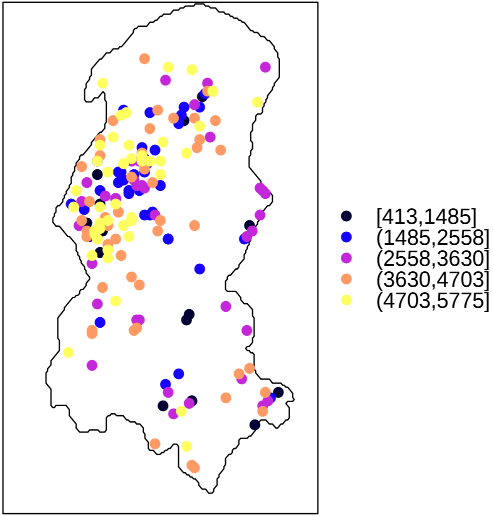
Point processes in space and time

- Data: (s,t) = location (probably in 2D), time of an event
- Examples:
  - Ecology: nest establishment, predation, death of tree
  - Plant pathology: infection of individual, death
  - Economics: farm consolidation, farm bankrupcy
  - Epidemiology: location, time of diseased individuals
  - Agronomy / soils: when a field planted in corn following corn
  - In general: where and when events occur
- Cressie calls these 'space-time shock point process' (p 720)
  - events at specific times and locations

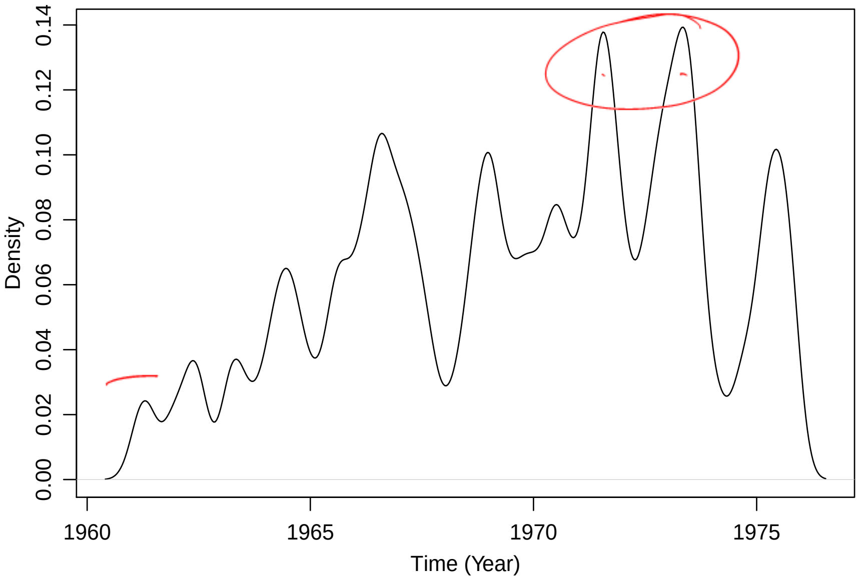
Point processes in space and time

- Not same as space-time survival point process
  - event at a location for an interval of time
  - e.g. presence of tree, infected plant
  - We'll focus on shock processes
    - events at specific locations and times
- Multiple ways to think about this
  - as point process in 3 dimensions (x,y,t)
  - as marked spatial point process (x,y), continuous mark (time)
- Many examples from spatial epidemiology, focus on clustering
  - Is event (disease) near other events?
  - near defined as near in space **and** near in time

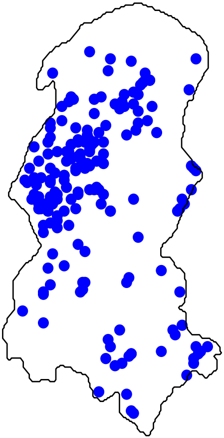
Example: Burkitt's lymphoma in Uganda



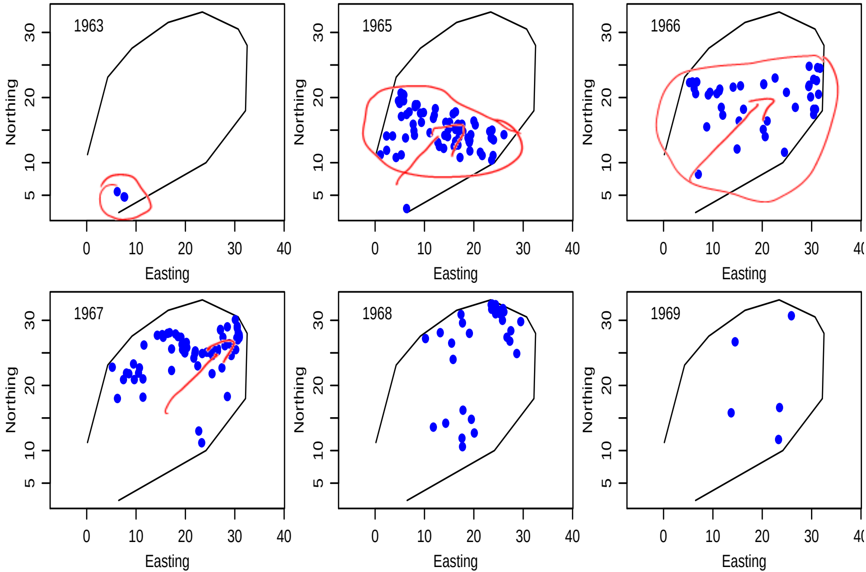
Example: Burkitt's lymphoma, Time ignoring space



Example: Burkitt's lymphoma, Space ignoring time



Example: Fox rabies in South Germany, 1963-1970, April



Biological Questions:

- Data are usually from observational or happenstance studies
  - questions not formulated before data are collected
- Many possible, some can be answered without new methods
- ignore time
  - are events clustered in space, without regard to time?
  - usual K(x) or g(x) analysis
- ignore space
  - are events clustered in time, without regard to space?
  - K(t) in one-dimension

Biological Questions:

- Some need new techniques
- classify events by year: relative clustering
  - are events in 1966 more clustered than those in 1967?
  - Compare two K(x) functions, not discussed this year
- classify events by year: spatial segregation
  - are events in 1963 in diff. places than those in 1964
  - use methods for point processes with two types of marks
  - not discussed this year
- are time and space independent?
  - If epidemic spread by contact, expect space-time clustering.
  - events close in space also are close in time.
  - How can we used space-time data to evaluate this?

Approaches / what I'll talk about:

- models (mostly simple) for space-time point processes
- mapping intensity in space x time (3D kernel)
- space-time K function
- Not discussing classical approaches for space-time clustering
  - Knox test
  - Mantel test = correlation between two distances
  - scan statistics for disease surveillance

Models:

HPR CSTR

- Space-time Poisson process:
  - notation: s spatial coordinates, t time
  - P[event in $(s, s+ds)(t, t+dt)] = \lambda(s,t)\, ds\, dt$
  - events are independent
- When interested in space-time patterns,
  - don't really care about marginal distributions in space or time
- focus on interaction, so model: $\lambda(s,t) = f(s)g(t)h(s,t)$
  - $f(s)$: marginal spatial intensity
  - $g(t)$: marginal temporal intensity
  - $h(s,t)$: interaction between space and time
- Space-time independence: above with $h(s,t) = 1$.
  - $\lambda(s,t) = f(s)g(t)$

λ = dependent interaction space time

$\log \lambda = S + T$

$\log \lambda = S + T + \text{interaction}$

Models:

- Complete Spatial-temporal randomness (CSTR):
  - Extend CSR to space-time volumne
  - $\lambda(s, t) = \lambda$
  - N = # events in box S,T $\sim$ Poisson(λST).
  - E N $= \lambda ST$.
  - Var N $= \lambda ST$.
- Very few general models, mostly Markov-like, for earthquake clustering
- Modern approaches rely on process models
  - How do locations of events at time t depend on events at time $t - 1$?
  - Or, how does $\lambda(s, t)$ depend on $\lambda(s, t - 1)$?
  - Details are all problem specific
- Process models are the future of space-time analyses
- Require close interaction between domain experts and statisticians
  - Domain experts: what is a appropriate model for the process?
  - Statistician: how can you fit that model and evaluate uncertainty?

Estimating intensity

- Data: Locations (space,time) of events
- want to estimate $\hat{\lambda}(s,t)$.
- Key question: How smooth is the intensity surface?
    - data are very 'rough'
        - shift from 0 (no event) to 1 (event) over short dist.
        - Intensity surface is (probably) not that rough

-
    - Very very smooth $\Rightarrow$ constant intensity
        - $\hat{\lambda}$ = #points/(area*time)
        - if want to produce a map, constant intensity makes a dull map

Local estimates of intensity

- Extend kernel smoothing to (S, T)
- Can be used with any form of ST data
  - Does not require contemporaneous or colocated events
- Contribution to $\hat{\lambda}(s, t)$ is $k(h)$
  - $k()$ is the kernel function
  - h is the space-time distance $\| \boldsymbol{\sigma} - \boldsymbol{\sigma}_i \|^2 / \tau_s^2 + (t - t_i)^2 / \tau_t^2$

Space   Spatial bandwidth

- Need to specify two bandwidths:
  - For spatial smoothing, τ_s
  - For temporal smoothing, τ_t

time   time bandwidth

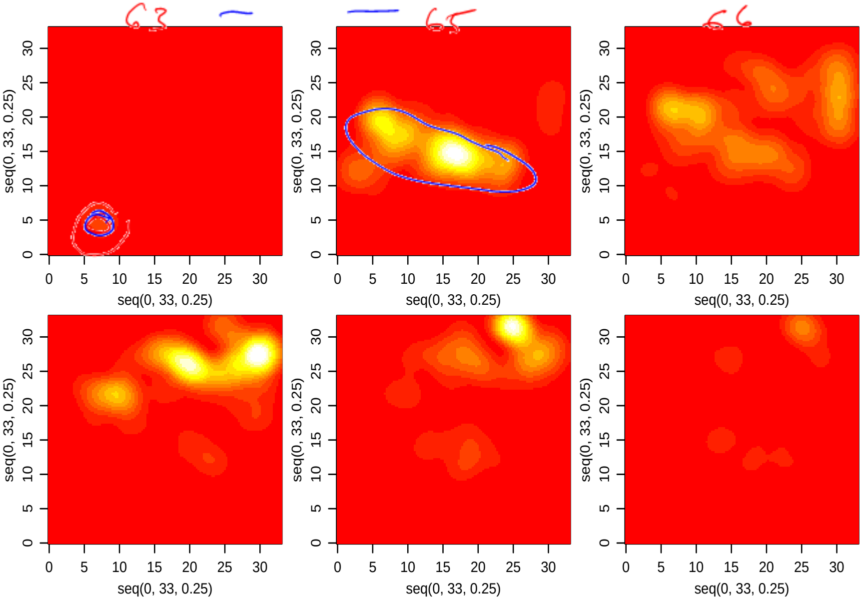
- Can define MSE or lnL, details harder

Alternative
years, spatial each year

- Practical: use what seems reasonable

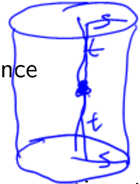
bandwidth

Fox rabies, $\tau_s = 4$, $\tau_t = 1.5$



Space-time clustering and space-time K function

- Q: is an event near other events? i.e. are events clustered?
  - Could look directly at $h(s, t)$, the interaction component of $\lambda(s, t)$
    - Spatial trend not same at each time point, or
    - Temporal trend not same at each location
- Easier to think of this as a second order property
  - evaluate using K(s,t) or g(s,t) pair correl
  - Focus here on K(s,t)
    - Only because software for this has been around longer
  - From here on, s is now a space distance, t a time difference
- $K(s, t)$ defined as:



$$K(s, t) \;=\; \frac{1}{\lambda}\,\mathrm{E}\ \#\text{ events within distance } s \text{ and time separation } t \text{ of a randomly chosen event}$$

- λ is average # events per unit of space and unit of time

K(s,t) under CSTR

- $K(s,t) = (\pi s^2)(2t)$
  - Why $2t$? Time is linear, look back and look ahead t units
- Estimator:

$$\hat{K}(s,t) = \frac{\mid A \mid T}{N^2} \sum_{i \neq j} w_{ij} v_{ij} I(d_{ij} < s) I(t_{ij} < t)$$

  - $\mid A \mid$ is area of the region
  - T is the total study time
  - I() are indicator functions, 1 if true, 0 if not
  - w_{ij} and v_{ij} are space and time edge corrections
- Properties:
  - $\hat{K}(s,t)$ approx. unbiased, esp. small s, t
  - Var $\hat{K}(s,t)$ not constant, increases with s, t
- Testing CSTR:
  - Compare $\hat{K}(s,t)$ to envelope for data simulated from CSTR.

Problem with CSTR as a null hypothesis

- When want to evaluate interaction, CSTR is too simple.
  - Marginal spatial pattern is CSR
  - Marginal temporal pattern is Poisson
- Really concerned about the interaction
  - Without specifying marginal patterns

no interaction

Independence of space and time

- If process in space and process in time are independent,

$$K(s,t) = K_{\text{space}}(s) \times K_{\text{time}}(t)$$

$\hat{K} > \pi s^2$?
$\Rightarrow$ clust.

- Suggests using

0 indep.

$$\hat{D}(s,t) = \hat{K}(s,t) - \hat{K}_{\text{space}}(s) \times \hat{K}_{\text{time}}(t)$$

- to evaluate independence
- Idea proposed and developed by Peter Diggle
- $\hat{D}(s,t) > 0 \Rightarrow$ space-time clustering at that distance and time domain
- Interpretation:
  - Remember $\lambda K(s,t)$ =E[# events w/i distance s and time t]
  - $\lambda D(s,t)$ = E[addn events due to space-time clustering]
  - $\hat{\lambda}\hat{D}(s,t)$ = est. # addn. events w/i s,t

CSTR : Simulated 2 poisson / space
— time

now : Space constant
Time constant) indp.

- Test using randomization:
  - randomly reassign times to locations.
  - compute envelope for $D(s, t)$
  - provides answers for each $D(s, t)$

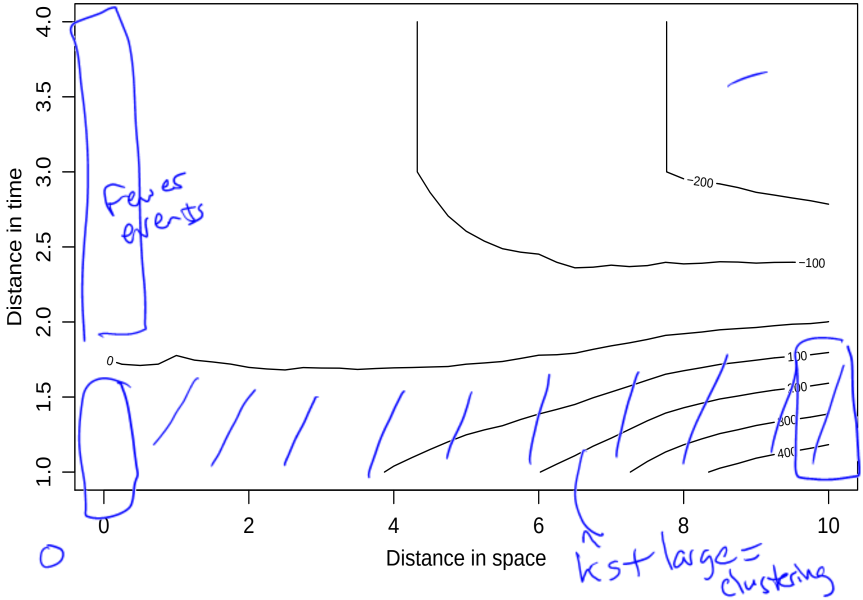
time

1	3
5	10

12 7 12

- These are point-wise tests (as with spatial problems)
- to get a single answer: compute a summary statistic
  - Diggle et al. 1995 suggest $\sum_s \sum_t \hat{D}(s,t)/\sqrt{\text{Var } D(s,t)}$
  - ~ 0 if no clustering, > 0 if clustering, < 0 if repulsion
  - range of s and range of t matter

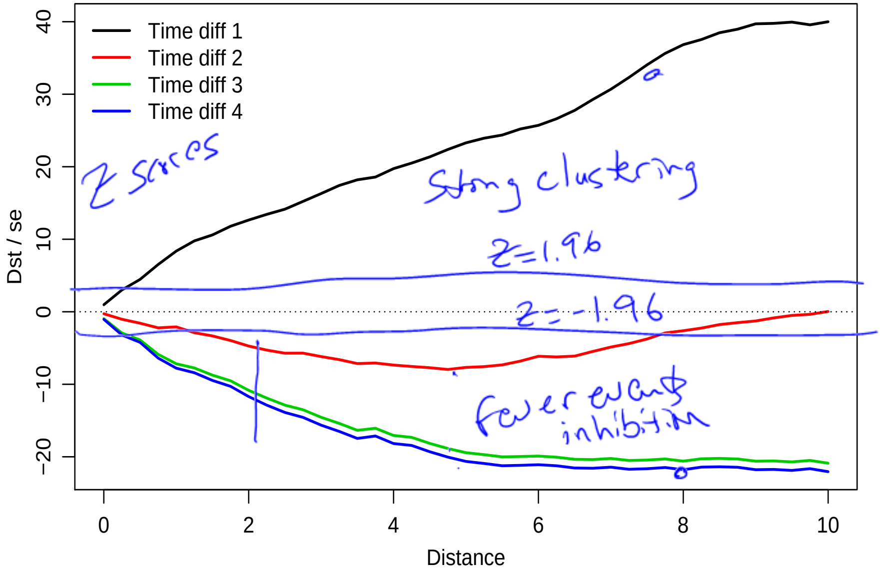
Fox rabies Dst



Conclusions for Fox rabies

- $D(s,t) > 0$ at T=1 year, especially large distances
  - case surrounded by more cases 1 time later at all distances
  - $\hat{D}(10,1) = 664$, $\hat{\lambda} = 0.0257$, 17.1 extra events
  - case surrounded by fewer cases 4 times later at all distances
  - both especially so for larger distances
- Support what seen in pictures
  - Space and time not independent
  - Positive ST clustering at one year
- Is this just random variation?
- Could use Monte-Carlo envelopes
- An approximate answer:
  can estimate se of $D(s,t)$ without simulation
- So plot $\frac{D(s,t)}{\text{se } D(s,t)}$ for t=1, 2, 3 or 4

Fox rabies Dst/se Dst



Conclusions for Fox rabies

- Is this just random variation?
- Use Diggle summary statistic
- At T = 1 year
  - obs $D(s, t)$ **larger** than all 99 random, p=0.01
- At T = 4 year
  - obs $D(s, t)$ **smaller** than all 99 random, p=0.01
- Consistent with a slowly moving outbreak
- Also says: if you have an outbreak here, it will clear in a couple of years.
  - so not spatially persistent
- Matches pictures, but analysis adds two useful things
  - Quantify intensity of the effects
  - Show they are more extreme than expected by chance

Summary of space-time analyses

- Lots of practically important questions
- many require new methods
  - not just many spatial analyses
- can combine information across times
  - Kernel smoothing of space-time point patterns
  - Space-time geostatistics
- And look at space-time independence
- Pictures/graphs are really really helpful
  - Provide reality check and help with interpretation
- My view of the future:
  - fitting subject-matter based models to dynamic data