceptibility rating was utilized that was based on the actual disease incidence of each of the 13 species cultivar categories within each respective site. Trees were also assigned four height categories and their canopies divided into 12 sectors (north, east, south, and west and top, middle, and bottom). The following three indices were calculated for each tree and average indices for each quadrat. The index of disease susceptibility was calculated as:

$$I_{sus} = \left( \sum_{i=1}^{t} c_i \right) / t \qquad i = 0, 1, \ldots . t, \qquad (6)$$

> where $c$ = the normalized cultivar susceptibility rating for each tree and $t$ = the number of trees in each quadrat. The index represents a composite estimation of susceptibility based on all trees and accounted for the diversity of cultivar mixture within a given quadrat.

It is assumed that the sentence beginning with “Trees were assigned four height categories” is not related to susceptibility index. Thus, two rating systems are defined, one based on published studies and an alternative system based on actual incidences of infection. On page 271 of the published article (1), it is stated:

> Susceptibility was best expressed as a function of proportion of diseased individuals in each cultivar/species category that became infected relative to each site (data not shown), and this index of susceptibility was used for all further comparisons.

Since prior to this statement, the only comparisons discussed were in results in Figures 6A and 6F, and did not involve the susceptibility index, it is assumed susceptibility ratings are based on actual disease incidences in each site. The other method, using prior published studies, is assumed not used.

Stepping through the procedure, consider a quadrat sample with 6 trees (t = 6) with each having an assigned susceptibility rating based on disease incidences. Suppose we have two grapefruit trees, with a top susceptibility rating, and four orange trees with a lower rating. With normalization, the 2 grapefruit trees will receive a “1” and the 4 orange trees will receive a “0”, resulting in a total score of 2. This would be divided by the number in the sample, resulting in a score of 2/6 = 0.30. Now, if there was another tree in the sample, with even a lower susceptibility rating, the normalized rating of orange trees would go up, so we have 0 + 4*0.5 + 1 = 3 and higher normalized susceptibility would go up.

This example illustrates why the susceptibility index has, through the procedure, not useful in making comparisons.

In the figures below, the results for Site D1 and D2 are shown. For purposes of comparison, Figures 6B and 6D represent Site D1.. Since they are very similar, only Figure 6D is shown. Figures 6G and 6I represent Site D2, and are similar.

Figure 5: Site D1, Susceptibility vs Disease Incidences (dotted line added)

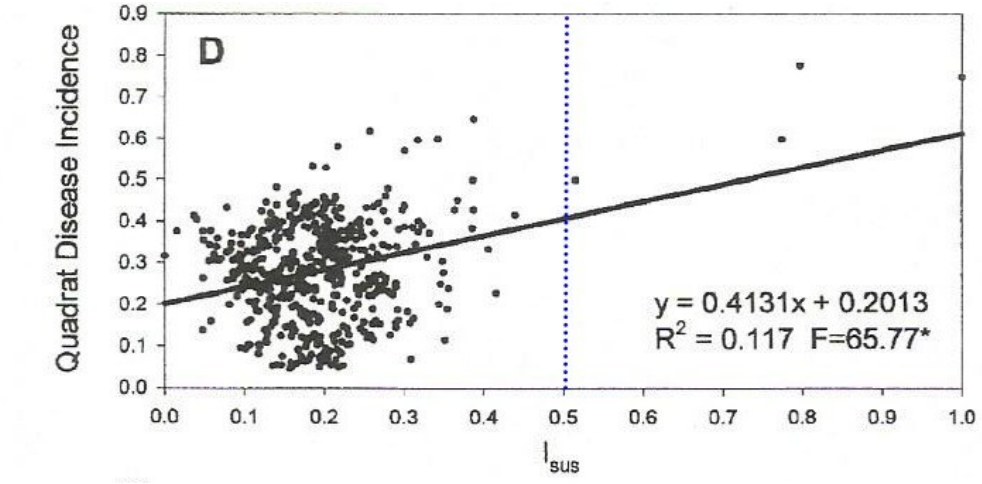
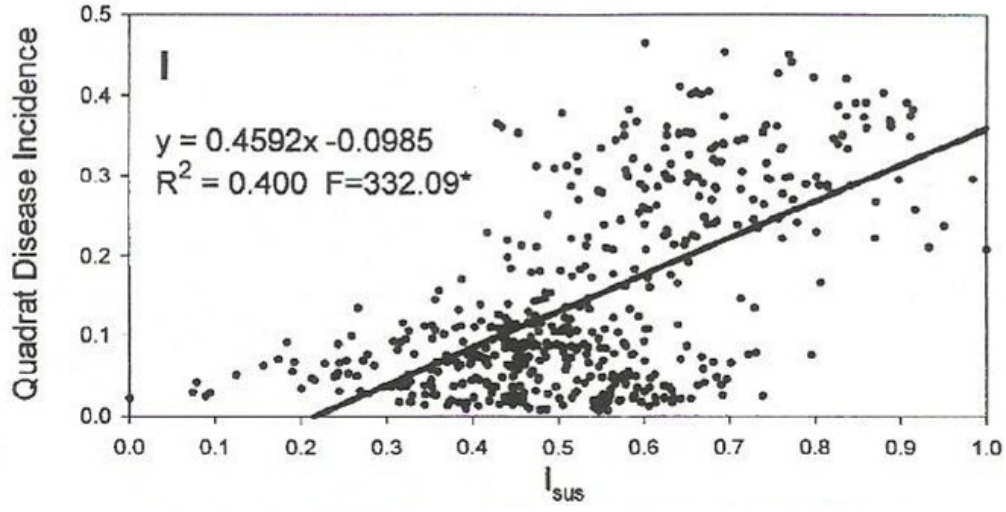



Figure 6: Site D2 Susceptibility vs Disease Incidences




Note the distribution of points in Site D1 are all contained within $I_{sus}$ = 0 to 0.5 with the exception of four points with $I_{sus}$ greater than 0.50. For Site D2, the $I_{sus}$ points cover the full range of the scale, with more points greater than 0.5. It is believed that the methodology for calculating the susceptibility index contributed to the difference between Site D1 and D2, The

use of normalization on the susceptibility ratings in effect, “dirtied” the data, by making the ratings dependent on the minimum and maximum ratings within the sample.

The quadrat disease incidences is informative as it shows that in many areas, 10% or fewer trees became infected by the end of the study. Of course contiguous quadrats would have been more informative as one could identify the general variations in disease intensity.

The authors could have provided the 13 categories of cultivar/ species, and information of how prevalent each of these categories are in each site. Also, it would be informative if the disease incidences for each of the 13 categories of cultivar/ species was given.

**--- Index of Host Density**

The authors state:

> The index of host density was calculated as:
>
> $$I_{hd} = t/(\mathrm{QS}) \qquad (8)$$
>
> :
>
> where QS = quadrat size in square kilometers. This index represents a calculation of the citrus tree population saturation (relative to the quadrat with the highest density) of a given quadrat. Each of the above indices was normalized, resulting in values from 0 to 1 for each quadrat sampled.

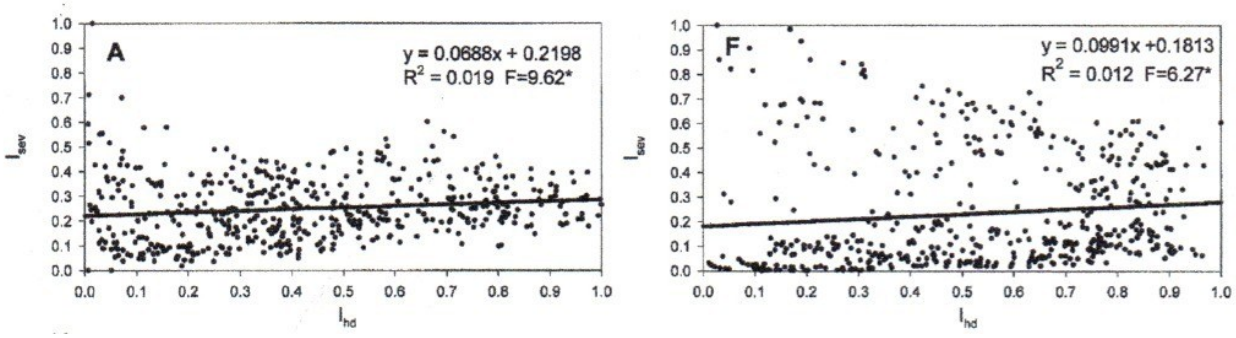
In prior discussion, the variable *t* is identified as the number of trees. The quadrat size is a constant (0.25 $km^2$). Normalization is unnecessary and makes comparison between sites impossible. As calculated before normalization, the host tree density would be in terms of number of citrus per 0.25 $km^2$, certainly information worth knowing to the regulators of the eradication program. However, once normalized, it is not possible to obtain actual densities.

## CORRELATION RESULTS

Normalization at the quadrat sampling levels for indexes of susceptibility and severity makes comparisons invalid. Quadrat disease incidences appears to be calculated correctly. Since in every graph in Figure 6, contains at least one variables which was normalized at the quadrat level, the comparisons are meaningless.

The graph below shows the correlation of index of host density verses index of severity for Sites D1 (Figure 6A) and D2 (Figure 6F).

**Figure 6: Correlation Results with host density as independent variable**




The authors state in the published article:

> Linear regression of disease incidence versus host plant density resulted in low coefficients of regression and slightly positive slopes for both D1 and D2, indicating little or no effect of host plant density on disease incidence. However, the variance associated with the relationship of disease incidence to host plant density decreased as density increased. That is, at lower host densities, disease incidence was more variable and variability decreased with increased host density. This decrease was more apparent for site D1 than for D2. Similarly, linear regression of disease severity (the proportion of diseased sectors of individual trees) versus host density indicated a slight positive slope for both sites, indicating little or no effect of host plant density on disease severity (Fig. 6A and F). The associated variance for this relationship also decreased as density increased. Linear regression of disease incidence versus the index of disease severity resulted in a positive slope and high $r^2$ values for both sites, indicating that much of the variation due to regression was accounted for and indicated a direct relationship between increasing disease incidence and increasing disease severity, as expected.

A correlation coefficients of 0.019 and 0.012, would typically be interpreted as no linear relationship exists between these two variables. A hypothesis test can be done on the calculated coefficient value with the null hypothesis of zero correlation.

# KRIGING ESTIMATE MAPS OF INDEX OF SEVERITY

*"Once a map is drawn, people tend to accept it as reality"* Bert Friesen, as quoted in reference x (Chiles and Definer).

## METHODOLOGY AND PRESENTED RESULTS

In the published article under Methods and Materials section, the authors state:

> Index of disease severity values were also used to perform a semivariance analysis followed by a kriging of the data by the block method to visualize the occurrence and position of foci and the development and spread of disease through time. Kriging was performed at four time periods selected to best represent periods following significant increases in disease.

Further, the authors state in the Results section:

> Kriging of the *Isev* demonstrated the occurrence and development of foci of disease in each of the urban areas (Fig. 7). For each of the study sites, foci of infection that established early in the study can be seen. As these foci continued to increase in severity and spread locally, additional foci began to appear. The effect of previously established foci on the establishment and evolution of new foci was seen. It was noted that early in the epidemic foci often became established at considerable distance from each other. These foci continued to enlarge while simultaneously new secondary foci began to fill in the previously uninfected areas between the original foci.

Similar statements of the filling in process are presented in the Discussion section of the published article and in the Abstract.

The maps were generated using 500 values from quadrat sampling. Each value of Isev corresponds to the center of the quadrat.

In Figure 7, there are a series of maps, identified as "Kriging estimates maps." Our preference is to refer to them as contour maps based on kriging interpolation. The parameters resulting from a semi-variance fit to a standard curve are a part of kriging method of interpolation. Further, the authors mention the "block method" which is typically identified as "kriging block method." This analysis should not be confused with the semi-variance analysis as reviewed in Appendix E.

Mapping can be subjective when sparse data are involved. Uncertainty or random noise in Z can make contouring more difficult. Additional discussion may be provided on this topic in a later appendix.

The maps presented by the authors in Figure 7 are colored in between contour lines. A constant color would be predict an area where the Z values would likely fall in a particular range. For example, if a contour map shows elevation, a constant color between the 100 and 150 ft contours would show where the elevation was likely to be in the range of 100 to 150 ft. However, the colors represent somewhat different values depending on the study site. For

instance, light blue represents Isev from 0.025 to 0.050 for Site D1 while for Site D2, the same color represent Isev from 0.035 to 0.070. It is assumed for Site D1, the color white represents valued from 0 to 0.025. The difference in color scale may affect the areas in white, but unlikey change the overall appearance.

The data used in the computing the maps is based on the DNC parsing method, as described in Appendix A. For time periods were selected, T50 (150 days), T10 (365 days), T14 (420 days) and T18 (540 days).

## STUDY SITE AREAS

The maps in Figure 7 provide estimates of the size of the study site. There are significant discrepancies between Figure 7 and the stated area within the text of the published article as follows:

Table 3: Comparison of Area as given in Figure 7 and Text of Published Article

| Site | Delta X (km) | Delta Y (km) | Figure 7 Area ($km^2$) | Figure 7 Area (sq miles) | Per Text Area (sq miles) |
|---|---|---|---|---|---|
| D1 | 3.00 | 1.80 | 5.40 | 2.08 | 4.00 |
| D2 | 2.00 | 2.50 | 5.00 | 1.93 | 2.00 |
| D3 | 1.20 | 1.20 | 1.44 | 0.56 | 3.14 |
| B1 | 4.70 | 1.70 | 7.99 | 3.08 | 1.00 |
| B2 | 1.00 | 1.10 | 1.76 | 0.68 | 6.00 |

It is reasonable to assume that in the text of the published article, sites B1 and B2 were inadvertently reversed, and the figures are correct. Major differences occur in all sites except D2.

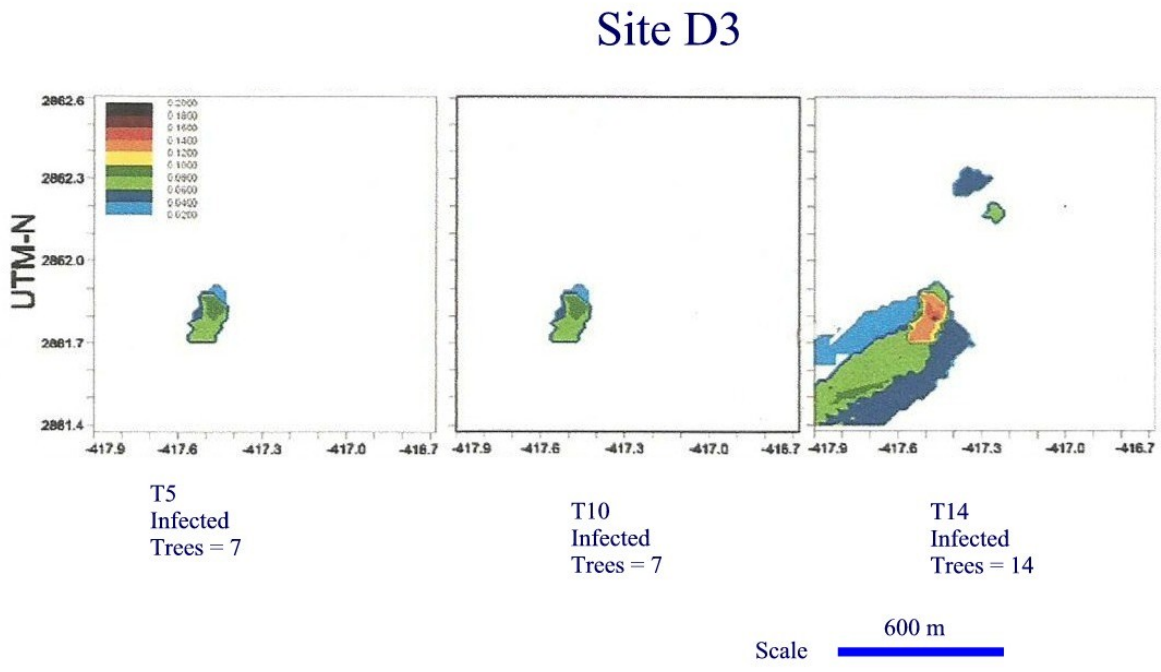
## REVIEW OF RESULTS

The underlying data used in the contour maps are generated using the 500 random quadrats, resulting in many overlapping squares. The results in a greatly exaggerate impression of the presence of citrus canker, if one views the maps in this manner.

If, for example, there exists a small group of infected trees, then the random quadrats will create many positive severity index values distant from this group of trees. The maximum distance between the quadrat centers and a small group of infected trees, all within a few meters of each other can be calculated. When infected trees are in the corner of the quadrat, the center is located at a distance of half the diagonal distance of the quadrat (352 m). If the trees are located at the center of one side of a quadrat, the the center of the quadrat is 250 m away from the infected trees. Thus, a large area can be created around a small grouping of trees, with quadrat centers located up to 250- 352 m from the small grouping of trees. A 250 m circle would encompass 49 acres. Since many of the houses are on ¼ acre spacing, this would be approximately 200 houses, if the area is completely residential.

The map of Side D3 shows an increase in "contoured area" as the number of infected trees increased from 7 to 14. The areal change is approximately 300 x 600 m, or 180,000 $m^2$. This equates to 44 acres. If we consider any lot with an infected tree to be affected, and one infected tree per lot, then perhaps 2- 3 acres would be considered affected.

Figure 7: Site D3, from published article, with scale and time period added (Figure 7 in published article)

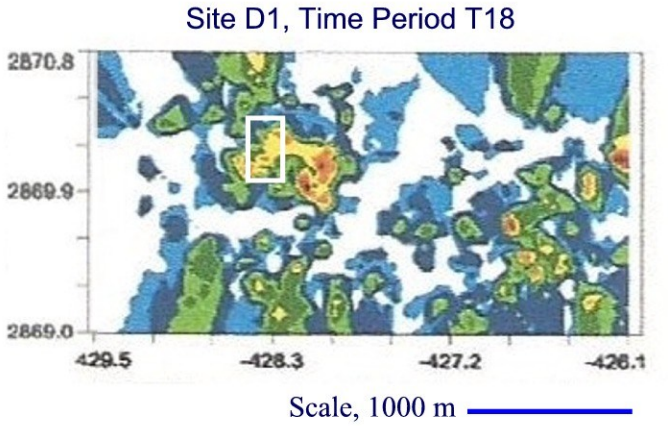


As reviewed in Appendix A, each site contains numerous “non-citrus” areas, including lakes, canals, parks, school yards, etc. The largest lake in Site D1, is a man-made lake, which is bounded by 192 St in Coral City, to the south. Coral City is a municipality within Miami-Dade County.

The outline of the lake is shown on the map as outlined in white as being within the contoured area with colors of orange, yellow, green and blue, representing all levels of the index of severity.

Obviously, it is not the infected trees which are filling in these areas, but a product of the random quadrats and kriging routine.

Figure 8: Site D1, Time Period T18, with location of lake at 192 Street, Carol City, Florida




The authors also indicate the appearance of many smaller circles or “disease focal points.” In development of contour maps, these are often considered undesirable anomalies, called bulleyes, and can be eliminated by adjustment of the nugget value of the fitted semi-variance curve. There are many alternative routines in contouring to smooth out the data to remove localized highs or lows.

## SUMMARY

The more important findings are summarized below:

1) The statistics reviewed in this appendix are based on random quadrat sampling. This type of sampling is generally appropriate to large populations where sampling is the only practical means to obtain unbiased estimates of variables. Random quadrat sampling generally is not done in areas that have been completely surveyed numerous times.

2) Given the geometry and area of Sites D1 and D2, with 0.25-$km^2$ random quadrats, approximately half of the quadrats will extend beyond the boundaries of the site. Also, the quadrats likely extend into many non-citrus areas as well.

3) As described in the article, normalization of variables at the quadrat sampling level used the minimum and maximum of the sample and not of the population, making these statistics non-comparable. The severity and susceptibility indices considered to be non-comparable from sample to sample as follows:.

| | |
|---|---|
| Index of Severity: | Heights of trees normalized within sample |
| Index of susceptibility: | Cultivar ratings normalized within sample |

The likely result is to introduce more variability into the calculated averages. This could lower the correlation coefficient. In any case, the sample averages can not be compared, and the correlation analysis is invalid.

4) The sizes of the study sites are inconsistent with the information provided in the text of th3 2002 published article. This problem was reviewed in Appendix A.

5) The maps based on kriging and semi-variance analysis requires a variable whose presence is continuous throughout the the area. The areas within the study sites where citrus canker can be present is highly fragmented as many non-citrus areas exist in every site. Both the random quadrat sampling routine and kriging are responsible for filling in areas without infected citrus trees, and is not capable of knowing that canker can not exist on roads, lakes, etc.

6) Localized high values were inferred in the published article as focal points. These points are likely the product of the nugget value of the semi-variance curve used by the kriging algorithm.

## CONCLUDING REMARKS

Quadrat analysis and related statistics as presented in the published article, may be considered to have two faces, a superficial face where all appears normal and documented, and a second face, coming from detailed investigation where nothing is right. The final result, as shown in the contour maps is bizarre as it shows incidences of citrus canker extending over parking lot, lakes, schools, shopping centers and other areas. Is this a joke?

But, our review is bottoms up, beginning with a detailed look at methodology. Granted, there are many transformations and clean up techniques in statistical data analyses, to make the raw data more useful. Anomalous points or outliers can be identified and removed, to improve relationships. Normalization is an important tool when used it a proper manner.

For the index of severity, the normalization of tree heights within the quadrat sample has nothing to do with data clean up or making the data more useful. It actually transforms useful information into useless information for comparison purposes. Tree heights in residential lots obviously vary considerably, likely in the range of 4 to 30 ft. Quadrats with less than two infected trees were rejected, which only sets a lower limit on the sample size range. It would be impossible to know how many diseased trees were in each quadrat sample, as the quadrats would sample many non-citrus areas and extend beyond the study site boundaries. Normalizing the heights within the sample, renders the index statistics on severity, host density and susceptibility non-comparable. A normalized height of 1 could be a 4 ft tree or 25 ft.

It is also noted that the index of susceptibility was also “dirtied” by the procedure to normalize the susceptibility ratings of the sample obtain with the quadrat. This results in all 10 plots in Figure 6 using variables which are transformed into non-comparable variables.

Why would anyone do this? It is extremely strange scientists to intentionally “dirtied” their data and make it useless for analyses. It is stated this was done intentionally, because at least two authors, Drs. Gottwald and Ferrandino are expert epidemiologists, and this was done on three sets of data, or 1500 sample sets. Since Dr. Ferrandino did not author epidemiology articles before or after this study on citrus canker, it is likely that Dr. Gottwald was solely responsible for the quadrat analyses.

It is suggested that as the article was being reviewed for publication, there was little attention given to the calculation of indexes. Why should one go through the methodology in detail, when ultimately only poor correlations resulted?

Scientists are often accused attempting to make data look better than it really is, through presenting partial sets of data and not presenting all details. No such accusation could possibly be considered in the case of the Florida field study. There was no clean up.

What was the purpose of the correlation analyses? Why the lengthy discussion between variables with regression correlation coefficients ($R^2$) of less than 0.02? Isn't this a bit odd? It also was odd that the relatively simple statistics seemed absent from this discussion. We know that in Site D3, only 3% of the trees became infected, while in Site D1 there was 29% infection rate. Why?

Certainly, by the manner in which the indexes were calculated with random number generation, it would be impossible to relate them back to the original collected data. If contiguous quadrats had been used, and normalization not done, then there would have been a direct connection between the collected data and the index values.

The random quadrats were not included in the Oct 1999 interim report, June 2000 presentation (transcript only) and the November 2000 court view graphs. However, the procedure is believed to be very much a part of what the Department considers the "Epidemiology Study" of which we believe the field study is one part. This is discussed further in Chapter 8.

# REFERENCES

1. Campbell, C.L, and Madden, L.V., 1990, Introduction to Plant Disease Epidemiology, John Wiley and Sons, New York.

2. 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.

3. Jeger, M.J, (Editor), 1989, Spatial Components of Plant Disease Epidemics, Prentice Hall, ISBN 0-13-824491-X. (Chapter 4, World Distribution of Soilborne Mycoparasites)

4. Madden, L.V., Hughes, G., and van den Bosch, F., 2007, The Study of Plant Disease Epidemics, American Phytopathology Society, Minnesota. ISDN 978-0-89054-354-2.

5. Schubert, T.S. , Gottwald, T.R., Rizvi, S.A., Graham, J.H., Sun, X., Dixon, W.N., 2001, Meeting the Challenge of Eradicating Citrus Canker in Florida- Again, Plant Disease, Vol. 85-4.