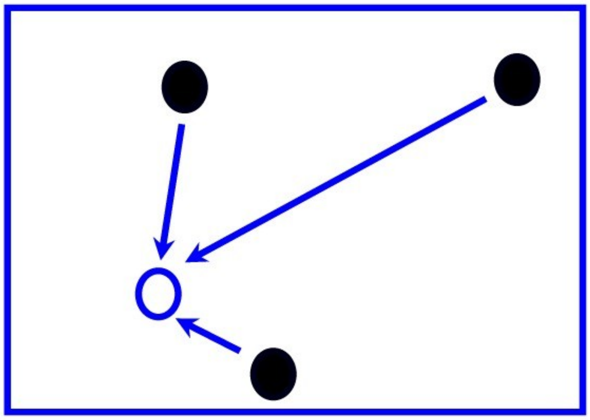
APPENDIX D

INTER-POINT DISTANCE ANALYSIS



Three potential source trees create 3 potential distances of transmission to a newly infected tree

INTRODUCTION

Inter-point distance analysis (IPDA) is a method of spatial point pattern analysis used in in the published article, but the method was not explicitly named. The technique has been referred to as point distance analysis in software programs (ArcGIS), as per reference 3 and inter-event distance (IED) analysis by Diggle (2).

This analysis is presented in the 2002 published article (1) on the Florida field study in the general discussion of "distances of spread." The distances as calculated by the DNC method, and presented in Appendix B/B1 were considered conservative estimates of distance of spread, because the nearest prior infected (PI) tree was considered the source tree within a time period. The authors indicate this is a method which can be an upper bound or an overestimation of the actual distance of spread of citrus canker.

The discussion of this method and results are presented on page 365, 371 and 373 of the published article. The set of calculated distances would represent all possible inter-tree distance for a set of diseased trees within as study site. . Thus, if one considers each tree was infected as a result of windblown rain from any other tree within the site, then the IPD analysis provides all the possible distances.

Spatial point pattern analysis using IPDA is also discussed in Appendix F. The topic in this appendix is in reference to the frequency diagram shown in Figure 3 of the article.

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 1)

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 exclude important details, so 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 365, **Materials and Methods,** Right hand column

A third VBA was used to calculate the distance from focal trees to all possible subsequent infected trees. For ease of calculation, these distances were parsed into contiguous 30.5-m (100-ft) distance categories. This calculation represented an “overestimate” of spread but provided the longest or maximum distance of spread that could have occurred. It is unknown which focal trees gave rise to infections in subsequent infected trees. Thus, the actual distance of spread intuitively would be between these two extremes described above.

Page 371, **Notation under Figure 3**:

Fig. 3, Frequency distribution of occurrence of new infections relative to all pre-existing citrus canker infected trees to the nearest 30.5 m (100 ft) for **A**, site D1, **B,** site D2**, C** site D3**, D** site B1**, E,** site B2**.**

Page 373, **Results Section,** Left hand column

Figure 3 represents frequency distributions for each study site for all possible distances of spread from minimum to maximum. Peaks in this distribution that represented the most common distance categories calculated were 640, 457, 60, 518, 243, and 335 m (2,100, 1,499, 197, 1,699, 797, and 1,099 ft) for D1, D2, D2, D3, B1, and B2, respectively, based on 30-day intervals. Note that D2 was bimodal and had two equivalent high frequency peaks. Maximum possible distances of spread based on the data from each area were 3,444, 2,133, 914, 4,754, and 1,432 m (11,299, 6,998, 2,999, 15,597, and 4,698 ft) for D1, D2, D3, B1, and B2, respectively.

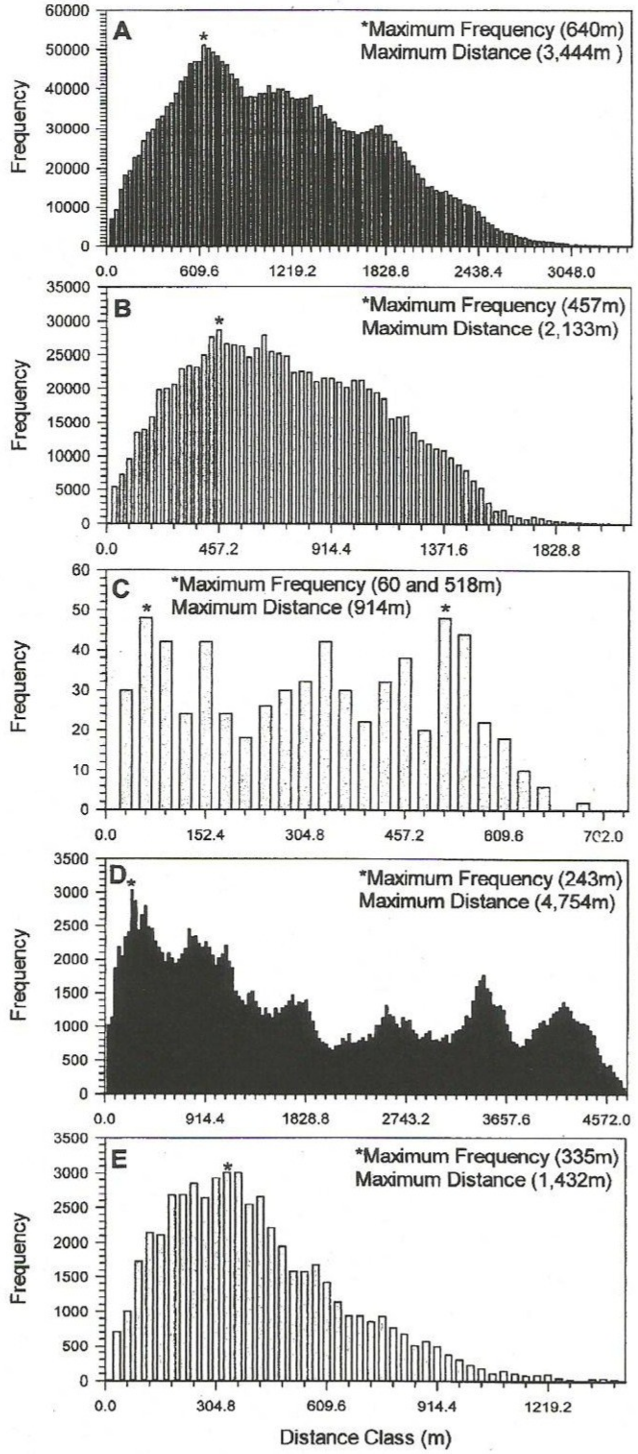


Fig. 3. Frequency distributions of occurrence of new infections relative to all possible preexisting citrus canker-infected trees to the nearest 30.5 m (100 ft) for **A,** site D1; **B,** site D2; **C,** site D3; **D,** site B1; and **E,** site B2. Distances represent the full range from minimum to maximum distances based on 30-day periods.

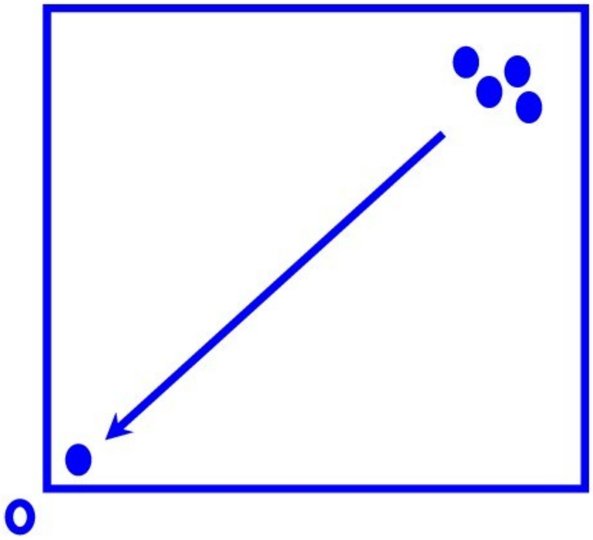
CALCULATION OF FREQUENCY DISTRIBUTIONS

The plots in Figure 3 of the article are constructed by calculating the frequency of occurrence based on class intervals. The class interval is 100 ft, and appear to begin at 15.2 m (50 ft) so the classes are likely [50- 150], [151 - 250], ... etc., if data are parsed to whole numbers. Peak frequency distances ranges from 197 to 2100 ft and the maximum distances range from 2999 to 15,597-ft.

CALCULATION EXAMPLE

Consider 5 infected trees in a site. Four of these trees are close together (i.e. < 30 ft) while a fifth tree is 2000 ft away. In a residential setting, this may be 10- 20 lots away.

Figure 1: Example of Inter-Point Analysis



If all infected trees are associated every other infected tree and duplicates excluded, inter-point distances calculated, there would be *n(n-1)/2* or (5 x 4)/2 = 10 distances calculated. Four of the distances would be approximately 500 ft, The other six distances would be approximately 30 ft. Therefore, the histogram would show about 40% of the distances were 500 ft, and 60% of the distances would be in the range of approximately 30 to 60 ft.

INTERPRETATION AND EQUIPROBABLE EVENTS

The authors do not refer to the frequency distribution as a histogram. In statistical analysis, histograms are often used to show the relative frequency of an attribute from a study. It may be inferred that relative frequencies are estimates of probability of events. Thus, statistical values, such as mean and mode values may be considered meaningful estimates in terms of a statistical description of the general population. The authors indicate the analysis was useful in determining a "non-conservative" estimate of the distance of spread of citrus canker. They make no claim that the frequency distribution is providing distances of spread for the general population.

The collected data in this case are the locations of infected trees. A set of $n(n-2)/2$ distance values result from connection of all infected trees and removing duplicates. Each distance represents an event of a successful infection from one tree to the next. If it is considered that each event has equal probability, then the normal statistical measures such as mean and mode are valid estimates of the distribution of distances.

Consider Site D1 with a maximum value of 3444 m (15,597-ft). The minimum distance is not given, but let us assume that it is 10-ft. If there are 5,000 calculated distances, then there is a a probability of 1/5000 that citrus canker can be transmitted 10-ft. But this is the same probability as 15597-ft. If there are 20 distance values between 10 to 20-ft and 20 distance values between 15557 and 15597-ft, then the calculated frequency (or expected probability) is 20/5,000 for the both short and long distance groups. The equiprobability assumption of each of the calculated distance results in unrealistic probabilities based on the known mechanics of disease dissemination.

Since equiprobable assumption is highly unrealistic, the peak values as shown in graph are not meaningful as modal estimates of distances of dissemination. The maximum values on the histogram are related to the diagonal distance in the site if the disease is scattered about the site. It is no accident that the largest distance (4754 m) comes from the largest site, B1, which is given as 6 square miles in the published article (1). For a widely dispersed disease, the larger the site, the longer the maximum distance. If a site were 10 miles by 10 miles square, the maximum distance would likely be close to the 14 miles, or the diagonal distance connecting the corners of the square.

Further investigation of the IPD pattern is provided in Appendix D1. The distance at which the maximum frequency occurs is theoretically equal to $0.38 \cdot D$, where D is the diagonal distance of the square, if the pattern is completely random and the site is a square. Problems with frequency counts are also identified in Appendix D1.

SUMMARY

1. Long distances (maximum distances) from the IPDA method were determined because the disease was widely dispersed within the study site by the end of the study. The maximum distances increase with increasing site areas, with the largest site, B1 having the highest value (4.8 km) , and the smallest site, D3, having the shortest maximum distance (0.9 km).

2. It should not be inferred that these distances relate to distances of spread. The distances are a function of the site area and geometry. If similar analysis was done on the State of Florida, it is likely the upper bound of canker spread would be hundreds of miles.

3. The maximum frequency value of the histogram would only be a valid descriptive statistic if equiprobable events could be assumed. However, this would be contrary to the basic mechanics of dispersal by windblown rain. A dissemination distance 50-ft away from an infected tree would have to be just as probable 15,594-ft.

4. The discussion of results refers to the graph as a frequency distribution, and the values as maximum frequency values. It does not refer to them as estimates of the mode of the distribution, so technically the discussion is correct, but it could be considered misleading.

5. Values are far beyond estimates from experimental studies. Travel distances of greater than 59 ft have not been shown to be possible in experimental studies. For rain water to first pass through the canopy of a tree then travel long distances, over buildings, lakes, parking lots, roads and other non-citrus area would seem impossible.

REFERENCES

1. 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.

2. Diggle, P.J. 2003, Statistical Analysis of Spatial Point Patterns, Oxford University Press, ISBN 0 340 74070 1.

3. The softtware ArcGIS is a product of ERSI, Incorporated.

More discussion can be found on Point distance analysis:

http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Point_Distance_(Analysis)