6. 時空交互作用與群聚

Space-Time Interaction and Clustering

https://ceiba.ntu.edu.tw/1062_Geog5016

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

Contents

- **Space-Time Interaction**
  - Space-time data: point event (locations)
  - Tests of Space-Time Interaction
    - Knox Test (1964) and Mantel Test (1967, 1970)
    - Jacquez K-Nearest Neighbor Test (1996)
  - Mapping the Areas with Space-Time Interaction (Hotspots)
  - Some Applications: Diffusion of crime and epidemics
- **Space-Time Clustering**
  - Space-Time K function and applications

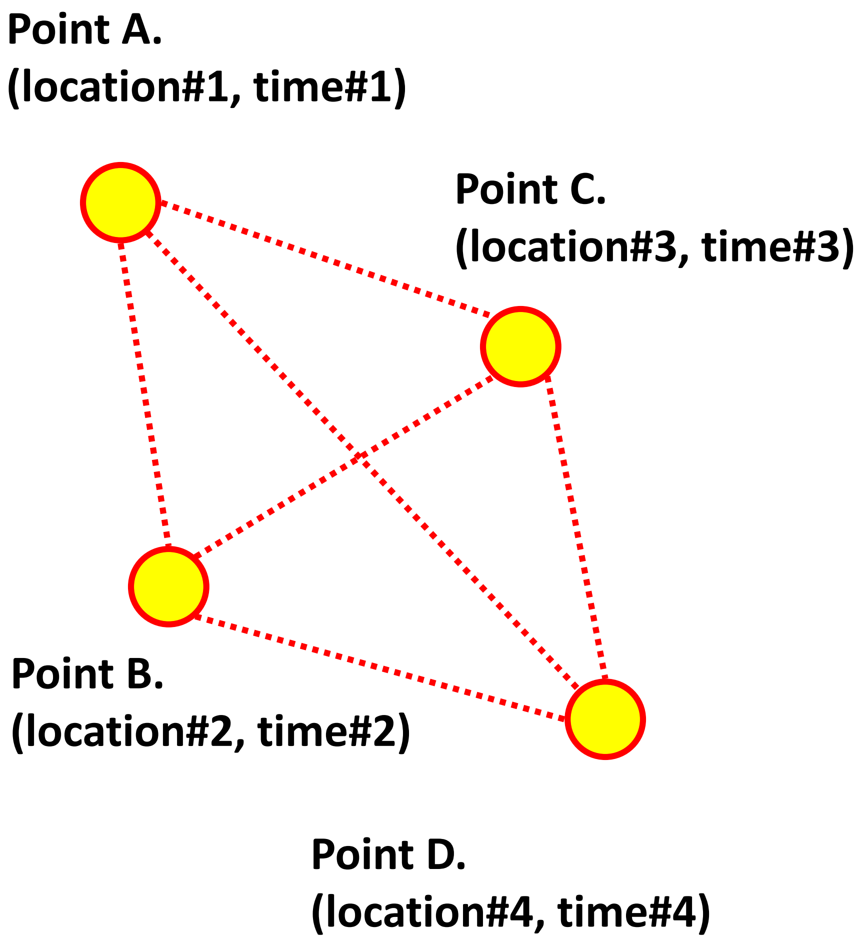
Space-Time Clustering

- *Spatial clustering* all the time
- *Spatial clustering* within a specific time period
  - Hot spot could occur during certain time periods
- *Space-time* clustering
  - A number of events could occur within a short time period within a concentrated area.
  - There is an interaction between space and time in that spatial hot spots appear at particular times, but are temporary.

本週課程的相關教材

- Test of Space-Time Interaction
  - Jacquez GM.(1996), A K Nearest Neighbor Test for Space-time Interaction, *Statistics in Medicine* 15:1935-49.
  - Grubesic TH and Mack EA (2008), Spatio-Temporal Interaction of Urban Crime, *Journal of Quantitative Criminology* 24:285–306.
  - Wen TH, et al (2012), Population Movement and Vector-borne Disease Transmission: Differentiating Spatial-temporal Diffusion Patterns of Commuting and Non-commuting Dengue Cases. *Annals of the Association of American Geographers*, 102(5):1026-1037.
- Space-Time Clustering
  - Kang (2010), Detecting Agglomeration Processes using Space-Time Clustering Analyses, *Annals of Regional Science* (2010) 45:291–311.
  - Arbia et al (2010), Detecting the Existence of Space–Time Clustering of Firms, *Regional Science and Urban Economics* 40:311–323.
  - Mstras et al (2011), Exploratory Space-Time Analyses of Rift Valley Fever in South Africa in 2008–2011, *PLOS Neglected Tropical Diseases* 6(8): e1808.

Point Data with Location and Time



Pairs of Points = N*(N-1)/2 = 6

Pair	Distance	Time Interval
A-B	Location#1-Location#2	Time#1-Time#2
A-C		
A-D		
B-C		
B-D		
C-D		

Sample Data

point3.rar

Data: point3.shp

View Data Table

Record	EVENT	X Coordinate	Y Coordinate	DAYNUMBER
1	1	3	1	1
2	2	1	3	3
3	3	1	5	5
4	4	3	7	7
5	5	1	9	9
6	6	1	11	11
7	7	3	13	13
8	8	1	15	15
9	9	1	17	17
10	10	3	19	19
11	11	1	21	21
12	12	1	23	23

# of Events (N) = 12

Pair of Points =

➔ N*(N-1)/2 = 66

Generating the Distance and Time Matrix (作業 1-1)

Time Matrix

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1	0	2	4	6	8	10	12	14	16	18	20	22
2	0	0	2	4	6	8	10	12	14	16	18	20
3	0	0	0	2	4	6	8	10	12	14	16	18
4	0	0	0	0	2	4	6	8	10	12	14	16
5	0	0	0	0	0	2	4	6	8	10	12	14
6	0	0	0	0	0	0	2	4	6	8	10	12
7	0	0	0	0	0	0	0	2	4	6	8	10
8	0	0	0	0	0	0	0	0	2	4	6	8
9	0	0	0	0	0	0	0	0	0	2	4	6
10	0	0	0	0	0	0	0	0	0	0	2	4
11	0	0	0	0	0	0	0	0	0	0	0	2
12	0	0	0	0	0	0	0	0	0	0	0	0

Distance Matrix

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1	0	2.828427	4.472136	6.000000	8.246211	10.198039	12.000000	14.142136	16.124515	18.000000	20.099751	22.090722
2	0	0.000000	2.000000	4.472136	6.000000	8.000000	10.198039	12.000000	14.000000	16.124515	18.000000	20.000000
3	0	0.000000	0.000000	2.828427	4.000000	6.000000	8.246211	10.000000	12.000000	14.142136	16.000000	18.000000
4	0	0.000000	0.000000	0.000000	2.828427	4.472136	6.000000	8.246211	10.198039	12.000000	14.142136	16.124515
5	0	0.000000	0.000000	0.000000	0.000000	2.000000	4.472136	6.000000	8.000000	10.198039	12.000000	14.000000
6	0	0.000000	0.000000	0.000000	0.000000	0.000000	2.828427	4.000000	6.000000	8.246211	10.000000	12.000000
7	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.828427	4.472136	6.000000	8.246211	10.198039
8	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.000000	4.472136	6.000000	8.000000
9	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.828427	4.000000	6.000000
10	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.828427	4.472136
11	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.000000
12	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

R Functions and Package

```
source("ST_functions.R")
```

Functions	
DiggleETAL.test	function (pts, time, polygon, range, s, t, Nrep)
Jacquez.test	function (x, y, time, k, Nrep)
KnoxA.test	function (x, y, time, del1, del2)
KnoxM.test	function (x, y, time, del1, del2, Nrep)
Mantel.test	function (x, y, time, c1, c2, Nrep)

```
source(splancs)
```

Tango T. *Statistical Methods for Disease Clustering*, Springer, 2010.

1. Concept of Knox Test (1964)

- The Knox method quantifies space-time interaction based on critical space and time distances.
- The test statistic, X , is a count of the number of pairs of cases that are separated by less than the critical space and time distances. The concept is that pairs of cases will be near to one another when interaction is present, and the test statistic will be large.

Knox Test (1964)

$$X = \sum_{i=1}^{n} \sum_{j=1}^{i-1} a_{ij}^{s} a_{ij}^{t}$$

where n = number of cases; δ = critical space distance; τ = critical time distance

$$a_{ij}^{s} = \begin{cases} 1 & \text{if the distance between cases } i \text{ and } j < \delta \\ 0 & \text{otherwise} \end{cases}$$

$$d_{ij}^{t} = \begin{cases} 1 & \text{if the distance between cases } i \text{ and } j < \tau \\ 0 & \text{otherwise} \end{cases}$$

Significance Test 1: Monte Carlo Simulation

H_o	The times of occurrence of the health events are distributed randomly across the case locations. This is another way of saying the time distances between pairs of cases are independent of the spatial distances between pairs of cases.
H_a	Pairs of cases near in space tend to be near in time.

- Monte Carlo significance test
  - The null hypothesis states that the times of occurrence of the health events are distributed randomly across the case locations.
  - This procedure is accomplished a fixed number of times, and the reference distribution is constructed by calculating X each time from the newly randomized data. The probability value is the proportion of the upper right hand tail of the reference distribution whose X values are as large or larger than the test statistic.

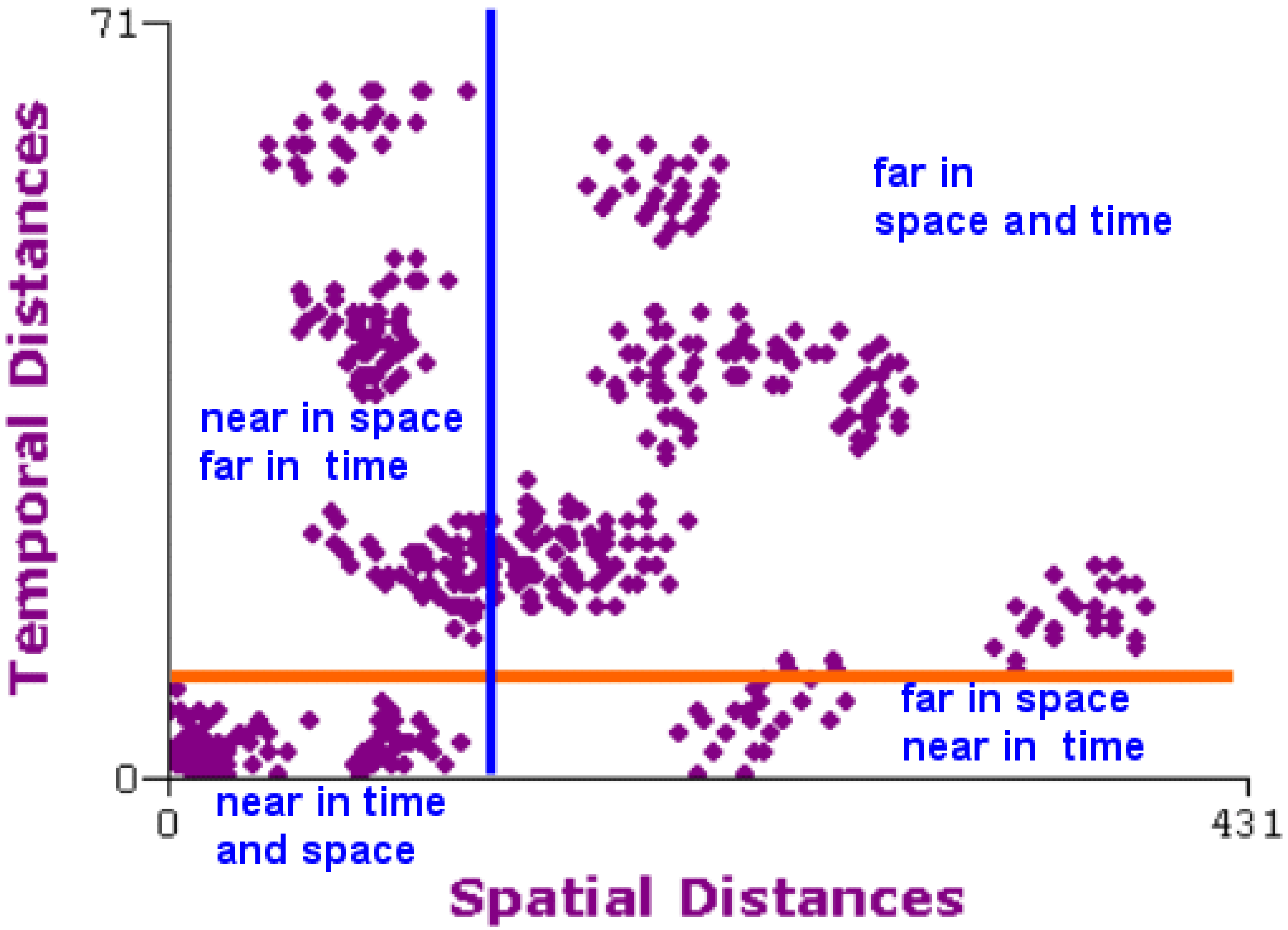
Significance Test 2: Chi-squared Statistic

- A 2 x 2 contingency table is used which classifies pairs of cases as near or far in both space and time, for a total of four possible outcomes.

		SPACE	
		Close	Not Close
TIME	Close	Space and Time	Time Only
	Not Close	Space Only	Not Close

卡方獨立性檢定：兩變數是否有相關

Illustration



Logical Structure of Knox Index

	Close in time	Not close in time	
Close in Distance	O_1	O_2	S_1
Not close in distance	O_3	O_4	S_2
	S_3	S_4	N

N = Total Number of Pairs

where $N = O_1 + O_2 + O_3 + O_4$

$$S_1 = O_1 + O_2$$
$$S_2 = O_3 + O_4$$
$$S_3 = O_1 + O_3$$
$$S_4 = O_2 + O_4$$

Observed vs. Expected Frequencies

Observed

	Close in time	Not close in time	
Close in Distance	O_1	O_2	S_1
Not close in distance	O_3	O_4	S_2
	S_3	S_4	N

where $N = O_1 + O_2 + O_3 + O_4$

$S_1 = O_1 + O_2$

$S_2 = O_3 + O_4$

$S_3 = O_1 + O_3$

$S_4 = O_2 + O_4$

Expected

	Close in time	Not close in time
Close in Distance	E_1	E_2
Not close in distance	E_3	E_4

where $E_1 = S_1 * S_3 / N$

$E_2 = S_1 * S_4 / N$

$E_3 = S_2 * S_3 / N$

$E_4 = S_2 * S_4 / N$

Chi-square Statistic

The difference between the actual (observed) number of pairs in each cell and the expected number is measured with a Chi-square statistic (equation 9.1).

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} \quad \text{with 1 degree of freedom}$$

How to decide the degree of "close"?

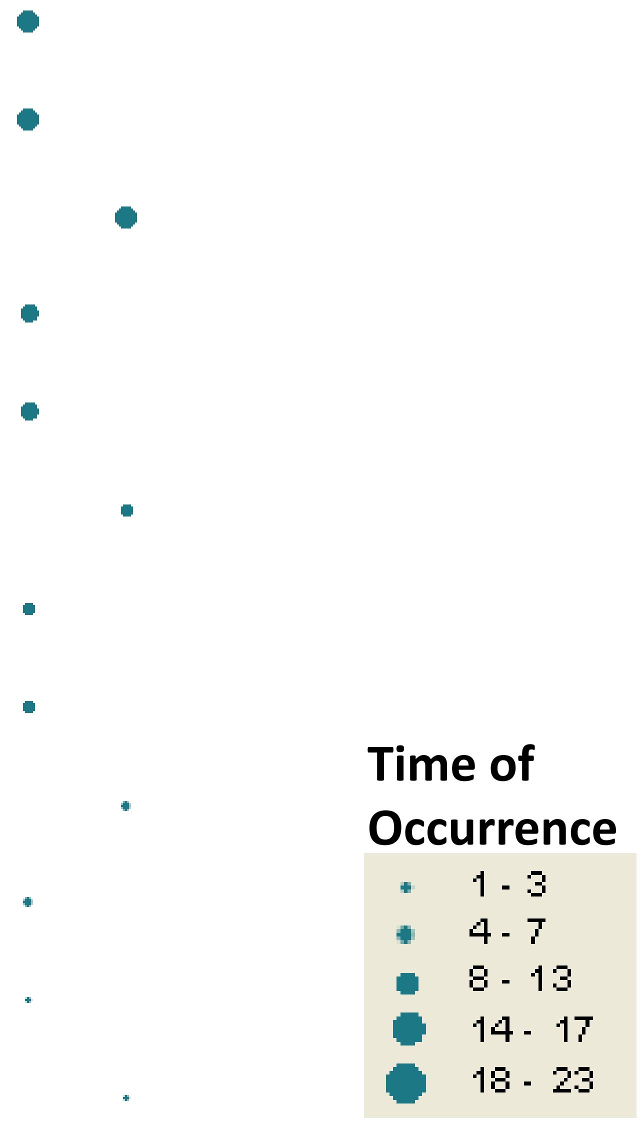
	Close in time	Not close in time	
Close in Distance	O_1	O_2	S_1
Not close in distance	O_3	O_4	S_2
	S_3	S_4	N

Methods for Dividing Distance and Time

In the *CrimeStat* implementation of the Knox Index, the user can divide distance and time interval based on the three criteria:

1. The mean (mean distance and mean time interval). This is the default.
2. The median (median distance and median time interval)
3. User defined criteria for distance and time separately.

Spatial-temporal patterns of the Sample data



The Contingency Table

```
"Close" time ........:            8.66667 days
"Close" distance ....:            8.85327 m

                     | Close in space(1)  | Not close in space(0) |
---------------------+-------------------+-----------------------+-----------------
Close in time(1)     |                38 |                     0 |              38
Not close in time(0) |                 0 |                    28 |              28
---------------------+-------------------+-----------------------+-----------------
                     |                38 |                    28 |              66

Expected:
                     | Close in space(1)  | Not close in space(0) |
---------------------+-------------------+-----------------------+-----------------
Close in time(1)     |          21.87879 |              16.12121 |        38.00000
Not close in time(0) |          16.12121 |              11.87879 |        28.00000
---------------------+-------------------+-----------------------+-----------------
                     |          38.00000 |              28.00000 |        66.00000

Chi-square ..........:          66.00000
P value of Chi-square:           0.00010
```

R Functions for Knox Test

```
KnoxM.test<-function(x,y,time,del1,del2,Nrep)
```

```
-------------------------------------------------------------------
 ARGUEMENTS
    x    : a vector of x-coordinates of case
    y    : a vector of y-coordinates of case
    time : a vector of observed times
    del1 : a measure of closeness in space
    del2 : a measure of closeness in time
    Nrep : The number of Monte Carlo replications, e.g., 999, 9999

 VALUES
    Knox.T            : test statistic
    Freq              : a vector of simulated test statistics under the null
    Simulated.p.value : simulated p-value
```

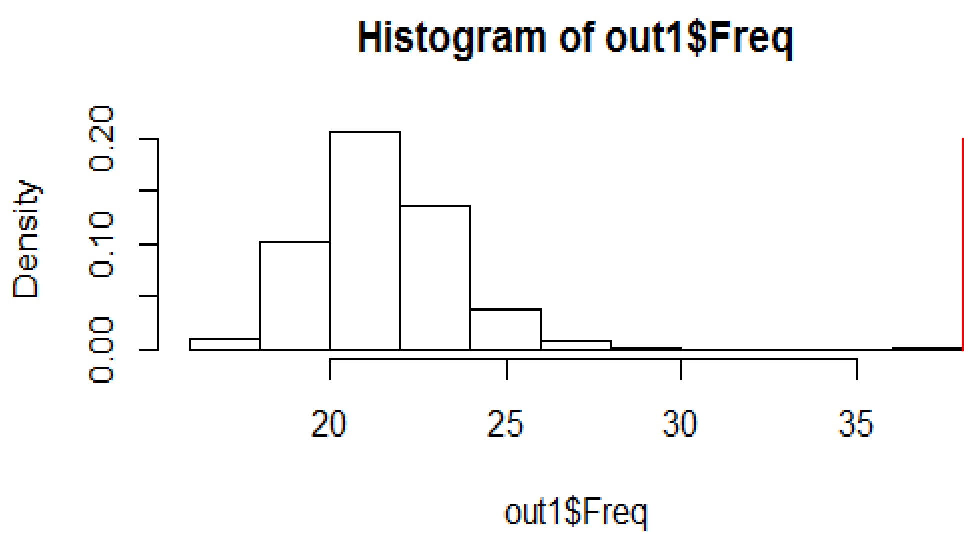
R code

```
source("ST_functions.R")
PT3 <- readOGR(dsn = "spacetime", layer = "point3",
encoding="big5")
head(PT3@data)

xcoord<-PT3@data$X
ycoord<-PT3@data$Y
time<-PT3@data$DAYS

out1<-KnoxM.test(xcoord,ycoord,time,8.8,8.6,99)
out1
```

R code: Results

```
> out1
$Knox.T
[1] 38

$Freq
  [1] 20 23 24 21 23 31 21 24 23 22 24 19 20 22 19 23 23 21 24 20 24 23 25 21 20 23 25
 [28] 22 21 22 22 21 26 24 19 21 20 17 23 21 22 24 20 21 20 20 23 24 24 23 23 21 22 22
 [55] 23 21 23 19 22 20 21 19 23 22 25 23 26 22 21 21 22 21 24 23 22 21 22 25 23 22 22
 [82] 21 21 22 24 24 19 28 25 27 27 20 26 21 20 25 22 21 21 38

$simulated.p.value
[1] 0.01
```

```
hist(out1$Freq, freq=F)

lines(x=out1$Knox.T, y=0.2,
col="red", type="h")
```



Application

Knox Index for Baltimore County Vehicle Thefts
Median Split
N = 1,855 with 1,719,585 comparisons

Month	Actual Chi-square	95 Percentile Simulation Chi-square	Approx. p
January	0.26	6.95	n.s.
February	0.00	6.61	n.s.
March	0.00	6.86	n.s.
April	0.50	6.56	n.s.
May	1.04	7.25	n.s.
June	0.01	6.02	n.s.
July	9.96	9.05	.05
August	5.91	5.55	.05
September	0.27	5.41	n.s.
October	3.33	6.43	n.s.
November	10.79	8.91	.01
December	0.00	6.87	n.s.
All of 1996	**8.69**	**41.89**	**n.s.**

Problems with the Knox Index

- **Subjective.** different results can be obtained by varying the cut-off points for distance or time.
- Not incorporate the changes of population-at-risk (assume that the population size does not change over time)
- Difficult to interpret
  - the observed and expected frequencies could occur in any cell or any combination of cells .
  - Finding a significant relationship does not automatically mean that events that were close in distance were also close in time; it could have been the opposite relationship. (using Chi-square test)
- However, a simple inspection of the table can indicate whether the relationship is as expected or not.

2. Concept of Mantel Test (1967)

- Mantel's statistic is the sum, across all case pairs, of the time distances multiplied by the spatial distances.

N: Number of cases.
d^{t}_{ij}: Distance between cases i and j in time.
d^{s}_{ij}: Distance between cases i and j in space.
$\bar{d}^{s}, \bar{d}^{t}$: Average space and time distances.
s_{s}, s_{t}: Standard deviations of the space and time distances.
Z: Test statistic, also called the Mantel product, $Z = \sum_{i=1}^{N} \sum_{j=1}^{N} d^{s}_{ij} d^{t}_{ij}$.
r: Standardized Mantel statistic, $r = \frac{1}{(N^2 - N - 1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(d^{s}_{ij} - \bar{d}^{s})}{s_{s}} \frac{(d^{t}_{ij} - \bar{d}^{t})}{s_{t}}$.

Mantel Test (1967)

H_o	The times of occurrence of the health events are distributed randomly across the case locations. This is another way of saying the time distances between pairs of cases are independent of the spatial distances between pairs of cases.
H_a	Pairs of cases near in space tend to be near in time.

$$Z = \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} t_{ij}$$

s_{ij} is the distance between i and j in space

t_{ij} is the distance between i and j in time

N is the number of cases

Reciprocal transformation

- For a contagious diffusion, we expect the **small space and time distances to be correlated, but not the large distances**. Mantel therefore recommended the use of the reciprocal transformation ($d' = 1/(C + d)$) to reduce the effect of large space and time distances. Here C is a constant and d is the distance to be transformed. Selection of the constant C is a matter of judgment and is subjective.

Mantel test statistic

The Mantel test statistic is

$$T = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{S} a_{ij}^{T} \tag{7.22}$$

where a_{ij}^S and a_{ij}^T denote the *clinal type measures of closeness* in space and in time, respectively, and are given by

$$a_{ij}^{S} = \frac{1}{d_{ij}^{S} + c_1} \quad (a_{ii}^{S} = 0) \tag{7.23}$$

$$a_{ij}^{T} = \frac{1}{d_{ij}^{T} + c_2} \quad (a_{ii}^{T} = 0) \tag{7.24}$$

and c_1 and c_2 are unknown parameters and have to be prespecified by the user. The expected value of T is given by (7.17).

Mantel Test

(Mantel and Bailar , 1970)

Pearson's Correlation (複習)

$$\mathrm{cov}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)],$$

$$\rho_{X,Y} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Resolves some of the problems of the Knox Index

探討「距離」與「時間」相似的程度

$$T = \sum_{i=1}^{N} \sum_{j=1}^{N} (X_{ij} - \mathrm{Mean}X)(Y_{ij} - \mathrm{Mean}Y)$$

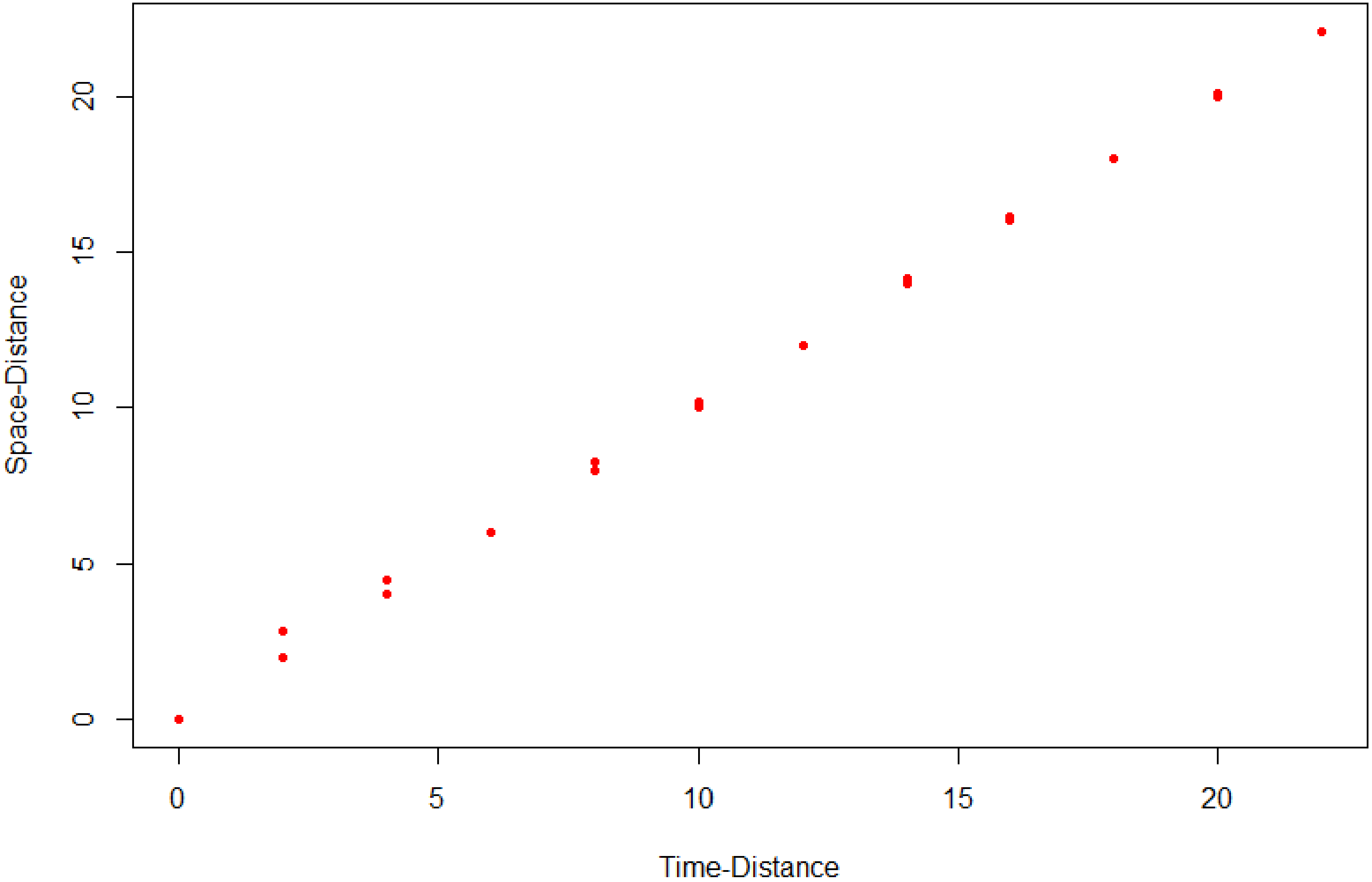
X: distance

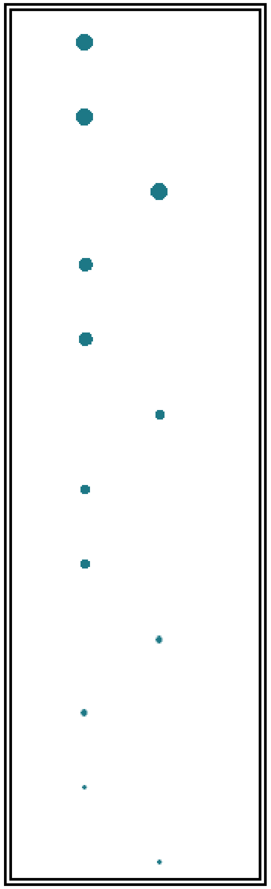
Y: time interval

Standardized Mantel statistic

$$r = \frac{1}{(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} (X_{ij} - \mathrm{Mean}X)/S_x * (Y_{ij} - \mathrm{Mean}Y)/S_y$$

Association between Space and Time (作業 1-2)





Application

Mantel Index for Baltimore County Vehicle Thefts
Median Split
N = 1,855 and 1,719,585 Comparisons

Month	r	Simulation 2.5%	Simulation 97.5%	Approx. p-level
January	-.0047	-0.033	0.033	n.s.
February	-.0023	-0.037	0.042	n.s.
March	-.0245	-0.032	0.039	n.s.
April	0.0077	-0.040	0.041	n.s.
May	0.0018	-0.038	0.043	n.s.
June	0.0043	-0.035	0.041	n.s.
July	0.0348	-0.034	0.033	.025
August	0.0544	-0.034	0.035	.01
September	0.0013	-0.044	0.046	n.s.
October	0.0409	-0.037	0.043	n.s.
November	0.0630	-0.042	0.040	.001
December	0.0086	-0.035	0.038	n.s.
All of 1996	0.0015	-0.009	0.010	n.s.

Limitations of the Mantel Index

- Pearson-type correlation coefficient 極端值極度影響兩變數線性關係
  - Extreme values of either space or time could distort the relationship, either positively, if there are one or two observations that are extreme in *both* distance in time interval, or negatively, if there are only one or two observations that are extreme in *either* distance or in time interval.
- Less intuitive
  - the correlations tend to be small
- The sample size needs to be fairly large to produce as table estimate

Summary: Deficiencies of Knox and Mantel tests

- First, selection of an appropriate data transformation for the Mantel test, and of critical distances for Knox's test, is subjective
- Second, the Knox space critical distance is invariant with changing population density.
- Third, the model underlying Mantel's test is linear, but the relationship between space and time distances for almost all disease processes are expected to be non-linear.
- Fourth, Mantel's statistic is the sum of the products of the space and time distances, which will cause large distances to have undue influence on the statistic.
- Finally, results of the Knox and Mantel test vary as population density changes.

Labs: Sample Data (Patients.csv)

Patients.rar

Table 7.1 Kaposi's sarcoma in the West Nile district of Uganda. Locations of the homes and the date of onset of the 22 patients (data from McHardy *et al.*, 1984).

	Coordinates (km)		
Case No.	Eastings	Northings	Date of onset
1	266.8	334.3	1958
2	304.4	379.3	1959
3	265.5	315.0	1960
4	265.0	314.0	1960
5	264.2	323.0	1962
6	288.7	265.2	1962
7	290.2	294.3	1964
8	265.6	318.2	1964
9	263.7	344.4	1965
10	271.3	333.5	1966
11	267.4	344.4	1968
12	267.4	344.4	1968
13	276.5	344.6	1968
14	260.2	358.2	1971
15	264.0	296.8	1972
16	263.8	344.3	1972
17	300.5	373.0	1972
18	270.8	326.1	1973
19	258.7	344.8	1974
20	282.7	322.3	1974
21	265.3	314.9	1974
22	285.3	261.0	1974

Tango T. *Statistical Methods for Disease Clustering*, Springer, 2010.

R Functions for Mantel Test

```
Mantel.test<-function(x,y,time,c1,c2,Nrep)
```

```
------------------------------------------------------------------------
ARGUEMENTS
   x    : a vector of x-coordinates of case
   y    : a vector of y-coordinates of case
   time : a vector of observed times
   c1   : a constant for Mantel's measure of closeness in space
   c2   : a constant for Mantel's measure of closeness in time
   Nrep : The number of Monte Carlo replications, e.g., 999, 9999

VALUES
   Mantel.T          : test statistic
   Expected          : expected value of test statistic
   Simulated.p.value : simulated p-value
   Freq              : a vector of simulated test statistics under the null
```

Labs: R codes

```
# Knox Test
del1 <- 2; del2 <- 0
Nrep <- 999
out<- KnoxM.test(x,y,time,del1,del2,Nrep)
hist(out$Freq)
out$Simulated.p.value  # p=0.029

# Mantel Test
c1<- 1; c2 <- 1/5
Nrep<- 999
out<- Mantel.test(x,y,time,c1,c2,Nrep)
hist(out$Freq)
out$Mantel.T  # T=12.1751
out$Simulated.p.value   # p=0.032
```

3. Concept of Jacquez's k-Nearest Neighbor test (k-NN)

- The test statistic, J_k, is the count of the number of case pairs that are nearest neighbors in both space and time. When space-time interaction exists J_k will be large, since nearest neighbors in space will also tend to be nearest neighbors in time.

Jacquez's k-Nearest Neighbor test (k-NN)

$$J_k = \sum_{i=1}^{n} \sum_{j=1}^{n} n_{ijk}^{s} n_{ijk}^{t}$$

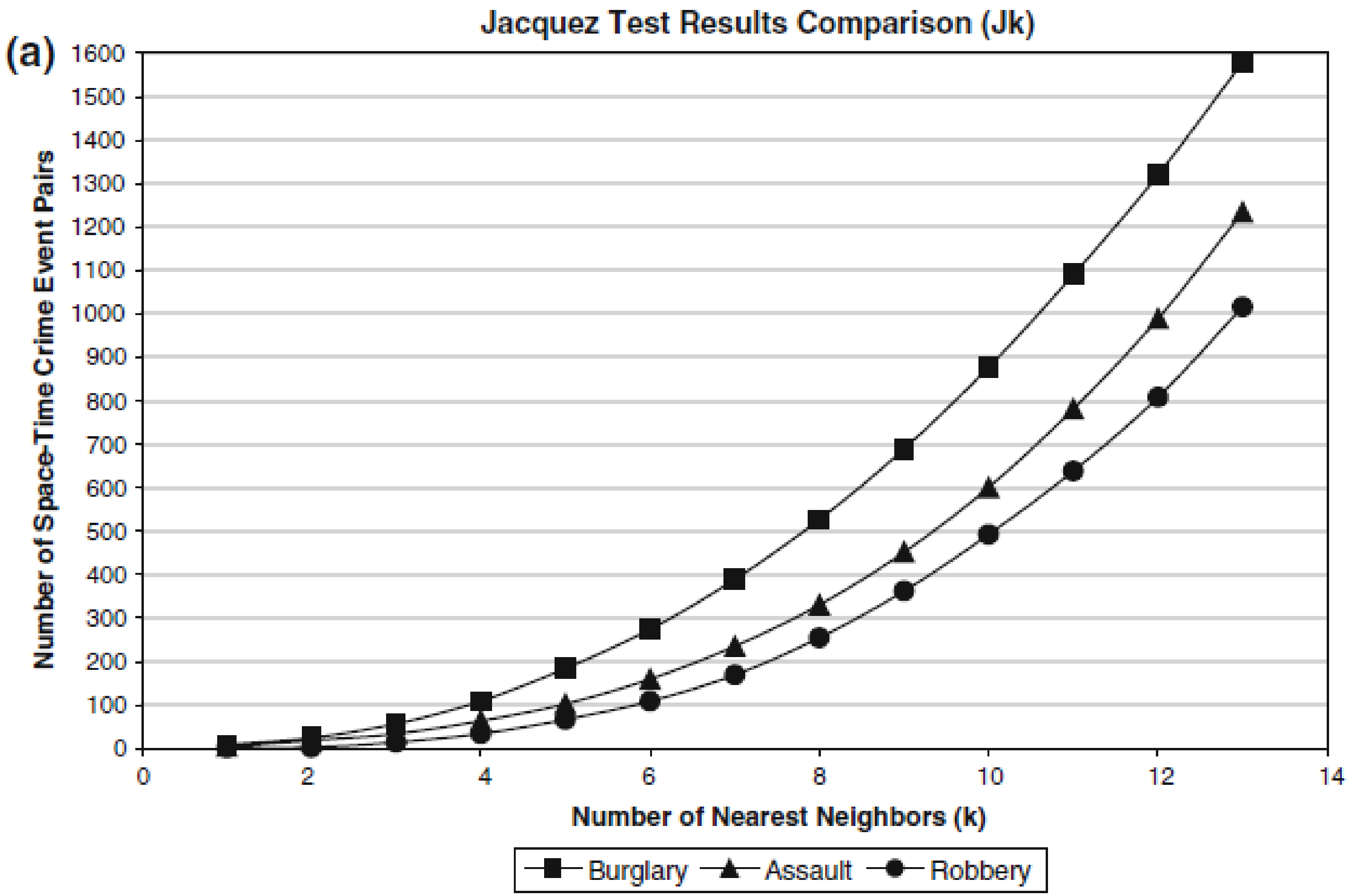
where: n = number of events; NN = nearest neighbor; k = the set of events as near or nearer to an event than the kth NN

$$n_{ijk}^{s} = \begin{cases} 1 & \text{if event } j \text{ is a } k \text{ NN of event } i \text{ in space} \\ 0 & \text{otherwise} \end{cases}$$

$$n_{ijk}^{t} = \begin{cases} 1 & \text{if event } j \text{ is a } k \text{ NN of event } i \text{ in time} \\ 0 & \text{otherwise} \end{cases}$$

H_o	Whether cases are nearest neighbors in space is independent of whether they are nearest neighbors in time
H_a	Nearest neighbors in space tend to be nearest neighbors in time.

Application



Grubesic and Mack (2008)

R Functions for Jacquez's k-Nearest Neighbor test

```
Jacquez.test<-function(x,y,time,k,Nrep)
```

```
--------------------------------------------------------------------
 ARGUEMENTS
    x    : a vector of x-coordinates of case
    y    : a vector of y-coordinates of case
    time : a vector of observed times
    k    : k of k nearest neighbors
    Nrep : The number of Monte Carlo replications, e.g., 999, 9999

 VALUES
    Jacquez.T           : test statistic
    Expected            : expected value of test statistic
    Simulated.p.value : simulated p-value
    Freq                : a vector of simulated test statistics under the null
```

Labs: R code

```
k<- 1; Nrep<- 999
time<-time+runif(22)/100 #small random numbers were added
out<- Jacquez.test(x,y,time,k,Nrep)
out$Jacquez.T  #J = 0.5
out$Expected # Exp[J]= 0.5238095
out$Simulated.p.value # p= 0.575
```

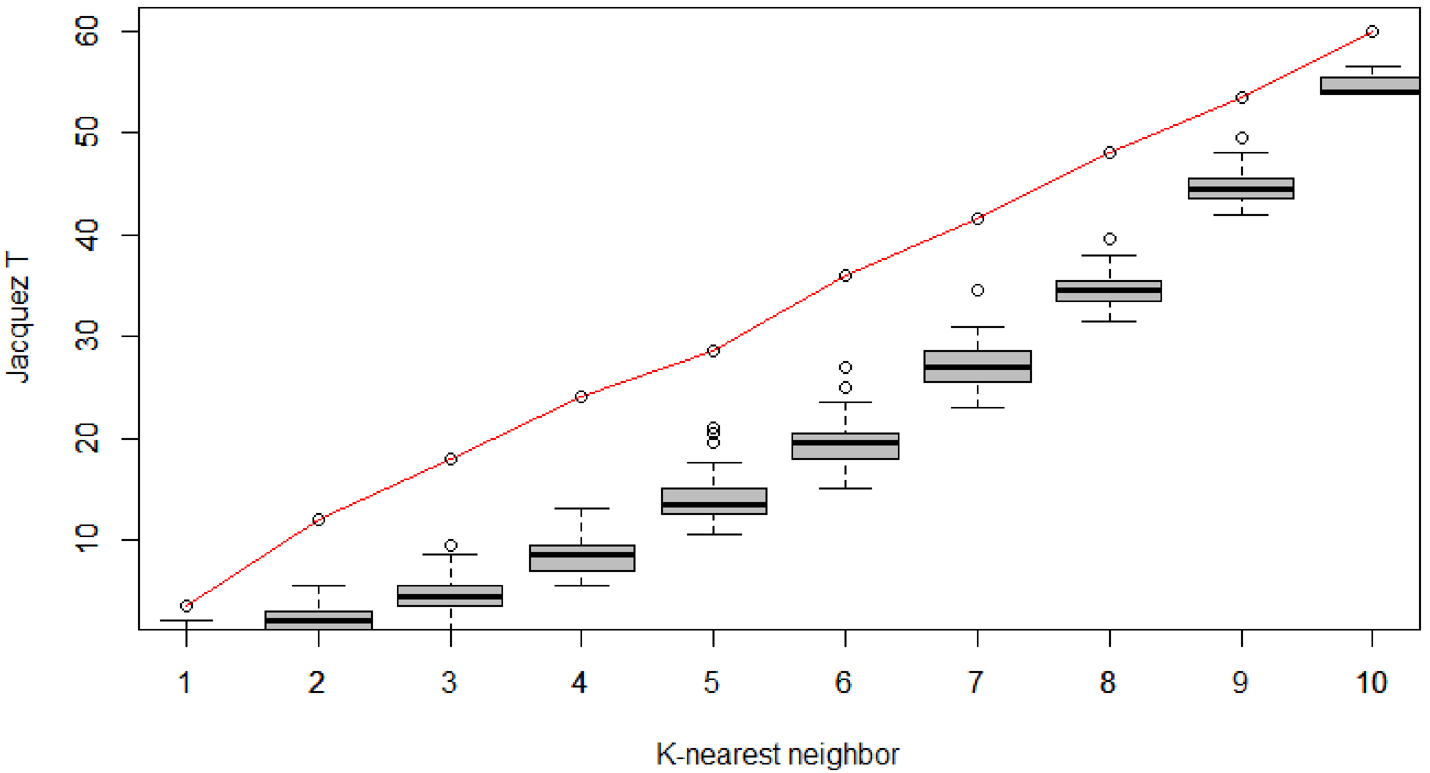
Table 7.9 Jacquez's k-NN approach with Monte Carlo p-value with $N_{\text{rep}} = 999$ for 22 patients with Kaposi's sarcoma in the West Nile district of Uganda (data from McHardy *et al.*, 1984).

	k nearest neighbors									
	1	2	3	4	5	6	7	8	9	10
T	0.5	2.5	4.0	8.0	13.5	20.0	26.0	34.5	43.0	52.5
Expected	0.52	2.09	4.71	8.38	13.09	18.85	25.66	33.52	42.42	52.38
Monte Carlo *p*	0.575	0.396	0.678	0.582	0.419	0.334	0.431	0.320	0.378	0.438

Jacquez's k-Nearest Neighbor test (k-NN)

(作業 1-3)

Data: point3.shp



Another k-NN test (k-specific test)

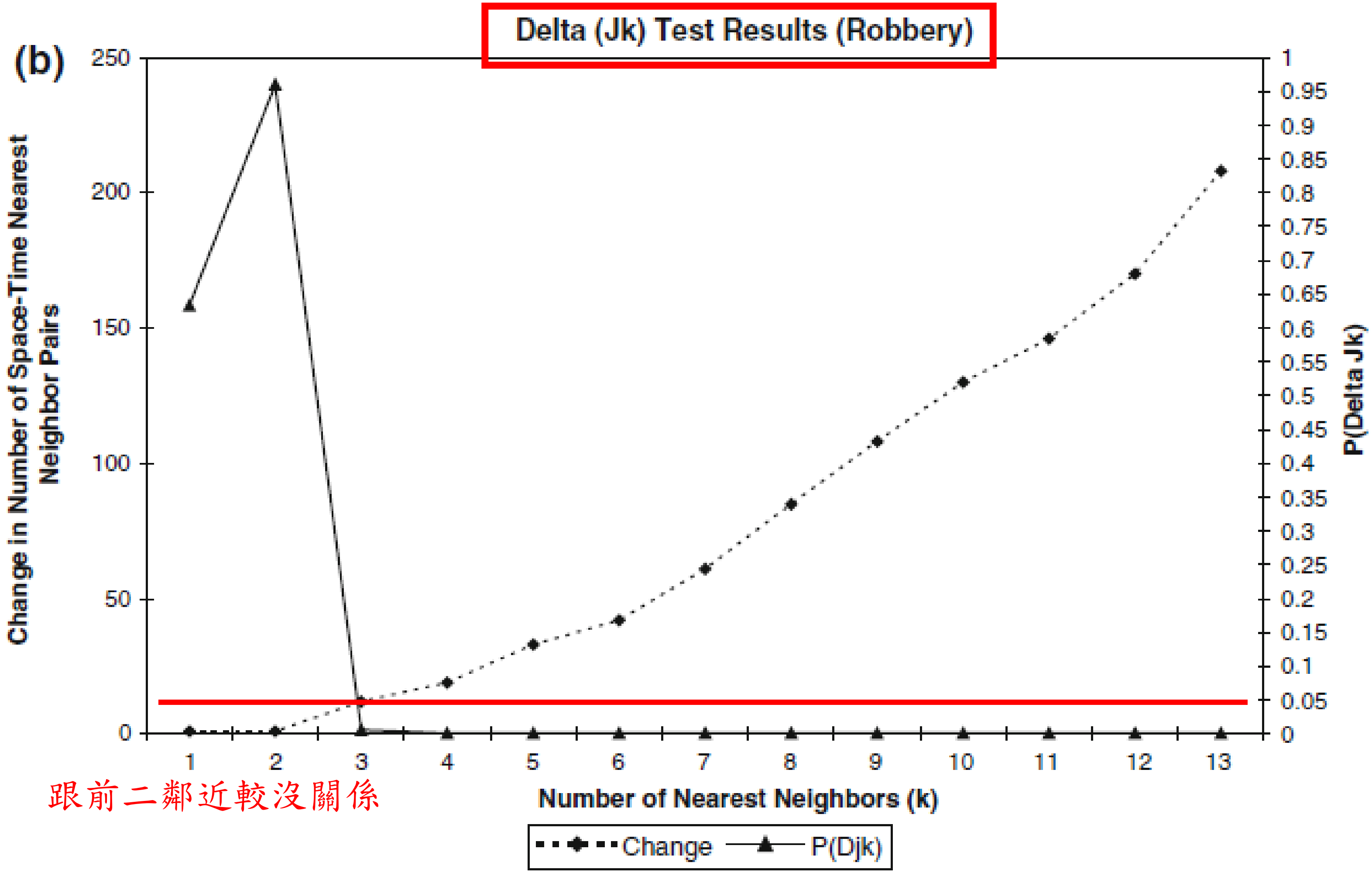
- **J_k are not independent** because all pairs of events included in smaller neighborhood definitions, k, are also included in subsequent, larger values of k.
- For example, as k increases from three to four, all event pairs that were included in k = 3 are also included in k = 4. Thus, larger values of k suggest increased levels of spatio-temporal interaction, which is not surprising given the increased likelihood of becoming a space–time neighbor as the number of events increase in the space and time matrices.
- To account for this statistical issue Jacquez (1996) provides a k-specific test statistic for measuring time–space interaction beyond that found for the k-1 nearest neighbors:

k-specific Test

$$\Delta J_k = J_k - J_{k-1}$$

- The index tracks any increases in the number of space–time nearest neighbors as k is increased and unlike the J_k, **the index are statistically independent**.
- In other words, if one was to change the neighborhood definition by increasing the number of neighbors under consideration (e.g. three to four), **the index monitors these changes for significance.**

Application (cont'd)



It means robbery begins to cluster at a different spatio-temporal scale (k=3).

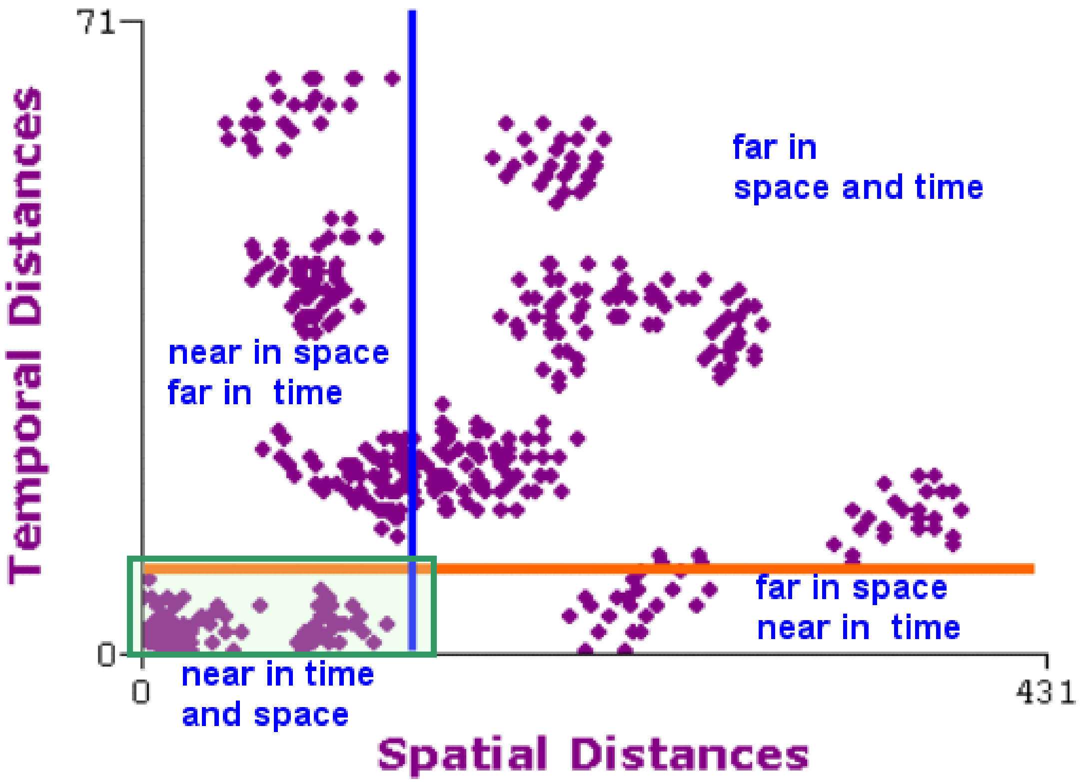
Grubesic and Mack (2008)

Applications: Implications

- Results: burglary: k=1; assault: k=2; robbery: k=3
- Burglary events have significant spatio-temporal interaction with their first-order nearest neighbors. It suggests that closeness in space and time exists between events.
- This can lead to an initial burglary at one property, followed by subsequent burglaries at neighboring properties.
- Predatory crime often occur near bars, taverns, bus stops and homeless shelters. These types of facilities have a different spatial distribution than residences, and thus, the crimes committed in these locations have unique temporal characteristics

Grubesic and Mack (2008)

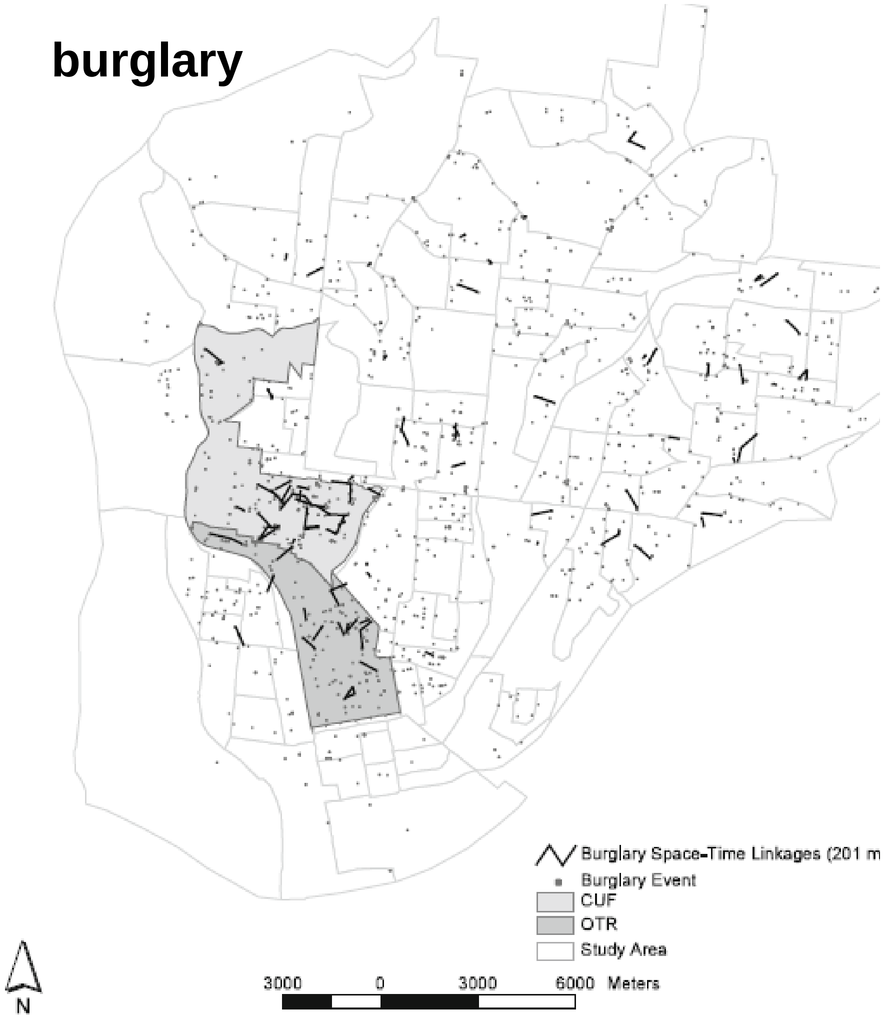
4. Mapping the Areas with Spatial-temporal Interaction: Space-time Linkages



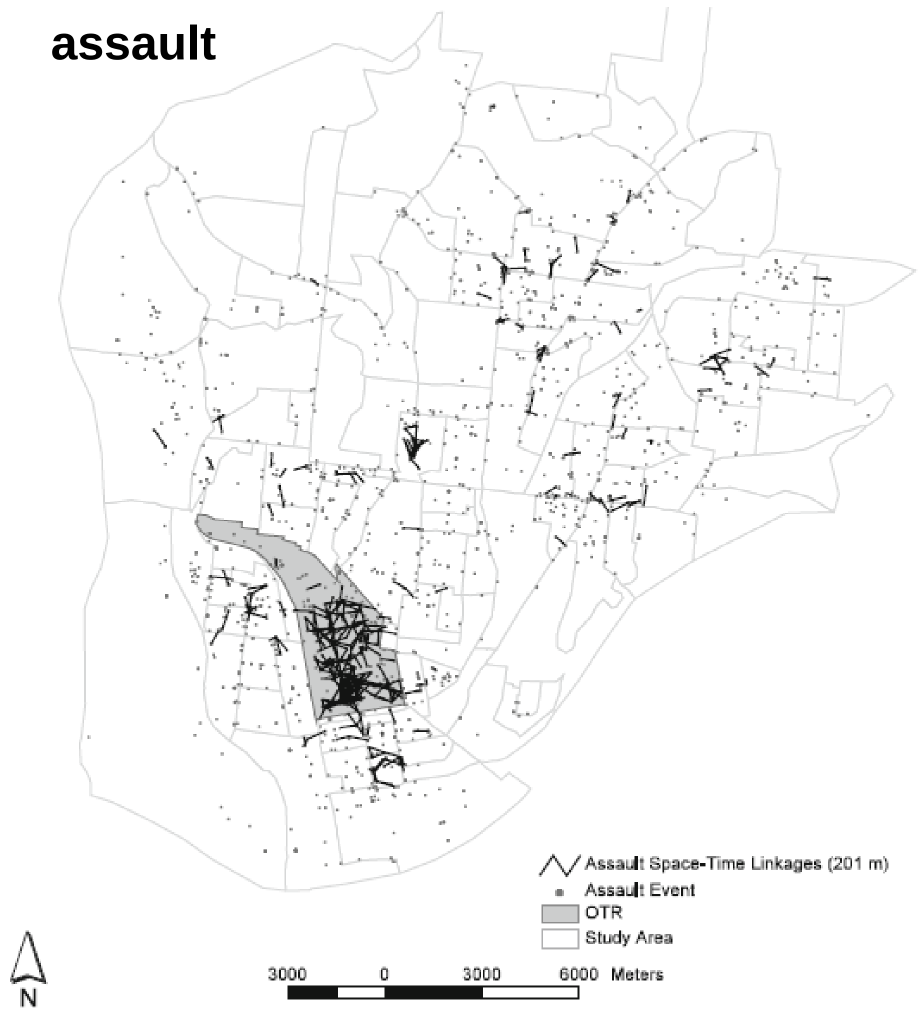
Each point of the space-time scatter plot means ?

Spatial Patterns of Space-time Links

burglary



assault

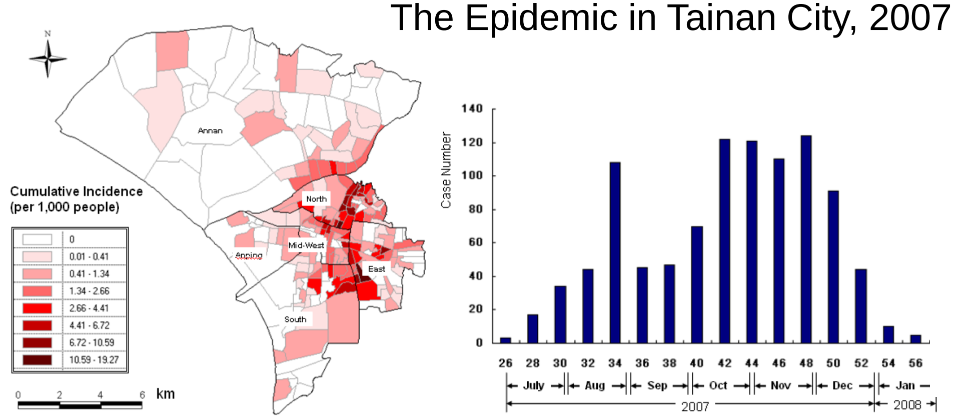


Grubesic and Mack (2008)

Analyzing the patterns of spatial-temporal diffusion of an epidemic

- **Population Movement** and Vector-borne Disease Transmission: **Differentiating Spatial-temporal Diffusion Patterns** of Commuting and Non-commuting Dengue Cases. *Annals of the Association of American Geographers,* 102(5):1026-1037

The Epidemic in Tainan City, 2007



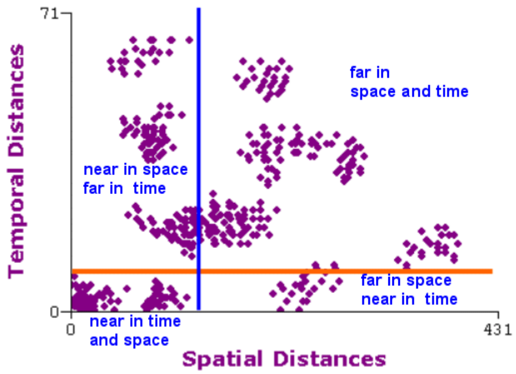
Hypothesis: Diffusion through Dengue-infected Commuter vs. Non-commuters

	Through Commuters	Through Non-Commuters
病例間的地理距離	較大	較小
病例間的發病時間	較長	較短
地理擴散的速度	快	慢
可能影響因素	人口接觸密度	環境因子

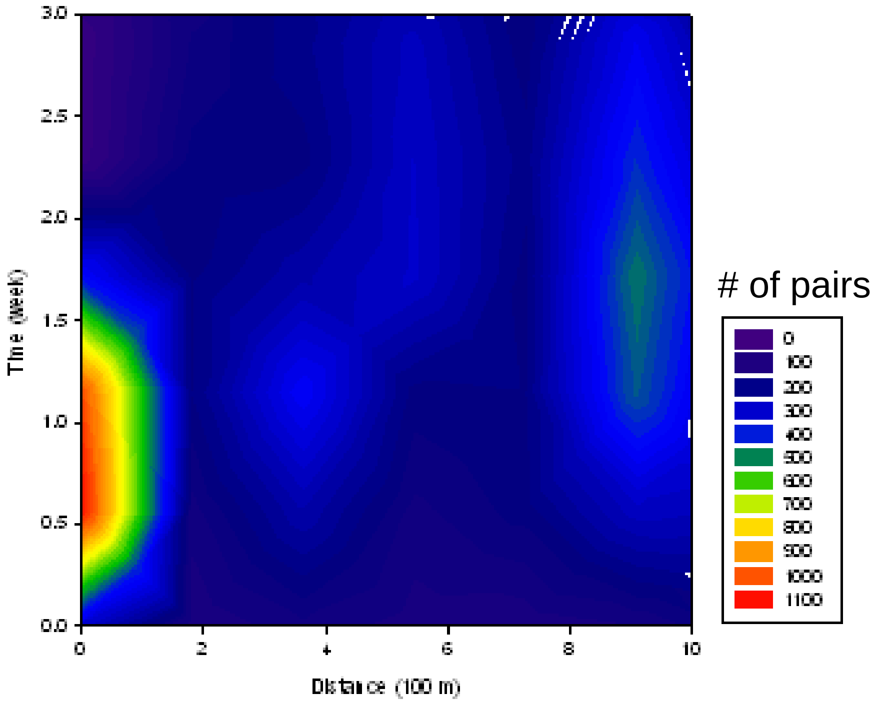
Space-time Clustering Analysis

Time-to-event Analysis

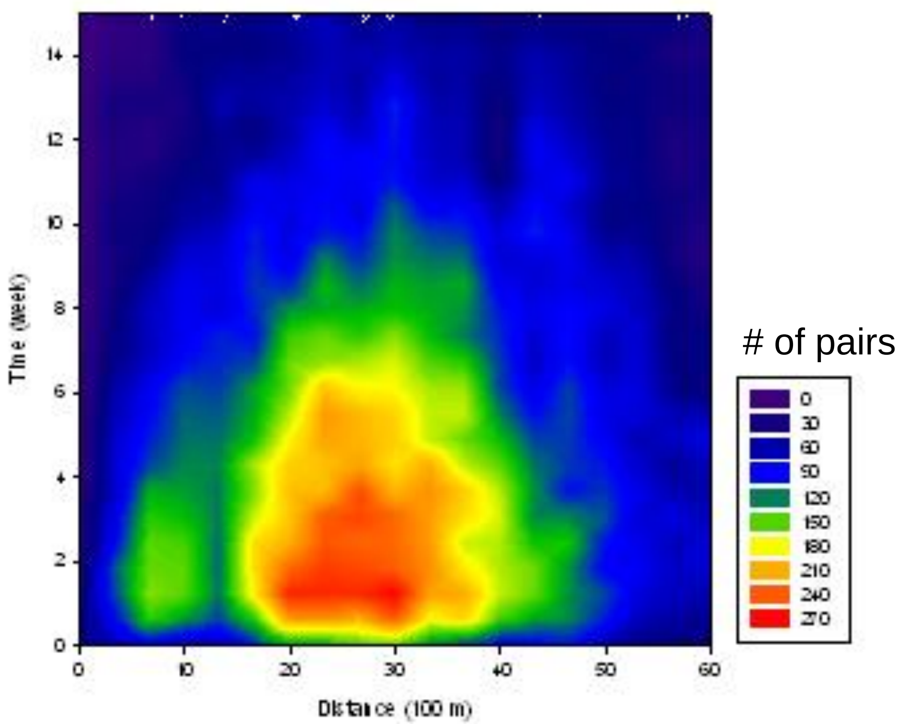
Spatial-Temporal Distance Diagrams



Non-commuting DF cases



Commuting DF cases

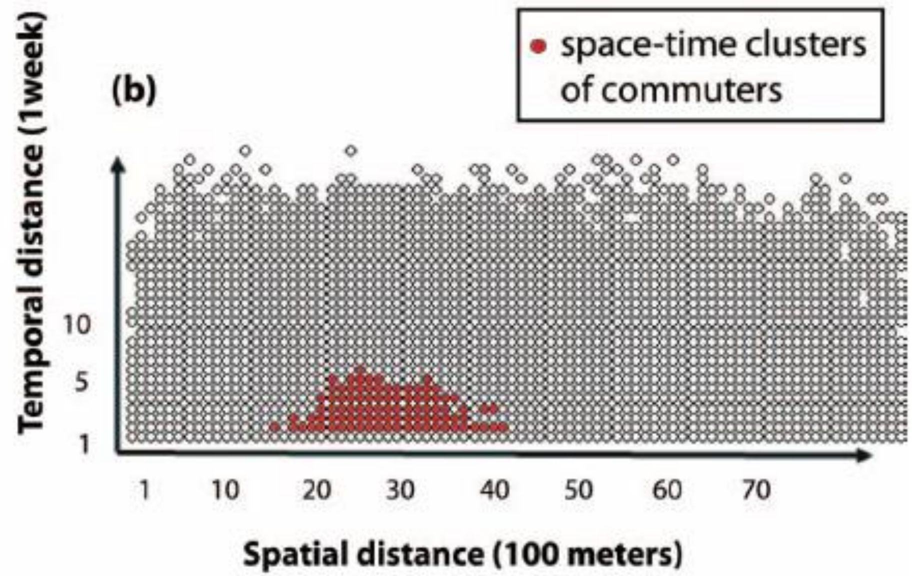


Time (0.5-1), Distance (0-1)
A total of 1308 pairs (links)

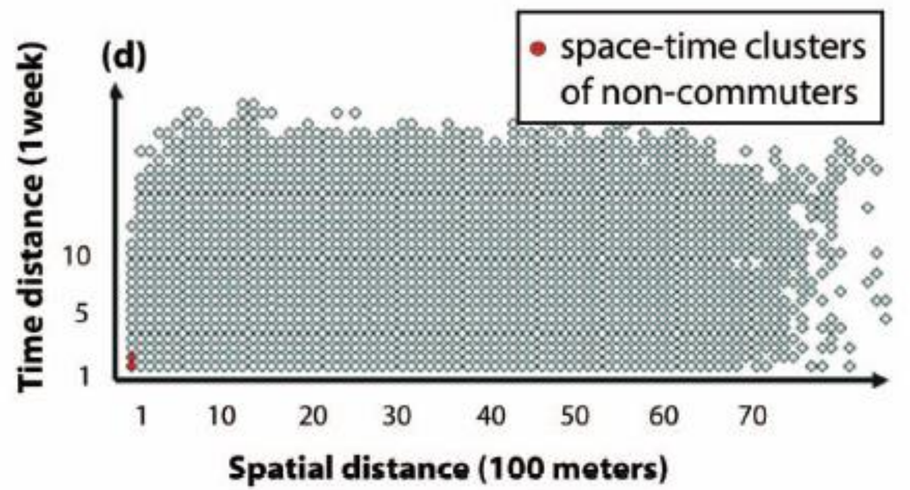
Time (0.5-4.5), Distance (20-30)
A total of 2764 pairs (links)

Identifying Space-time Clusters

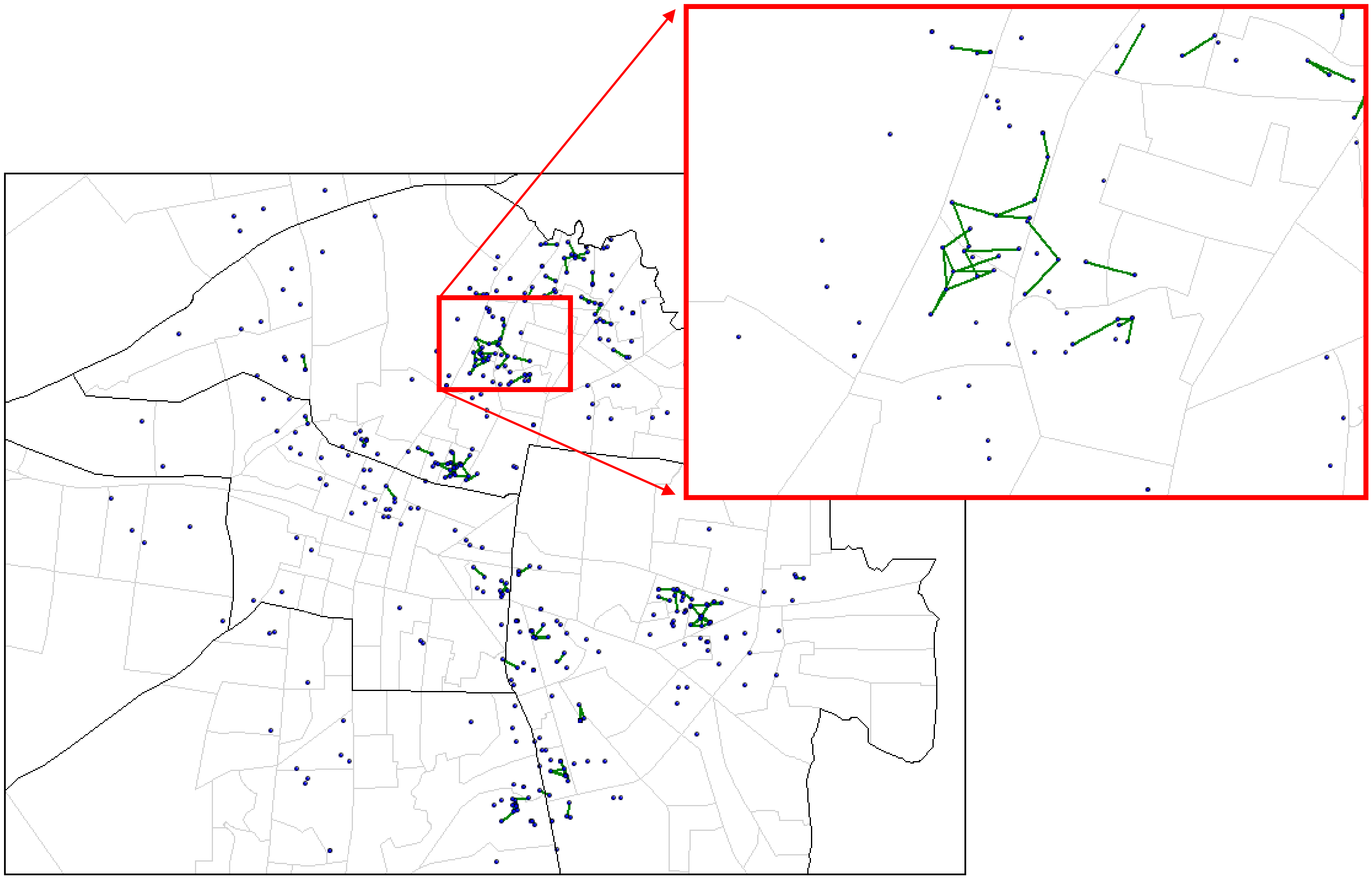
Time (0.5-4.5), Distance (20-30)
A total of 2764 pairs (links)



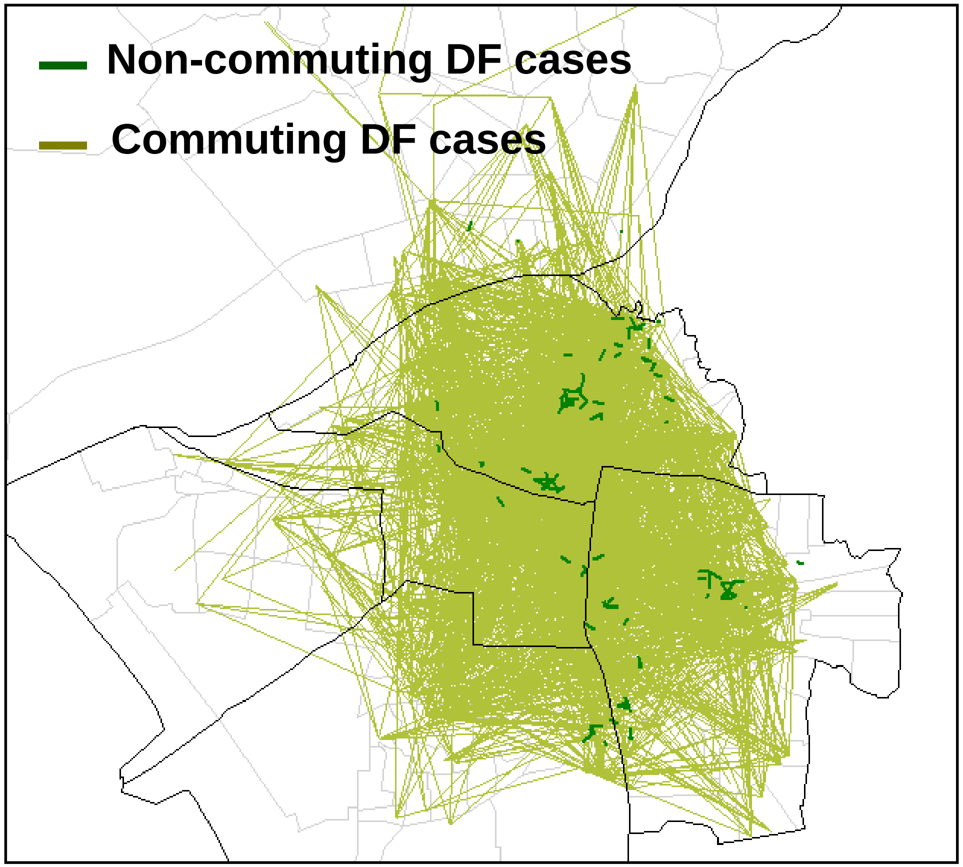
Time (0.5-1), Distance (0-1)
A total of 1308 pairs (links)

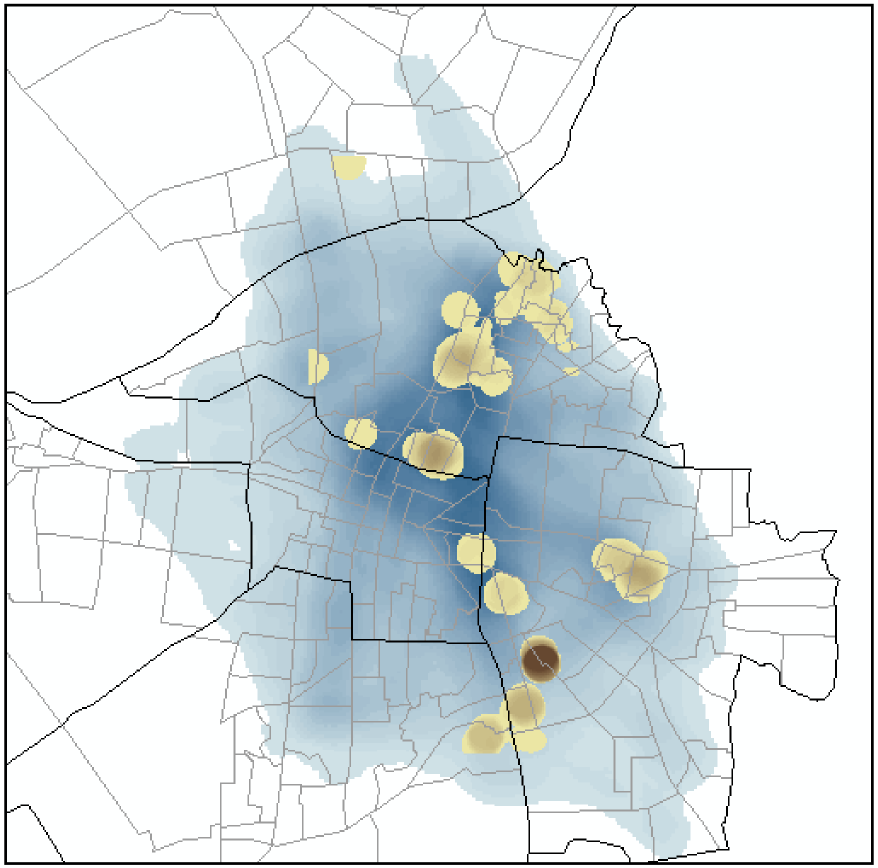


Space-time Links



Spatial Patterns of Space-time Links

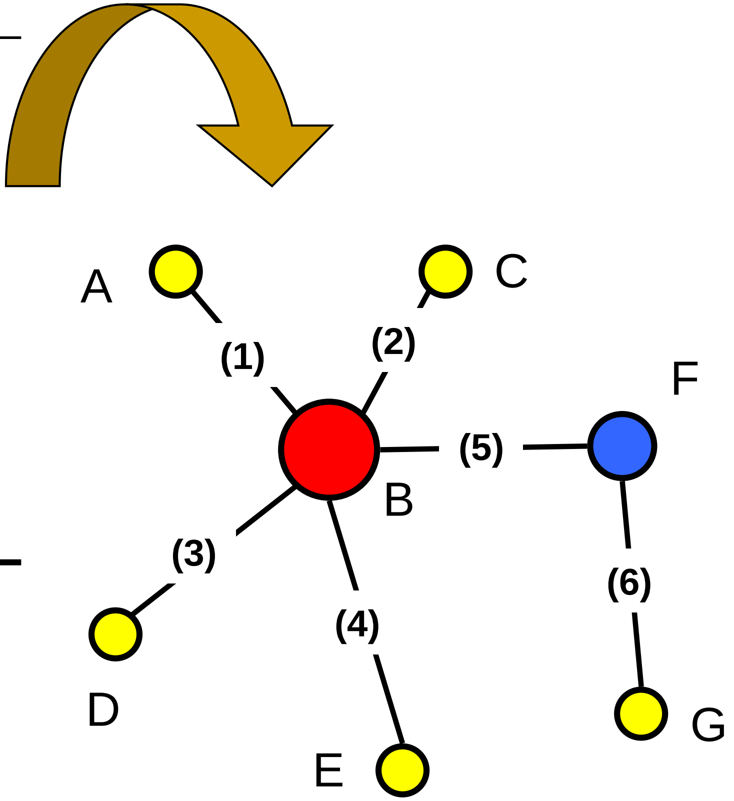




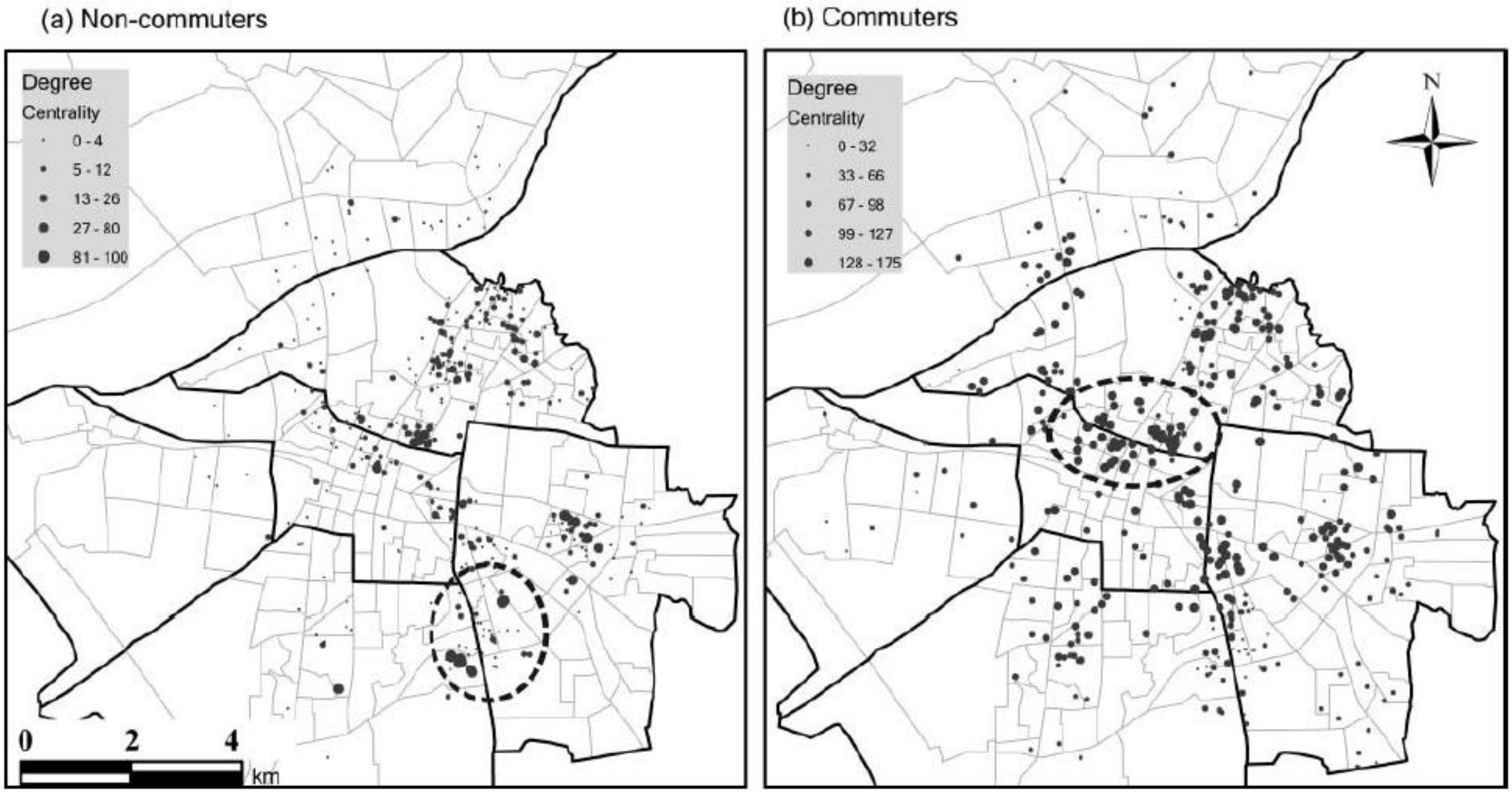
Network Formation (An Example)

Pairs of selected space-time links

Pair #	Node ID	Node ID	Space (km)	Time (weeks)
(1)	B	A	0.7	1
(2)	B	C	1.2	1.5
(3)	B	D	1.1	0.8
(4)	B	E	1.3	1.1
(5)	B	F	1.5	1.6
(6)	G	F	1.8	0.7



Identifying the origins of infection



5. Space-time K-function: Concepts

Traditional (spatial) K-function (Ripley, 1977)

$$K(h) = \frac{1}{\lambda} E(\#(\text{events w/in distance h of randomly chosen event})$$

Space-time K-function (Diggle, et al. 1995)

$$K(h,t) = \frac{1}{\lambda} E(\#(\text{events w/in distance h and time t of randomly chosen event})$$

Space-time K-function: Equations

Spatial K-function

$$\hat{K}_D(d) = \frac{A}{n^2} \sum_i \sum_{j \neq i} \frac{I_d(d_{ij})}{w_{ij}}$$

Temporal K-function

$$\hat{K}_T(t) = \frac{T}{n^2} \sum_i \sum_{j \neq i} \frac{I_t(t_{ij})}{v_{ij}}$$

Space-time K-function

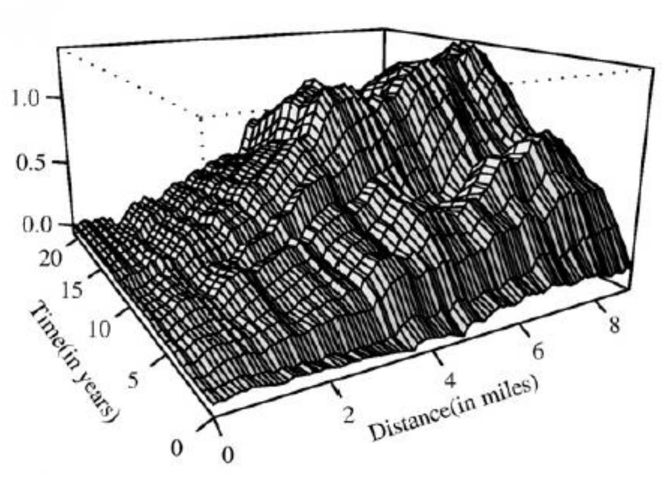
$$\hat{K}(d,t) = \frac{AT}{n^2} \sum_i \sum_{j \neq i} \frac{I_d(d_{ij}) I_t(t_{ij})}{w_{ij} v_{ij}}$$

Measuring Space-time Interaction

A, B 獨立事件 $P(A \cap B) = P(A) \cdot P(B)$

No Space-time Interaction: $K(d,t) = K_D(d)K_T(t)$

Performance Indices:



Residuals:

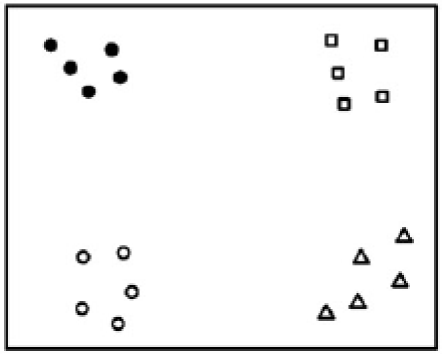
$$\hat{D}(d,t) = \hat{K}(d,t) - \hat{K}_D(d)\hat{K}_T(t)$$

Relative function:

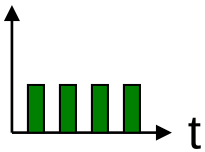
$$\hat{D}_0(d,t) = \hat{D}(d,t) \,/\, \{\hat{K}_D(d)\hat{K}_T(t)\}$$

Some stylized space–time distributions

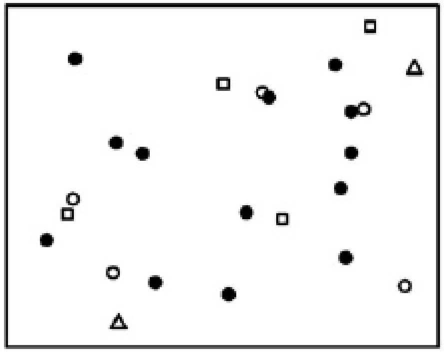
(1)



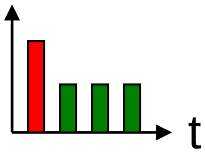
Spatial clustering (+)
Temporal clustering (-)
Space-time interaction (-)



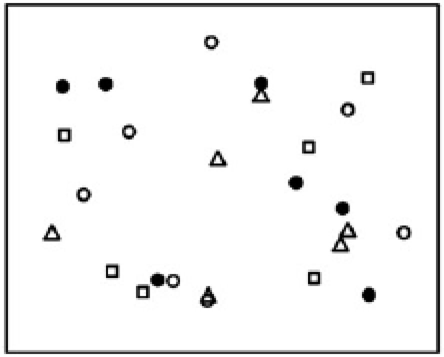
(2)



Spatial clustering (-)
Temporal clustering (+)
Space-time interaction (-)

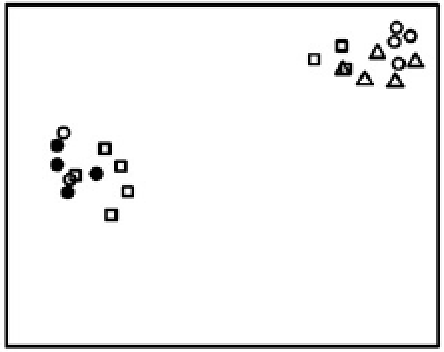


(3)

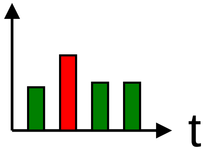


Spatial clustering (-)
Temporal clustering (-)
Space-time interaction (-)

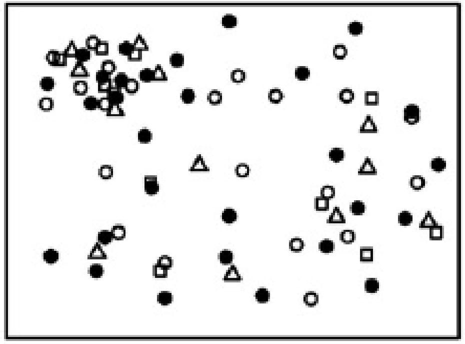
(4)



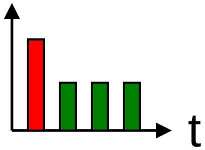
Spatial clustering (+)
Temporal clustering (+)
Space-time interaction (+)

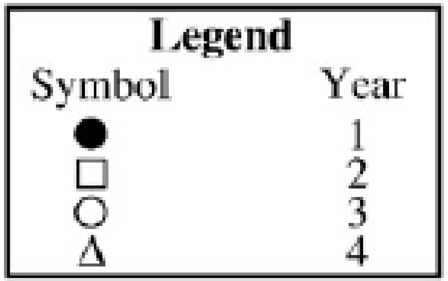


(5)



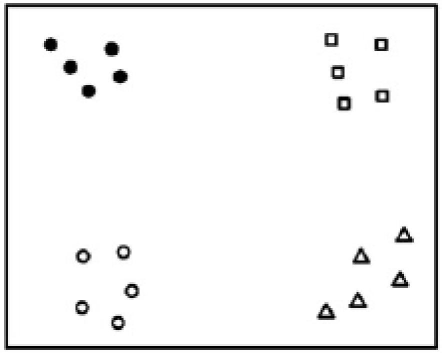
Spatial clustering (+)
Temporal clustering (+)
Space-time interaction (-)

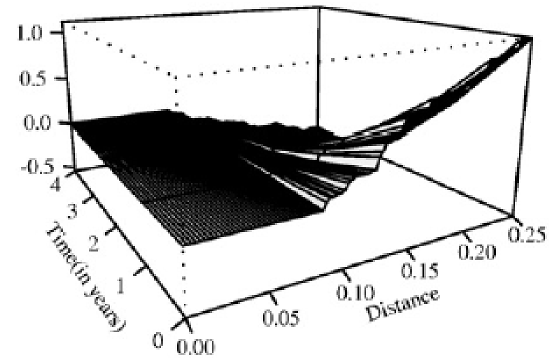


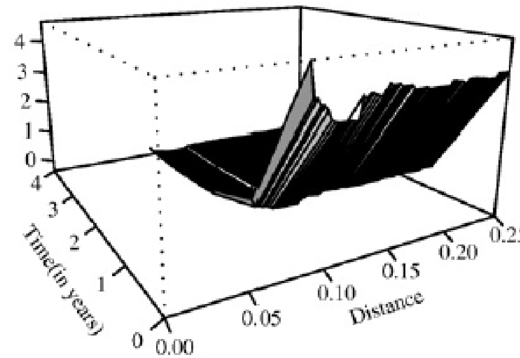


(1)

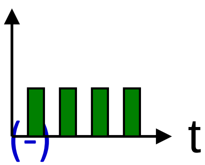
$D(d,t)$ $D_0(d,t)$



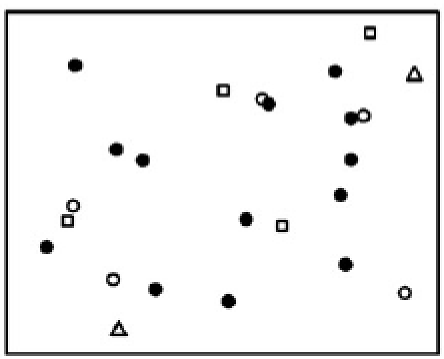


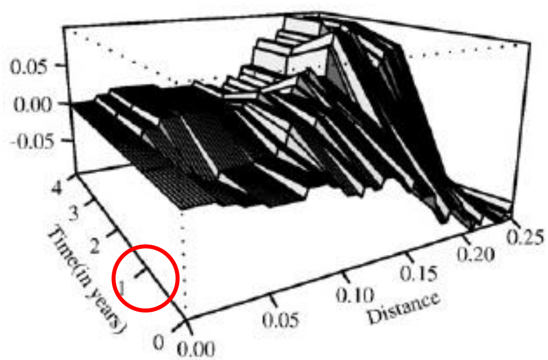


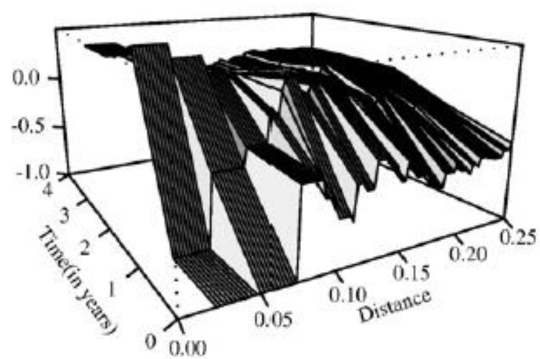
Spatial clustering (+)
Temporal clustering (-)
Space-time interaction (-)



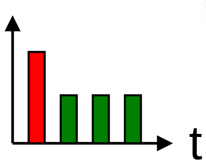
(2)





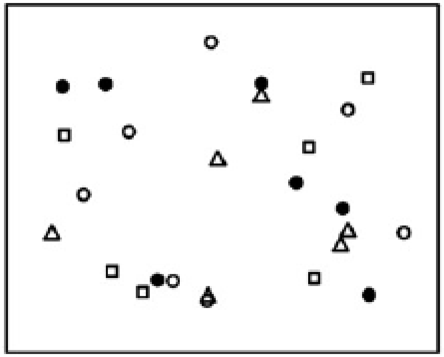


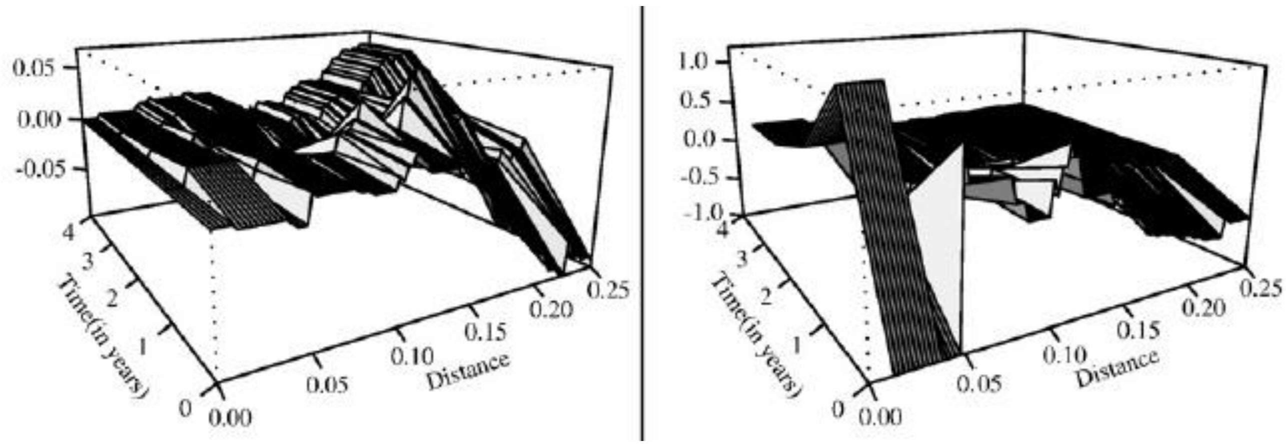
Spatial clustering (-)
Temporal clustering (+)
Space-time interaction (-)



(3)

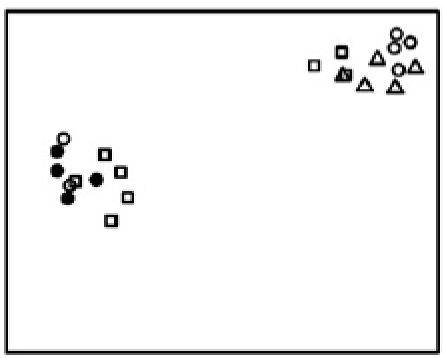
$D(d,t)$ $D_0(d,t)$

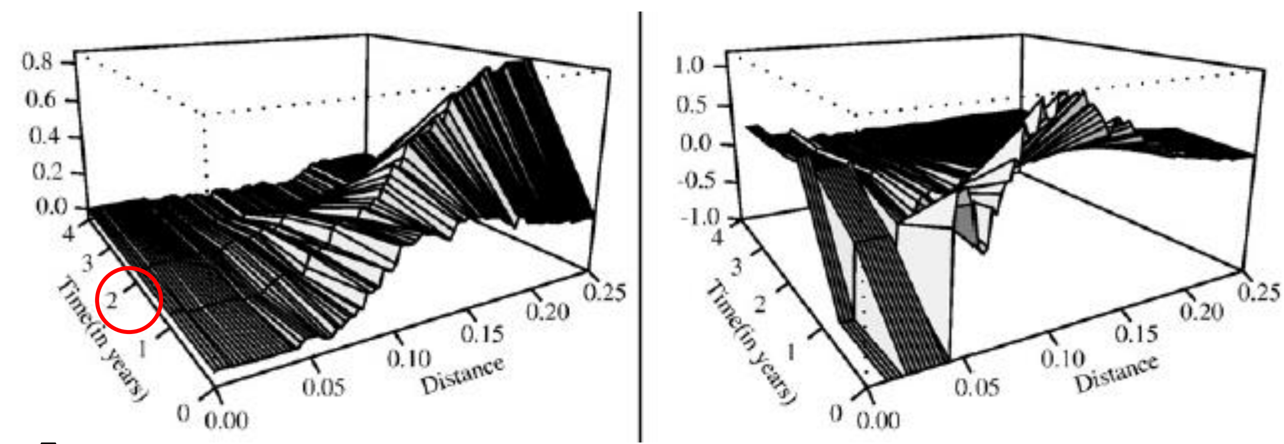




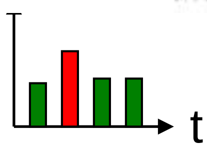
Spatial clustering (-)
Temporal clustering (-)
Space-time interaction (-)

(4)

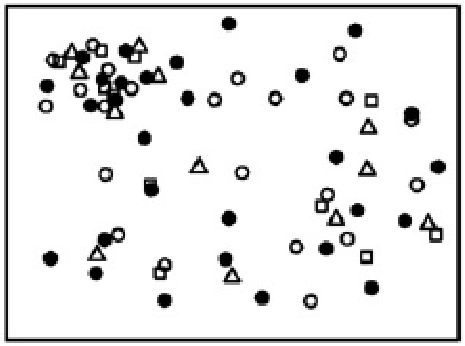




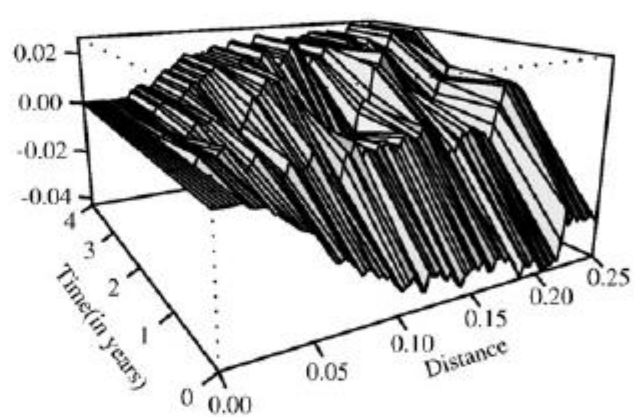
Spatial clustering (+)
Temporal clustering (+)
Space-time interaction (+)



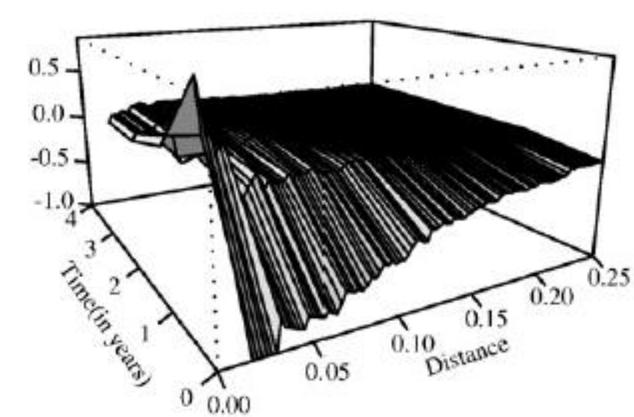
(5)



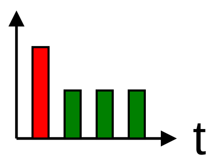
$D(d,t)$



$D_0(d,t)$



Spatial clustering (+)
Temporal clustering (+)
Space-time interaction (-)



Each year individually produces a visual impression of clustering; however, looking at the whole map without distinguishing between the different time periods the visual impression is that of randomness.

Testing Statistical Significance

1. standardized residuals

Standard error

$$\hat{R}(d,t) = \hat{D}(d,t) \,/\, \sqrt{\hat{V}(d,t)}.$$

In the absence of any space–time interaction, these residuals have zero expectation and a variance equal to one and approximately 95% of the values of *R(d,t)* would lie within two standard errors (French et al., 2005).

2. Monte Carlo testing

Actual observed sum of $D_i(d,t)$ overall d and t. If the observed sum is ranked above 95 out of 100 simulated values then the probability that the observed space-time interaction occurred by pure chance is less than 5%..

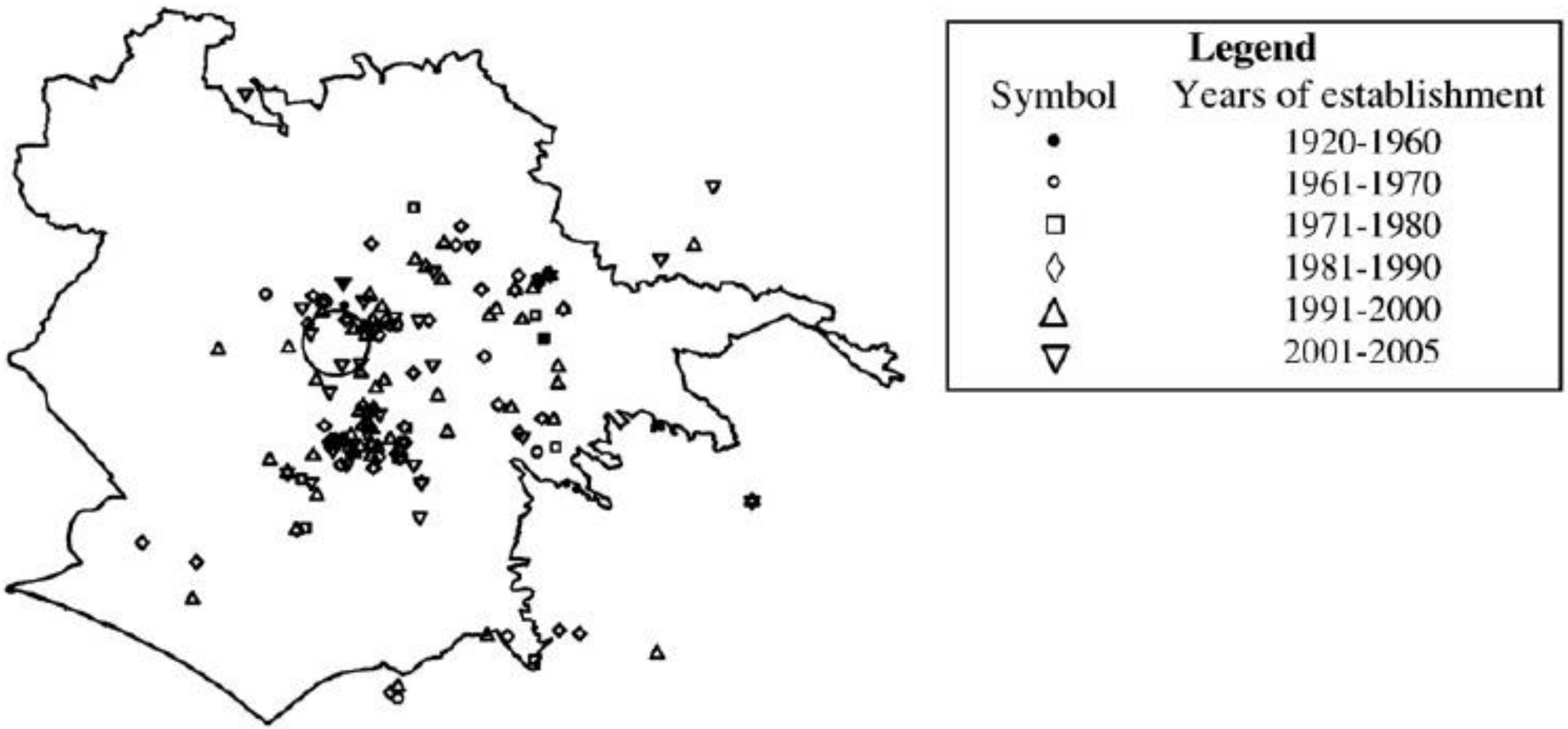
Application: ICT industries in Rome (Italy) 1920–2005

Arbiaa et al (2010)

Year of establishment	Number of firms		
	(1) Electronic and Communication	(2) Information Technology	(3) ICT = (1) + (2)
1920–1960	5	0	5
1961–1970	5	1	6
1971–1980	4	10	14
1981–1990	15	30	45
1991–2000	29	37	66
2001–2005	8	25	33
Total	66	103	169

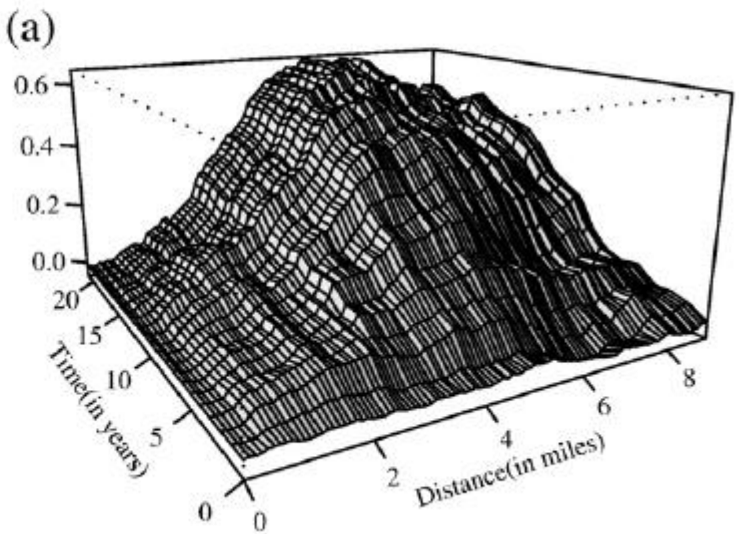
1995, widespread internet → 1991–2000

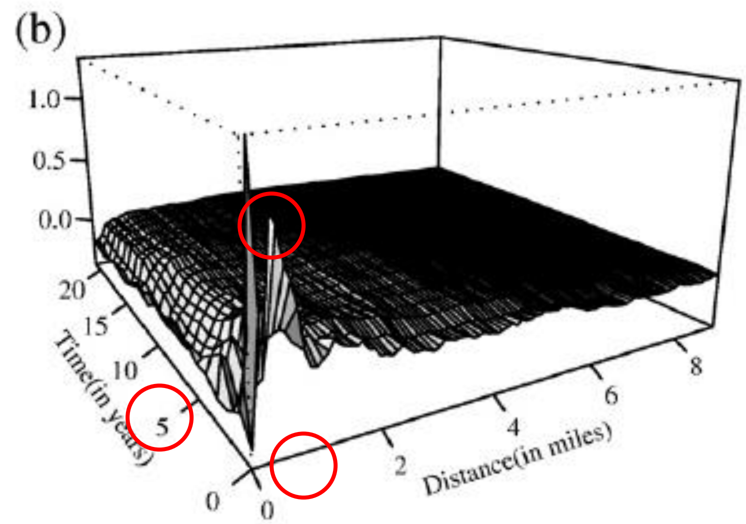
Spatial-temporal distribution of ICT industries



Space-time K-function

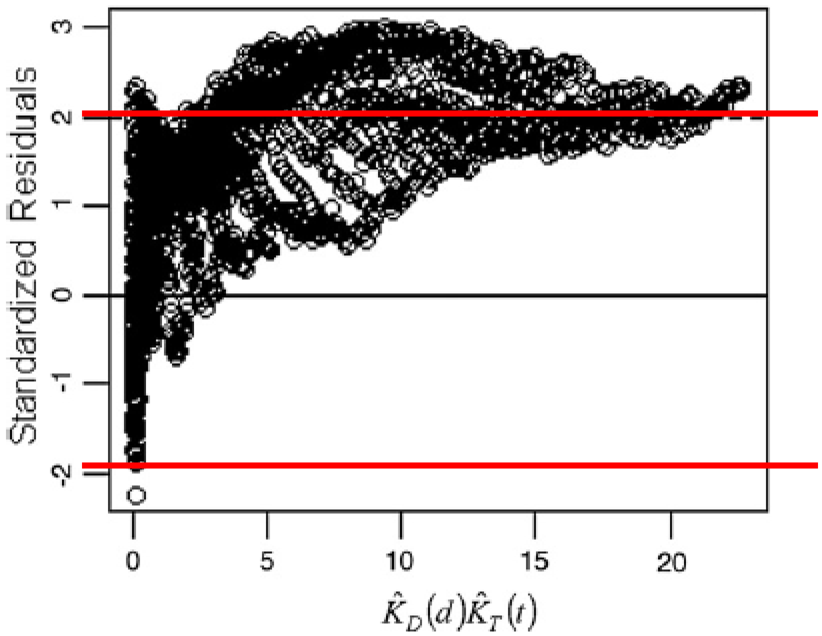
D(d,t)　　　　*$D_0(d,t)$*

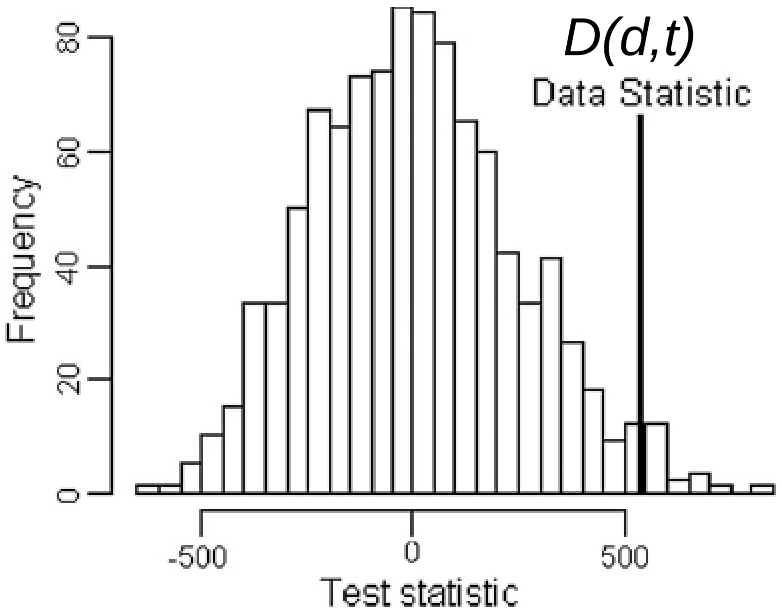




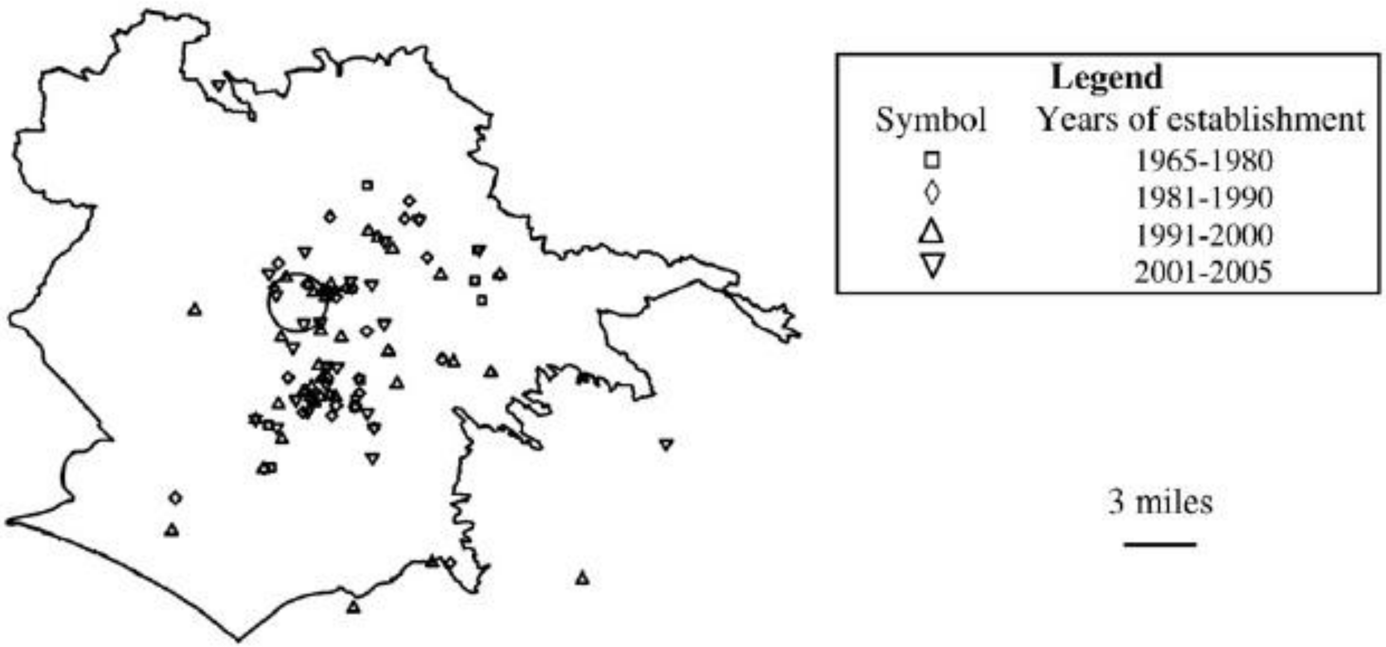
This shows that the underlying concentration phenomenon tends to drive clusters with a small spatial magnitude (circles with radius of 1 mile) and where the firms are temporally correlated in terms of the year of establishment.

Standardized residuals and Monte Carlo testing



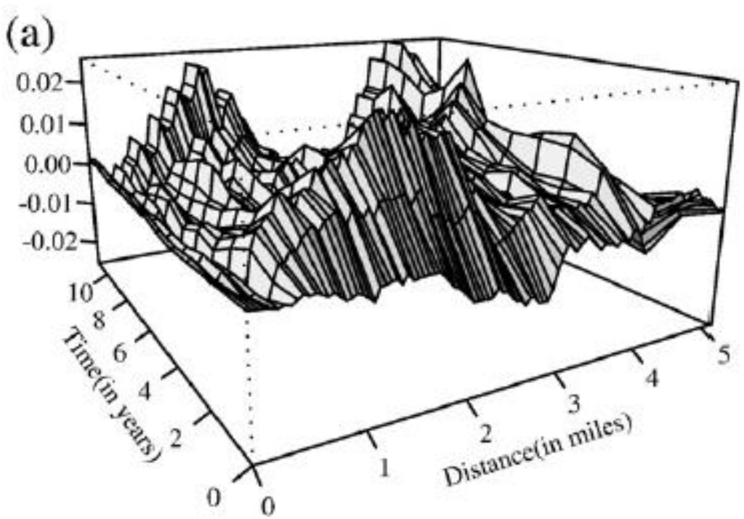


Spatial-temporal distribution of Information Technology sector

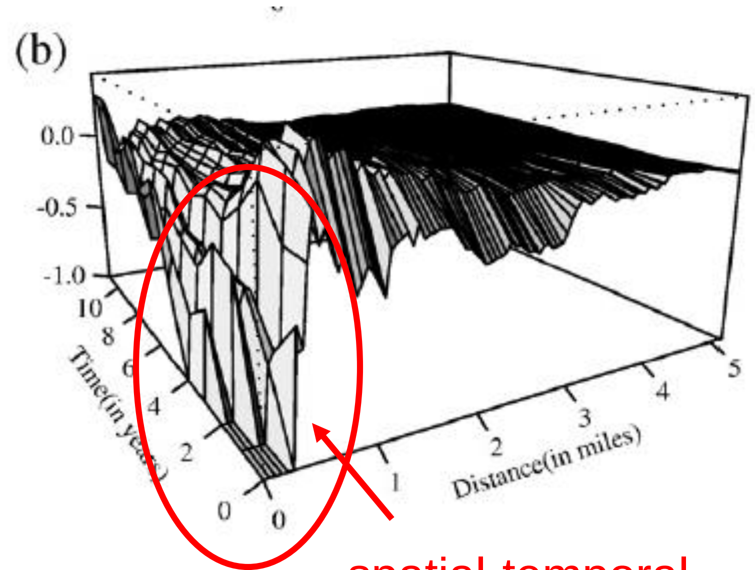


Space-time K-function

D(d,t)



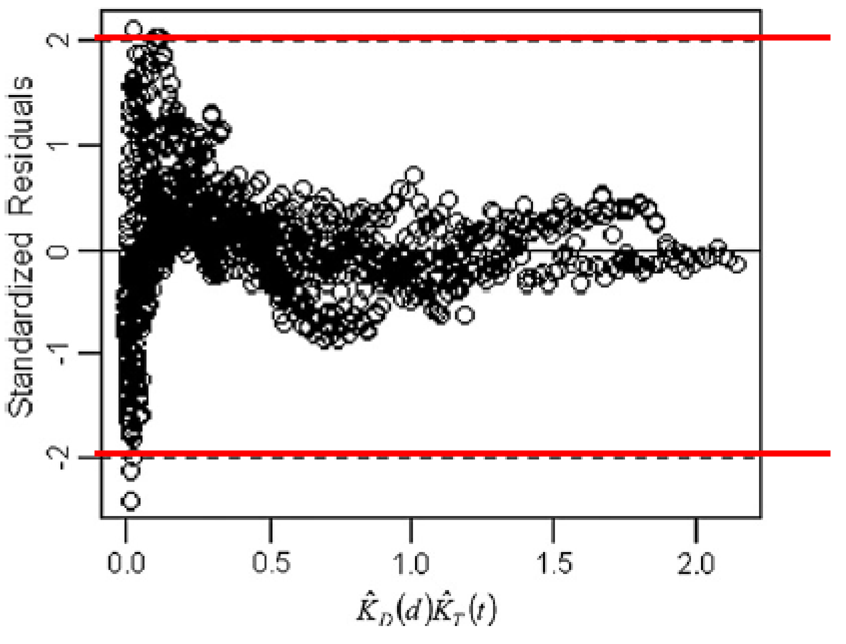
$D_0(d,t)$

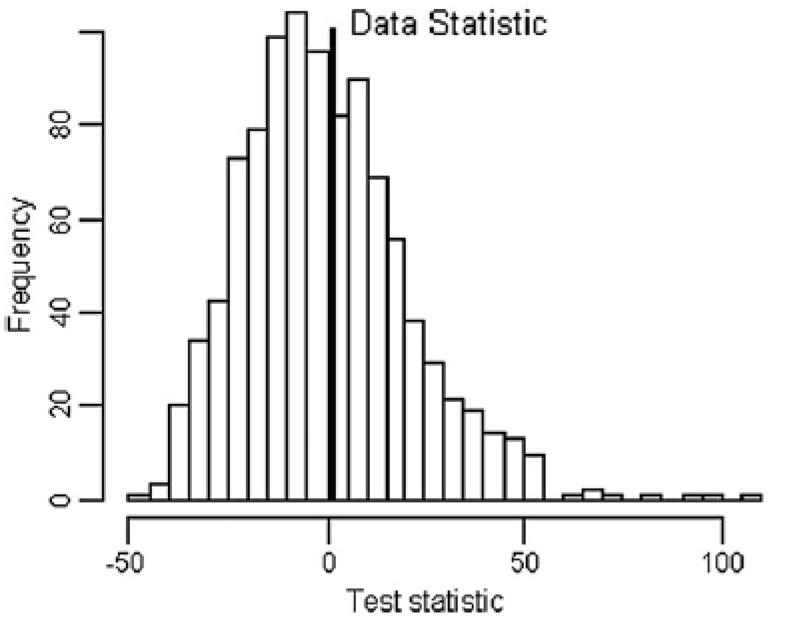


spatial-temporal segregation

$D_0(d,t)$ displays a rather less marked spatial clustering and a negative time cluster.

Standardized residuals and Monte Carlo testing





Space-time clustering situation of the ICT sector before and after widespread internet availability

Characteristics	Before widespread internet (before 1995)	After widespread internet
Spatial lag	1 mile	No space–time interaction
Temporal lag	1 year	–
p-value Monte Carlo test	0.011	0.592
% of estimated residuals out of ±2 SE	40.1%	0.50%

Application: RVF disease transmission in 2008-2011

Metras R, et al (2012)

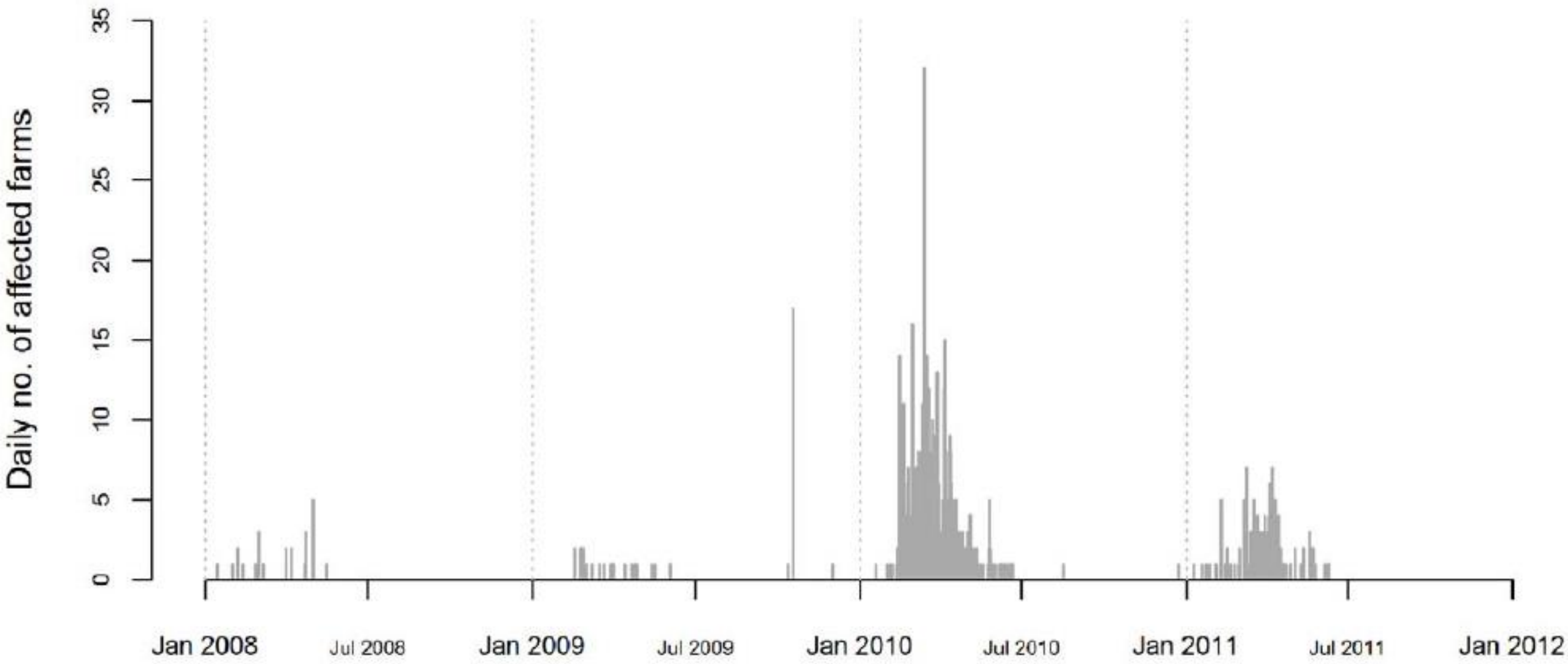
- Rift Valley fever (RVF) is a zoonotic arbovirosis for which the primary hosts are domestic livestock (cattle, sheep and goats).
- Mechanisms for short and long distance transmission have been hypothesized, but there is little supporting evidence.
- We investigate the presence of a contagious process in order to generate hypotheses on the different mechanisms of transmission.

Disease data in time and space

Table 1. Number of affected farms (%) per outbreak wave, by on-farm species.

	Number of affected farms (%)					
On-farm species	**2008**	**2009, wave 1**	**2009, wave 2**	**2010**	**2011**	**All years**
CA	21 (87.5)	18 (90.0)	6 (31.6)	62 (13.2)	19 (15.3)	126 (19.1)
SR	3 (12.5)	2 (10.0)	3 (15.8)	232 (49.3)	100 (80.6)	340 (51.7)
SR+CA	-	-	10 (52.6)	177 (37.6)	5 (4.0)	192 (29.2)
Total per year (100%)	24	20	19	471	124	658

SR = small ruminants, CA = cattle.



Results

$D_0(d,t)$: Excess Risk

$$\hat{D}_0(d,t) = \hat{D}(d,t) \,/\, \{\hat{K}_D(d)\hat{K}_T(t)\}$$

Table 3. Excess risk attributed to the space-time interactions ($D_o(s,t)$), and corresponding *p*-values, by wave.

	Separating distances		Results			
Year (wave)	**Time (60 days)**	**Space (300 km)**	$D_o(s,t)$	**Upper time window**	**Upper space window**	***p*-value**
2008	2 days	5 km	>2	9 days	15 km	0.091
			>1	35 days	50 km	
2009 (1)	5 days	10 km	>3	1 day	20 km	0.008
			>2	11 days	30 km	
			>1	31 days	40 km	
2009 (2)	5 days	10 km	>2	-	-	n.a.*
			>1	-	-	
2010	2 days	5 km	>3	1 day	5 km	<0.001
			>2	3 days	5 km	
			>1	13 days	90 km	
2011	2 days	5 km	>3	3 days	15 km	0.050
			>2	5 days	20 km	
			>1	13 days	35 km	

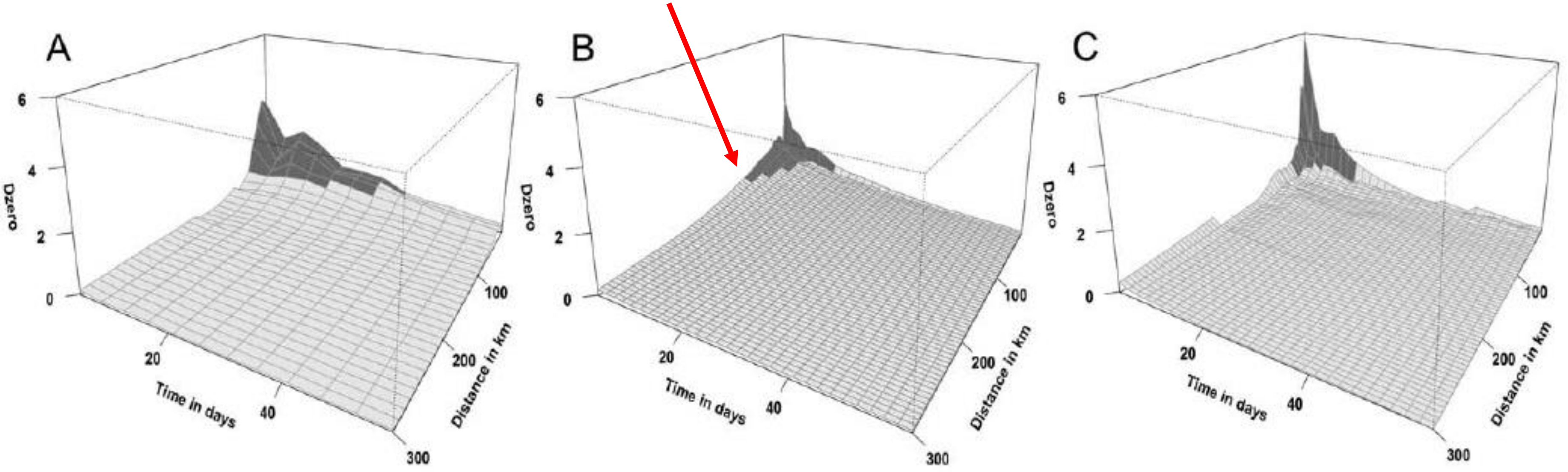
*n.a.: not applicable: $D_o(s,t)$ values were below unity.

doi:10.1371/journal.pntd.0001808.t003

long-distance transmission

Plot of excess risk attributed to space-time interactions

long-distance transmission



Findings and discussions

The study detected the presence of an additional spatiotemporal process, with RVF potentially spreading to distances up to 40 to 90 km, within about 2 weeks.

This appearance of long-distance spread could be explained by the existence of several RVF virus emergences. This suggests that RVF spread over distances larger than the assumed range of active vector dispersal could be explained by the movement of domestic or wild viraemic and therefore infectious animals.

Other **mechanisms of long-distance** spread could also be incriminated, such as wind-borne vector dispersal.

International Regional Science Review
33(2) 134-156

DOI: 10.1177/0160017609354760
http://irsr.sagepub.com
SAGE

A Space–Time Approach to Diffusion of Health Service Provision Information

Irene Casas,[1] Eric Delmelle,[2] and Alejandro Varela[3]

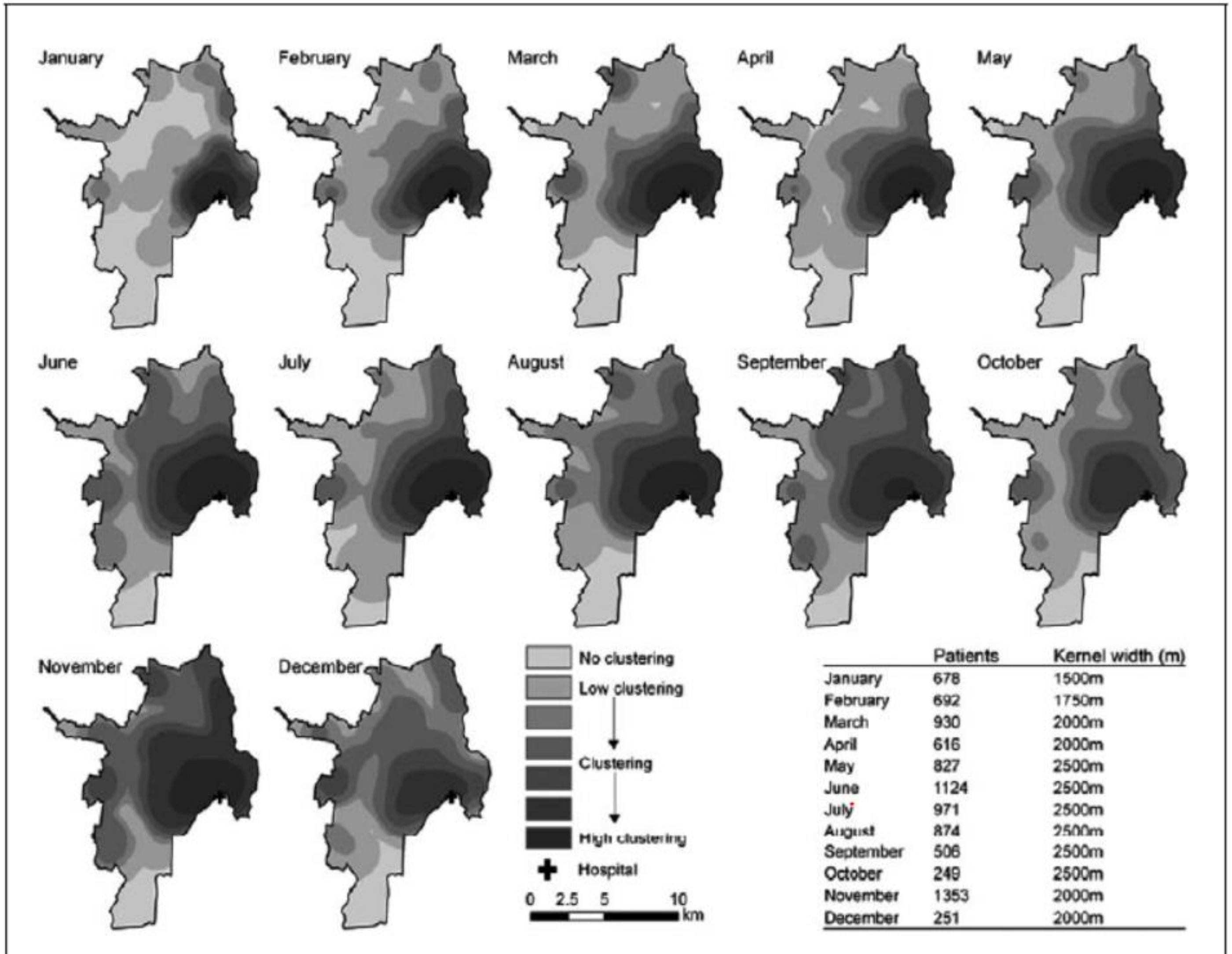


Figure 8. Monthly variation in the patients' density for the year 2004.

Space-time K-function

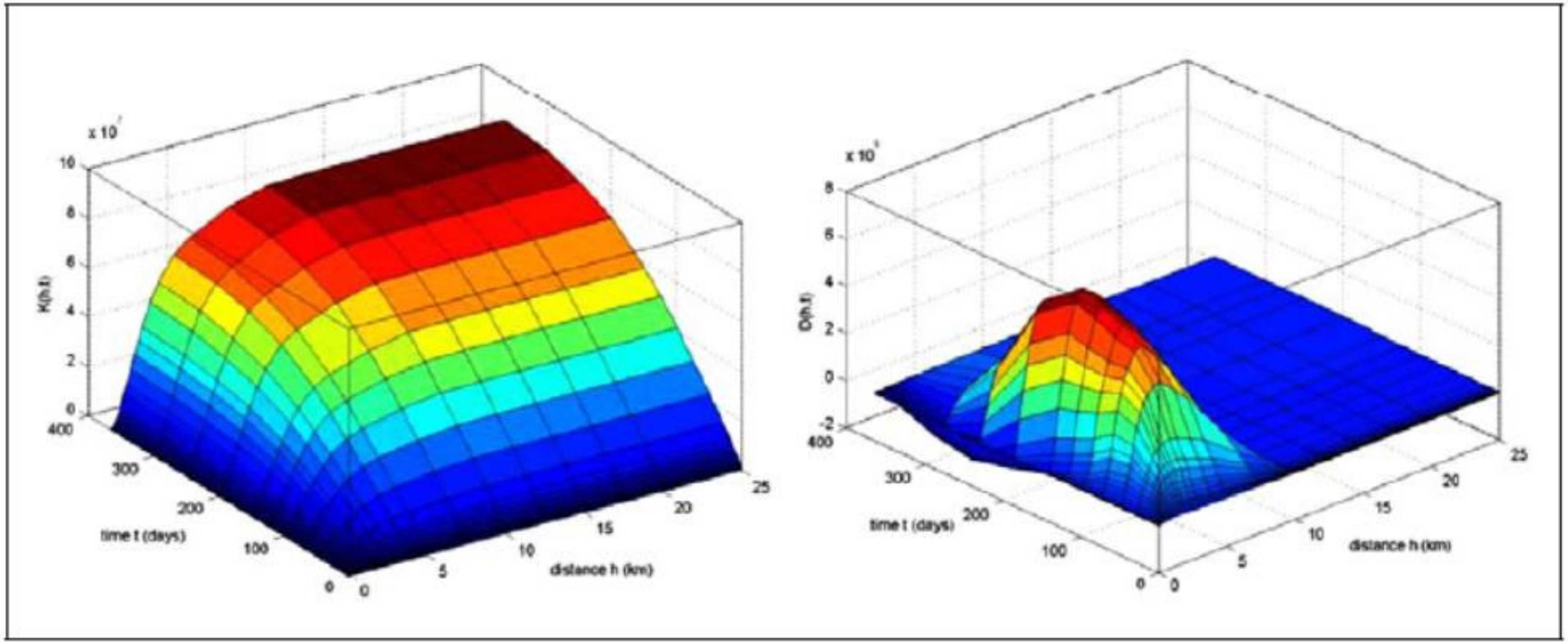


Figure 9. The space–time K-function $\hat{K}(h,t)$ on the left and the test for space–time interaction $\hat{D}(h,t)$ on the right.

Findings

$$\hat{D}(h,t) = \hat{K}(h,t) - \hat{K}(h) * \hat{K}(t). \quad (9)$$

Figure 9 to the right shows the change in the value of $\hat{D}(h,t)$ with time and distance variation. A clear space–time interaction is noticed at short distances between 1 and 10 km, with a peak at 3 km. This comes very close to the results obtained with the monthly variation of the space K-function, where clusters were noticed in the 1–2 km range. Along the time axis, it can be observed that the surface reaches a maximum at approximately 180 days but remains very significant between 50 and 200 days. There is a clustering of patients who live within a small distance from one another (<5 km) and visit the hospital within 6–7 months from one another. This coincides with the cumulative distribution reported in figure 4, where it was observed that most patients were originating from about 4.5 km of the hospital. These results are an evidence of a diffusion process taking place, where a focal point in the vicinity of the hospital originates, and as time goes by, it begins to radiate outward spreading to neighborhoods with similar characteristics to the intended service area.

R Functions for Space-time K-function

Usage `library (splancs)`

```
stkhat(pts, times, poly, tlimits, s, tm)
```

Arguments

`pts`	A set of points as defined in Splancs
`times`	A vector of times, the same length as the number of points in `pts`
`poly`	A polygon enclosing the points
`tlimits`	A vector of length 2 specifying the upper and lower temporal domain.
`s`	A vector of spatial distances for the analysis.
`tm`	A vector of times for the analysis

Value

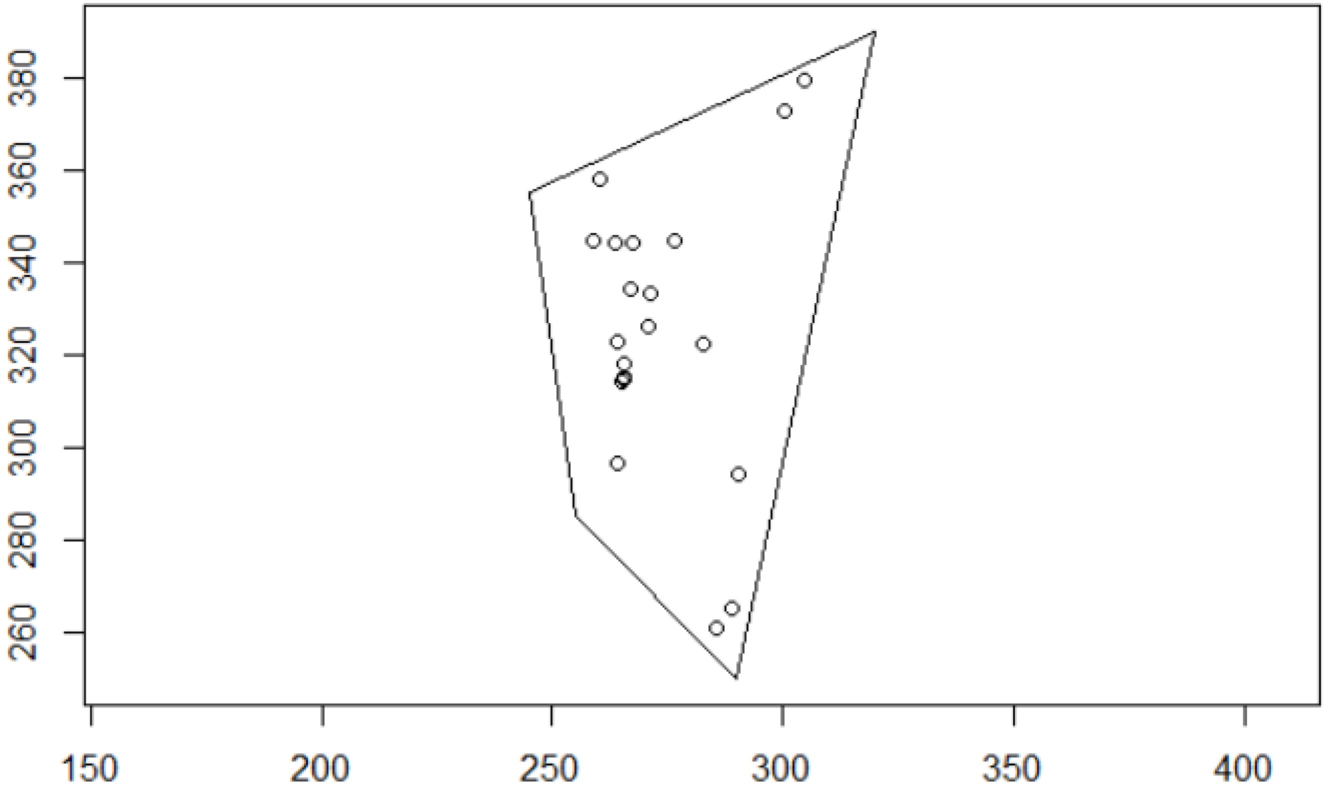
A list with the following components is returned:

`s, t`	The spatial and temporal scales
`ks`	The spatial K-function
`kt`	The temporal K-function
`kst`	The space-time K-function

R code: Space-time Data

```
library (splancs)
source("ST_functions.R")


ptdata <- read.table("pts_data/Patients.csv", header=TRUE,
sep=",")
Pts_Loc <- as.points(ptdata[,2], ptdata[,3])
Pts_time <- ptdata[,4]
ptbnd <- read.table("pts_data/Paitents_BND.csv", header=TRUE,
sep=",")
Pts_BND <- as.points(ptbnd[,2], ptbnd[,3])
polymap(Pts_BND)
pointmap(Pts_Loc, add=T)
```

Mapping Space-time Data



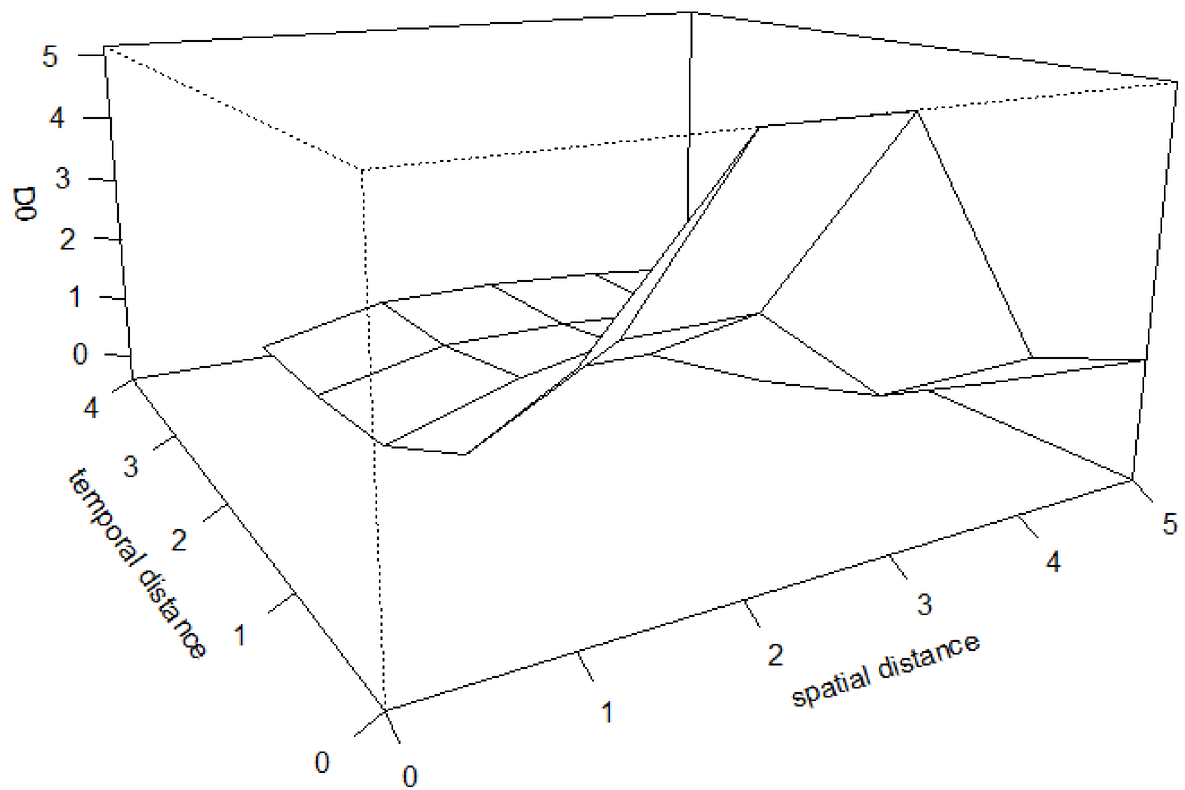
R code: Space-time K function

```
## Plotting D0(s,t)

kap1<- stkhat(Pts_Loc, Pts_time, Pts_BND, c(1955,
1980),seq(1,5),seq(0,4))

g1<- matrix(kap1$ks)

g2<- matrix(kap1$kt)

g1g2<- g1 %*% t(g2)

turD<- kap1$kst - g1g2

persp(kap1$s, kap1$t, turD, theta=-30, phi = 15, expand = 0.5,
xlim=c(0,5), ylim=c(0,4), xlab="spatial distance", ylab="temporal
distance", zlab="D", ticktype ="detailed" )

turD0<- kap1$kst/g1g2-1.0

persp(kap1$s, kap1$t, turD0, theta=-30,phi = 15, expand = 0.5,
xlim=c(0,5), ylim=c(0,4), xlab="spatial distance", ylab="temporal
distance", zlab="D0", ticktype ="detailed" )
```

$$\hat{D}(d,t) = \hat{K}(d,t) - \hat{K}_D(d)\hat{K}_T(t)$$

$$\hat{D}_0(d,t) = \hat{D}(d,t) \;/\; \{\hat{K}_D(d)\hat{K}_T(t)\}$$

Results: Distribution of Do



Standardized residuals

$$\hat{R}(d,t) = \hat{D}(d,t) \,/\, \sqrt{\hat{V}(d,t)}.$$

Standard error

Standard error for space-time clustering

Description

Computes the standard error for space-time clustering.

Usage

```
stsecal(pts, times, poly, tlim, s, tm)
```

Arguments

`pts`	A set of points, as defined in Splancs.
`times`	A vector of times, the same length as the number of points in `pts`
`poly`	A polygon enclosing the points
`tlim`	A vector of length 2 specifying the upper and lower temporal domain.
`s`	A vector of spatial distances for the analysis
`tm`	A vector of times for the analysis

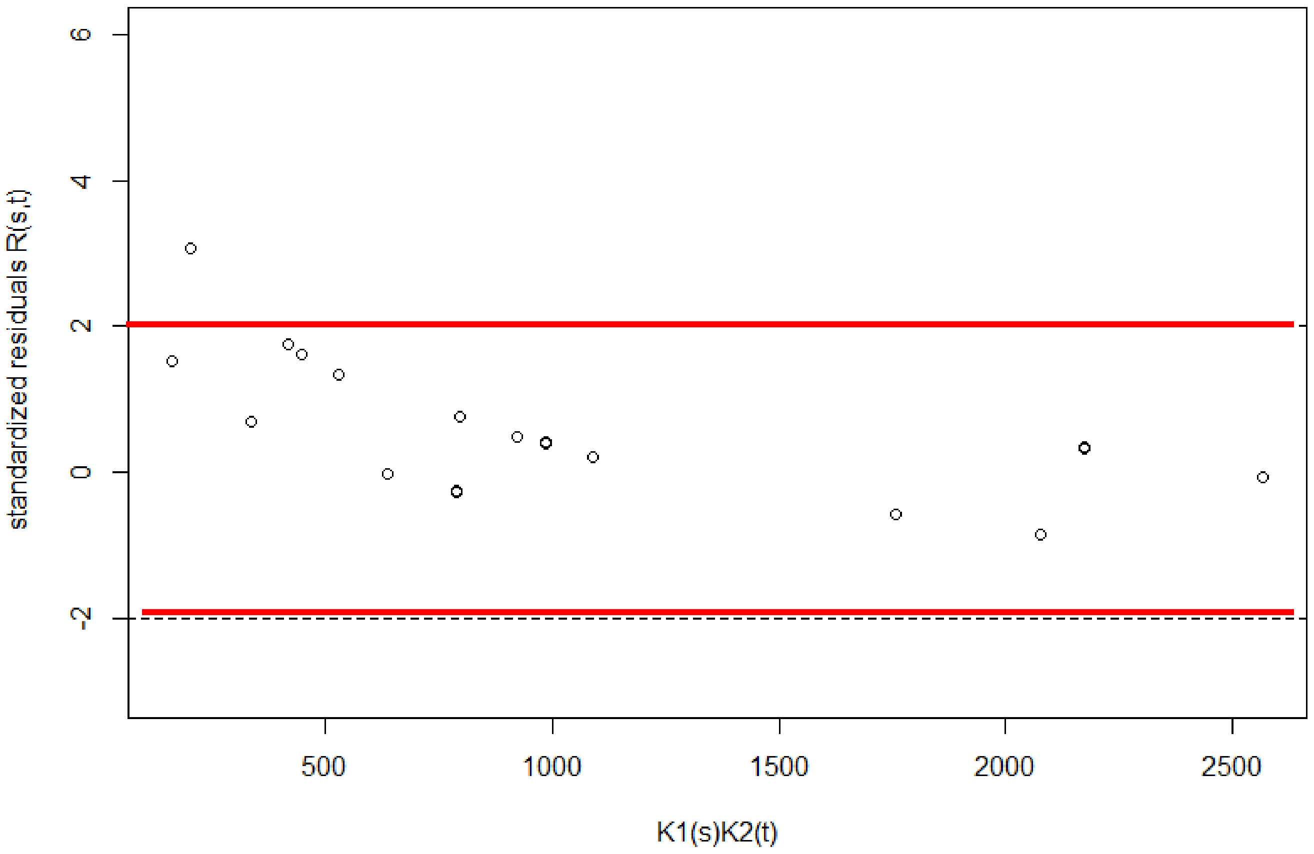
R code: standardized residuals R(s,t)

Standard error

$$\hat{R}(d,t) = \hat{D}(d,t) / \sqrt{\hat{V}(d,t)}.$$

```
# plotting standardized residuals R(s,t)

se<- stsecal(Pts_Loc, Pts_time, Pts_BND,c(1955,
1980),seq(1,5),seq(0,4))

Res<- turD / se

plot( glg2, Res ,ylim=c(-3, 6), xlab="K1(s)K2(t)",
ylab=" standardized residuals R(s,t)")

abline(h=c(-2,2), lty=2)
```

Results: Standardized residuals R(s,t)



作業

- 1. R實作：建立自訂函數
  - 函數1：建立距離與時間矩陣 (p. 7)
  - 函數2：建立時空關連圖 (p. 30)
  - 函數3：建立 Jacquez‘s k-NN曲線，含隨機區間 (p. 42)
- 2. R實作：時空分析(時空K函數)
  - 利用台北市某疾病的疫情資料 (Pts_tpe.rar)，進行時空風險評估