

Example:
Spatial and Spatio-temporal
Patterns of Yellow Crinkle
Disease in Papaya

Dixon, P.M. and Esker, P.
Department of Statistics,
Iowa State University

Spatial point patterns

- Locations of events in space or space x time
- Goals / questions include:
 - Visualizing probability of an event
 - Are events clustered?
 - Are events clustered with other types of events?
 - In space or space x time
- Illustrate using Papaya Phytoplasma data

Phytoplasmas in Papaya

- In Australia, three predominant types of economic importance:
 - Papaya dieback
 - Papaya mosaic
 - **Papaya yellow crinkle**
 - **Tomato big bud (TBB)**
 - **Sweet potato little leaf V4 (SPLL-V4)**
 - Visual scouting for symptoms
 - confirm / identify with PCR and RFLP

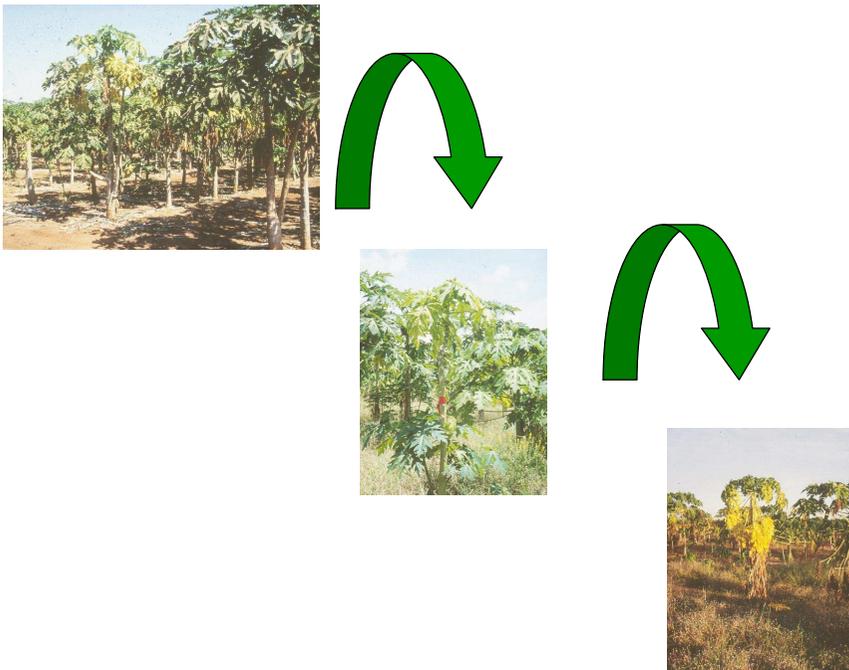
Notes

- Phytoplasmas: specialized bacteria lacking cell walls
 - Transmitted by sap feeding insects, e.g. leafhoppers, psyllids
 - Responsible for a variety of yellowing or wilting diseases in large number of plant species
- Identification usually by molecular methods, e.g. sequencing



Notes

- Data from papaya plantation in Northern Territory, Australia, outside Katherine Australia, 14 S, 132 E
- Location of Katherine within Australia



Notes

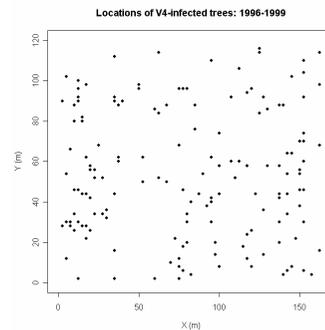
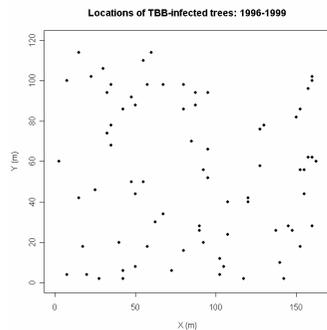
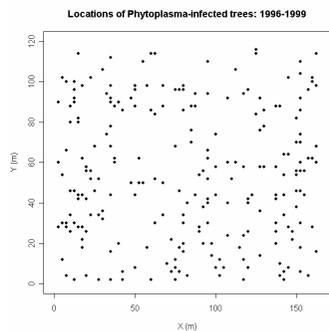
- Photos show uninfected and two stages of phytoplasma infection.
- Upper left photo: uninfected plants and the general layout of the plantation.
- Middle photo: papaya with early symptoms of yellow crinkle disease.
- Lower right photo: plant that is near death due to a phytoplasma infection.
- Infected plants remained in the plantation for this study.

Field Study: May 1996 – April 1999

- Padovan and Gibb (2001)
- Time, incubation to death: Esker et al. (2006)
- Plantation measured: 65 x 58 = ~ 3,800 plants
- Planted: January 1996
- Monthly census of all plants (began May 1996 through April 1999)
- Plants left in plantation until plant death (month of death then noted)
- 154 cases of V4 infection; 76 cases of TBB

Notes

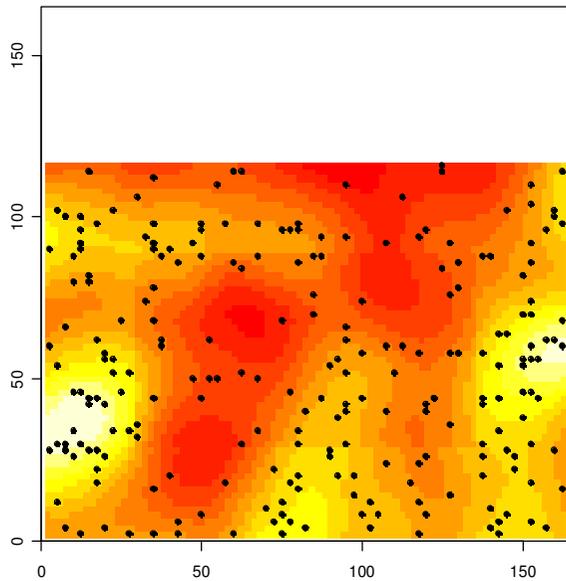
- Padovan, A.C. and Gibb, K. S. 2001. Epidemiology of phytoplasma diseases in papaya in Northern Australia. *J. Phytopathology* 149:649-658
- Esker, P.D., Gibb, K.S., Padovan, A., Dixon, P.M, and Nutter, F.W. Jr. 2006. Use of survival analysis to determine the postincubation time-to-death of papaya due to yellow crinkle disease in Australia. *Plant Disease* 90:102-107.



Notes

- These plots show locations of diseased plants. Time when symptoms first noticed is ignored.
- Is there any general trend in disease incidence? E.g.:
- Is disease more frequent in one part?, or on the edge?
- Answers not obvious
- Are locations of diseased plants clustered? I.e. are diseased trees surrounded by other diseased trees?
- Hard to tell from the plots of locations. Appear to be places with many and places with few, but the eye is easily fooled.
- Even harder to tell if there is any relationship between the two types

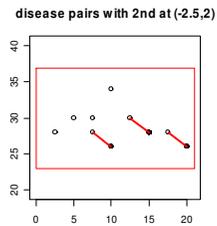
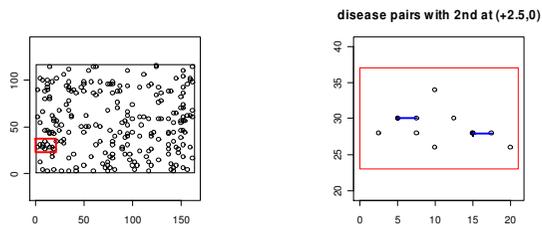
Disease locations and smoothed intensity



Notes

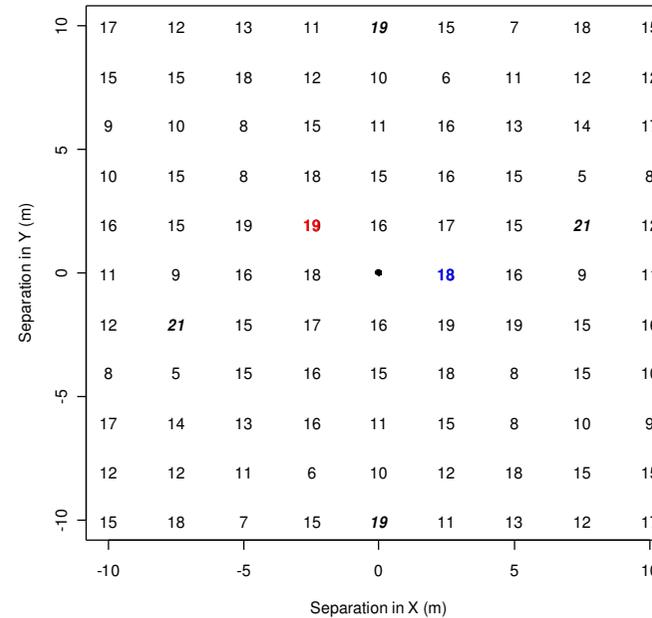
- Estimate intensity (average number per m^2) using a non-parametric kernel smoother
- On color plot, white = highest intensity, yellow = intermediate intensity, red = lowest intensity
- Black and white plot has contour lines to show the same thing.
- These plots show average number; don't look at relationships between points
- Clustering is a property of sets of points.
 - Clustering: neighborhood of a diseased tree is unusually likely to have another diseased tree
- When set of possible disease locations is a grid (e.g. trees in a plantation), can count distance-direction pairs: Look at each diseased location. Count number of diseased tree with another diseased tree 'next door' to the east, 'next door' to the north, ..., two trees away to the east, ..., one north and one east, ... Use all combinations of lag distance and direction pairs up to some maximum distance and you get the next plot.

Distance-Direction Pairs



Consider each diseased tree
 How many diseased trees are 2.5 m to right?
 2 in this small part of the data
 How many are 2.5m to left and 2m up?
 3 in this small part of the data.

Pairs of diseased trees separated by (x,y)

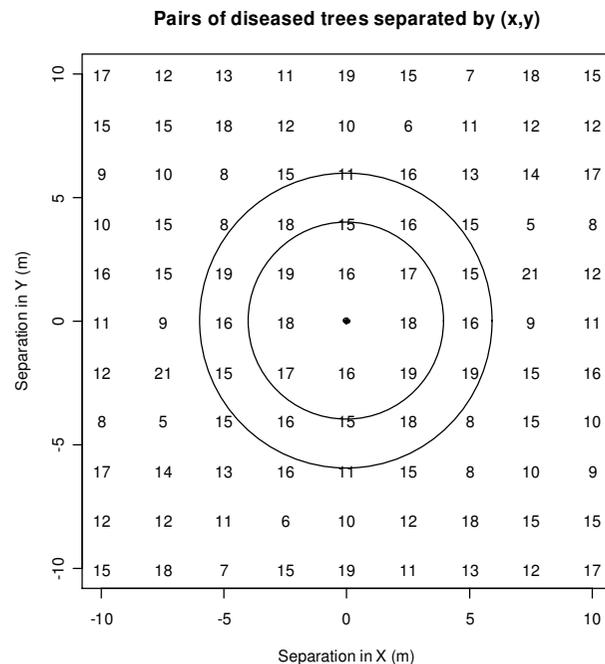


Notes

- Counts in red and blue are the full data set equivalents of the counts illustrated on previous slide
- Visually, no obvious pattern to the counts.
- Disease not more frequent among adjacent individuals or in one specific direction
- A very simple model: Complete Spatial Randomness
 - Locations of diseased trees are independent of each other (no clustering)
 - Probability that a tree is diseased is constant across the study area (no trend)
- Number of diseased individuals in any distance/direction count has Poisson (m) distribution
 - Ignoring edges: $m = n(n-1)/(t-1)$ $n = \# \text{ diseased trees, } t = \# \text{ trees}$
 - $m = 230(229)/3769 = 13.97$
 - $P[X > 20 | m = 13.86] = 0.047$
 - Should account for edge effects: fewer pairs of points separated by 5 trees than by 1 tree.
 - Values in bold italic are significantly larger than expected for that distance and direction

Distance / direction pairs

- Consider each pair of diseased tree
- Tabulate distance/direction between them
- Counts larger than expected (ca. 13)
 - But not unusually so
- Direction doesn't seem to matter
- Consider only distance to increase power



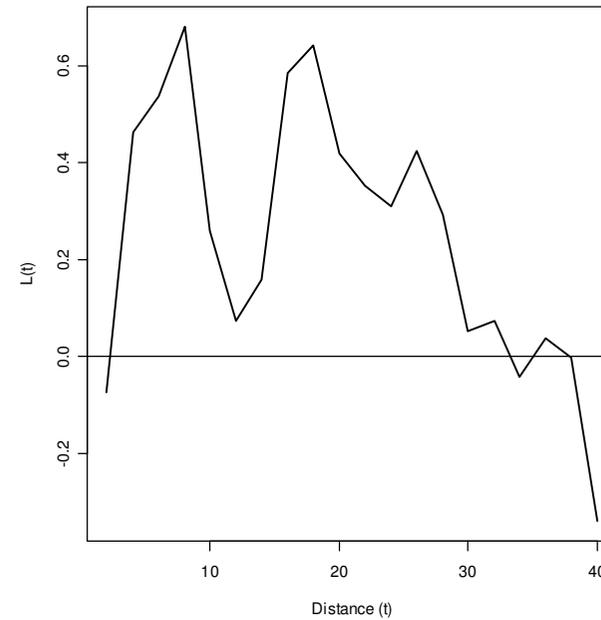
Ripley's K(t)

- Ripley's K(t) combines information across distances, ignores direction
- $K(t) = E \# \text{ add'n points w/i dist. } t / \text{intensity}$
- Often easier to work with $L(t) = \sqrt{K(t)/\pi} - t$
- Compare K(t) to πt^2 or L(t) to 0
- Clustering: $L(t) > 0$
- Segregation: $L(t) < 0$

Notes

- E is shorthand for Expected value of. This is the theoretical average.
- Estimate by counting number of points w/i distance t of each point, divide by intensity
- Estimate intensity as # points / total area
- Again, need to account for edge effects, details in many books and papers
- Under CSR: Complete Spatial Randomness, defines a few slides ago:
 - $K(t) = \pi t^2$
 - $L(t) = 0$

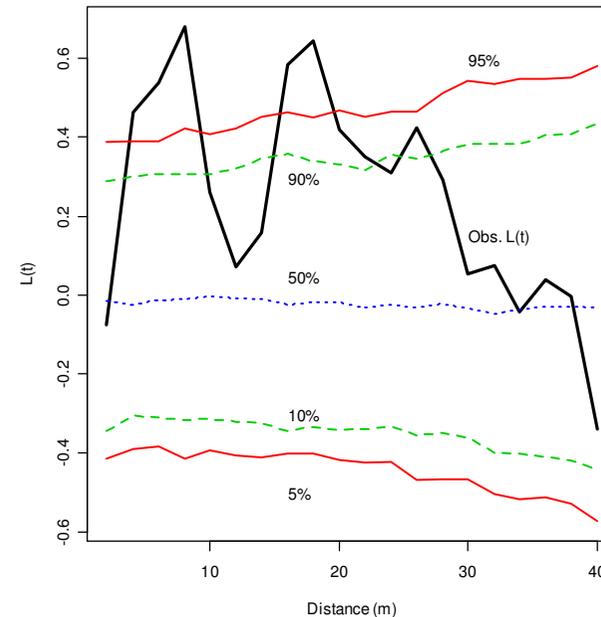
L(t) for all diseased trees



Are diseased trees clustered?

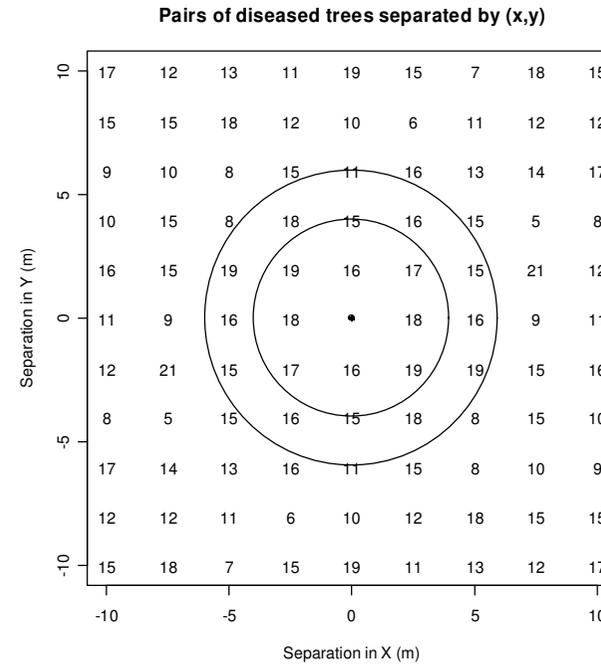
- Data: estimated $L(t) > 0$ from 4 – 30 m
- But: $L(t)$ estimated from 230 points. How large is the sampling variation?
- Easy to construct test of $H_0: K(t) = 0$ at that specific distance t.
- Simulate complete spatial random process, estimate $K(t)$ and $L(t)$, repeat many times. Calculate quantiles.

Comparing L(t) to point-based simulation

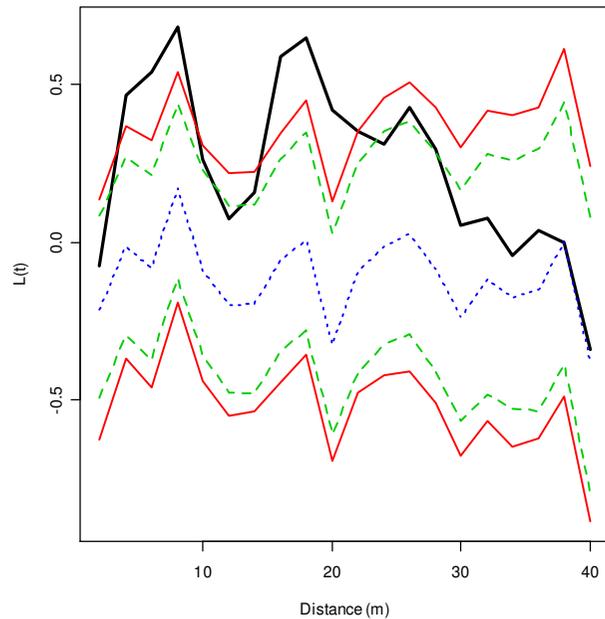


Interpretation of L(t) plots

- Focus on clustering, so one-sided test(s)
- More diseased trees than expected within 4m, 6m, 8m, 16m, and 18m of other diseased trees.
- But, trees aren't anywhere: on a grid
- Randomly choose 230 'diseased' locations from the 3770 possible grid positions



Comparing L(t) to grid-based simulation

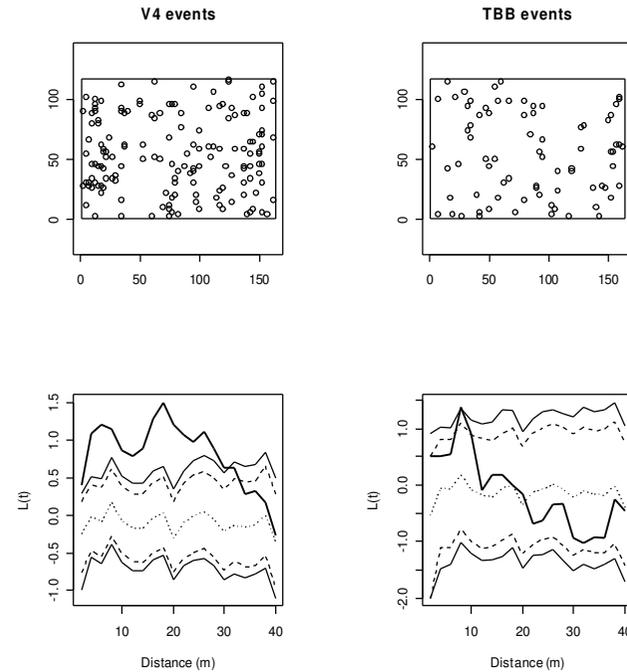


Notes

- Same legend as slide 23:
- Black line = observed L(t)
- Red lines = 0.05 and 0.95 quantiles
- Green lines = 0.10 and 0.90 quantiles
- Blue line = median (0.5 quantile)
- Lines are jagged because there are lots of trees separated by 2m, lots separated by 2.5m, lots separated by 4m
- But none separated by 1.5m, or 3.5m, because of the planting grid

Grid-based simulation

- Similar conclusions:
- More diseased trees than expected within 4m, 6m, 8m, 16m, 18m and 20m of other diseased trees.
- What about each type separately?



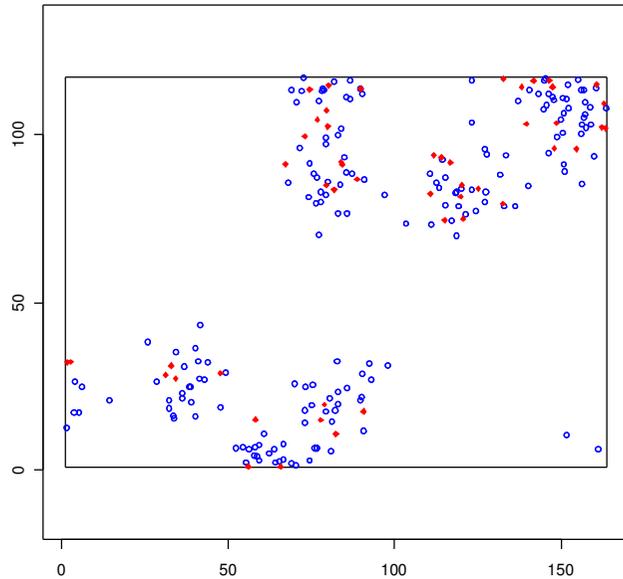
Notes

- Top pair of figures are locations of each type of phytoplasma, plotted separately
- Bottom pair of figures are L(t) for each type plotted separately
- Notice spread between 5% and 95% quantiles.
- Much larger for TBB events (only n=76) compared to V4 events (n=154)
- Can do the analysis for very small sample sizes (e.g. 30 locations), but power is very low.

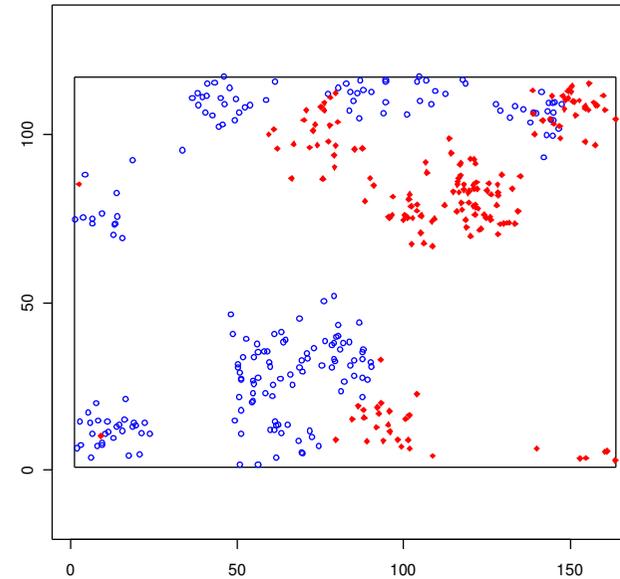
Association of types

- Are V4 events surrounded by (more, fewer) TBB events?
- Are there clusters of diseased trees? Or, separate clusters of V4 and TBB?
- Concerns relationships between two (or more) processes, not the characteristics of each process.

Clusters of diseased trees



Separate clusters of each type



Bivariate K functions

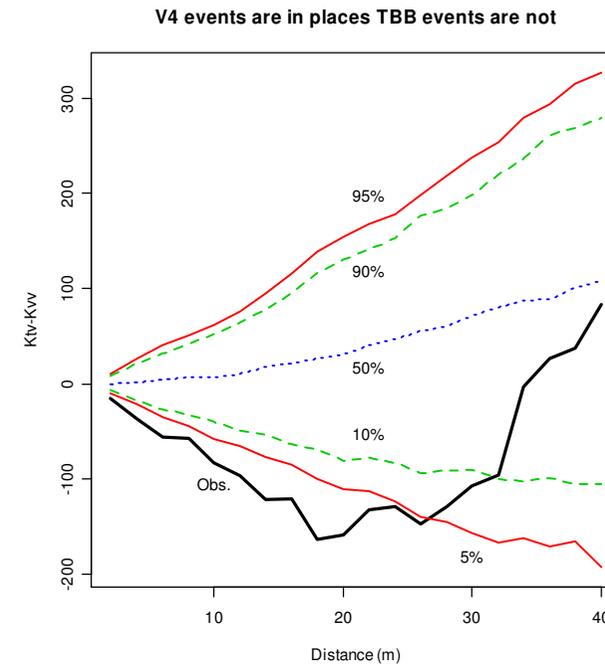
- Univariate K function: center circle on a location, count # additional events
 - 3 of these, All events. $K(t)$
 - V4 events to themselves. $K_{VV}(t)$
 - TBB events to themselves. $K_{TT}(t)$
- Bivariate K function: center circle on each V4 location, count number of TBB points in that circle. $K_{VT}(t) = K_{TV}(t)$

Notes

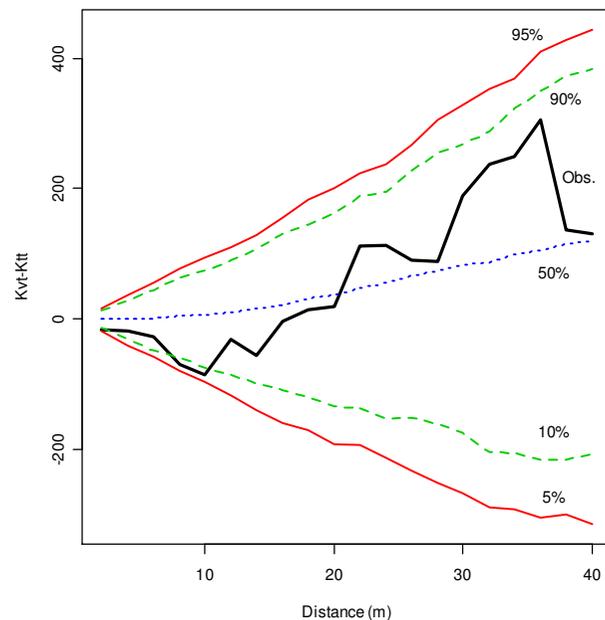
- Estimate each univariate $K(t)$ by considering all locations, only V4 locations or only TBB locations
- Estimate bivariate K function by generalizing the estimate of $K(t)$
- Because of edge effects, estimated $K_{VT}(t)$ is not the same as the estimated $K_{TV}(t)$. Usually averaged.
- $K_{VT}(t)$ is sometimes called the cross-K function
- Two commonly used null hypotheses:
 - Random labeling: Labels (TBB or V4) are randomly assigned to disease locations
 - Under random labeling: $K_{VT}(t) = K_{TT}(t) = K_{VV}(t) = K(t)$,
 - Independence: process generating TBB locations is independent of that generating V4 locations
 - Under independence: $K_{VT}(t) = \pi t^2$
- Simulate random labeling by randomly assigning labels to observed disease locations
- Simulating independence is more difficult. Toroidal rotation is one possibility
- I used random labeling

Association between types

- Examine using differences of K functions
- Q: Are V4 events surrounded by (more, fewer) TBB events? $K_{VT}(t) - K_{VV}(t)$
- Q: Are TBB events surrounded by (more, fewer) V4 events? $K_{VT}(t) - K_{TT}(t)$:
- Null hypothesis:
 - points are randomly labeled,
 - both differences = 0



But, can't tell what's happening with TBB events

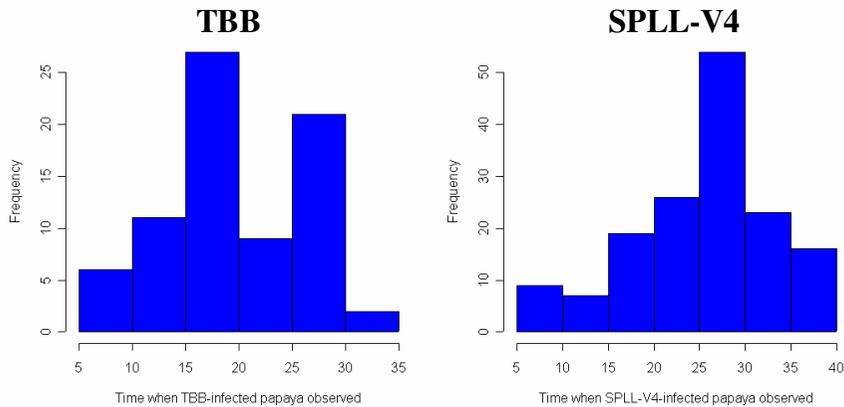


What about aggregation in time?

- Data includes the time an infection first noticed.
- So far, analyses have ignored time
- Is the number of newly infected trees constant over time?
- No.

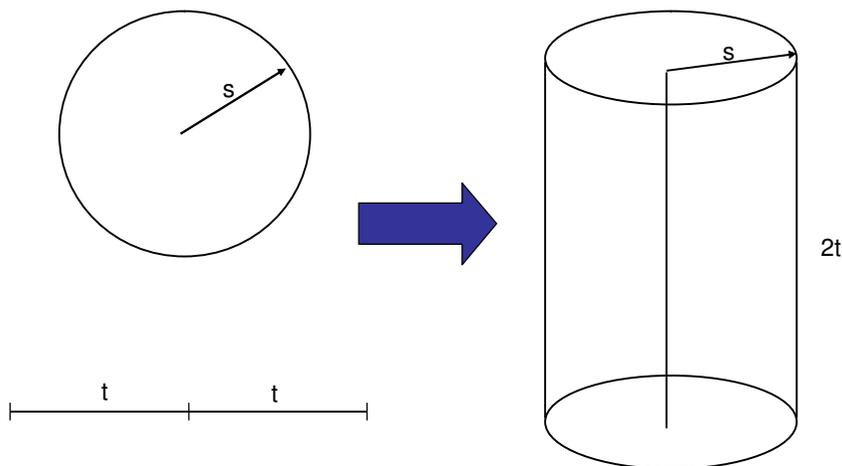
Are space and time independent?

- We have shown Spatial Aggregation
 $K(s) \neq CSR$
- And Temporal Aggregation
 $K(t) \neq CTR$
- But are events close in space also close in time?
- Expect this if disease spreads by local transfer by insects.
- Define $K(s,t)$ in terms of # events within s in space and t in time



Evidence that there is a non-uniform frequency of the time when a papaya was found to be infected with either phytoplasma strain

Space-Time “Windows”



Courtesy: Dale Tessin, Depts. EEOB and Statistics, ISU

Space-time independence

- If space and time are independent,
 $K(s,t) = K(s) K(t)$
- Measure departure from independence by

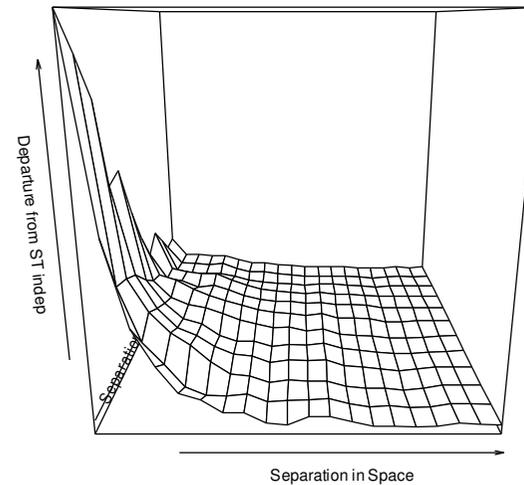
$$D_0 = \frac{K(s,t) - K(s)K(t)}{K(s)K(t)} = 0?$$

- $D_0 > 0$ when space-time contagion

Notes

- Division by $K(s,t)$ is to equalize (at least approximately) the variance of $K(s,t) - K(s)K(t)$

$D(s,t)$ to assess space-time independence



Notes

- (0,0) is the forward left corner
- D_0 is $\gg 0$ in the forward lefthand corner, that is for pairs of points close in space and close in time.
- Randomization test not shown, but results are highly significant. Space and time are not independent

What have we learned?

- Locations of diseased trees are clustered
 - Don't know whether contagion (infection) or consequence of environmental variation across plot
- Seem to be two processes,
 - one for each type (V4 and TBB)
- Rate of new infection varies over time
- Space-time assoc. suggests contagion
 - Events close in time tend to be close in space

Additional Materials

- Day-long short course on spatial point pattern analysis for ecologists on the web at:
- <http://www.ci.uri.edu/projects/geostats/Theory.pdf> concepts and examples
 - Large bibliography, emphasizing ecological applications at the end. Not updated since 2003.
- <http://www.ci.uri.edu/projects/geostats/Hands-on.pdf> computing using Splus (R is very similar)
- My favorite text is:
- Diggle, P.J. 2003. Statistical Analysis of Spatial Point Patterns, 2nd ed. Arnold / Oxford Univ. Press.

