

6. Space-Time Point Processes

Point events (such as crimes or disease cases) occur in *time* as well as space. If both time and location data are available for these events, then one can in principle model this data as the realization of a *space-time point process*. As a prime example, recall that the **Burkitt's Lymphoma data** (examined in Assignment 1 of this class) contains both onset times and locations for 188 cases during the period 1961-1975. Moreover, the original study of this data by Williams et al. (1978)¹, (here referred to as [W]) focused precisely on the question of identifying significant *space-time clustering* of these cases. Hence it is of interest to consider this data in more detail.

The cases occurring in each five-year period of the study are displayed in Figure 6.1 below (with green shading reflecting relative population density in West Nile), and correspond roughly to Figure 5 in [W].² Here it does appear that cases in subsequent periods tend to be clustered near cases in previous periods. But the inclusion of population density in Figure 6.1 was done specifically to show that such casual observations can be deceptive. Much of the new clustering is seen to occur in more densely populated areas where one would expect new cases to be more likely based on chance alone.

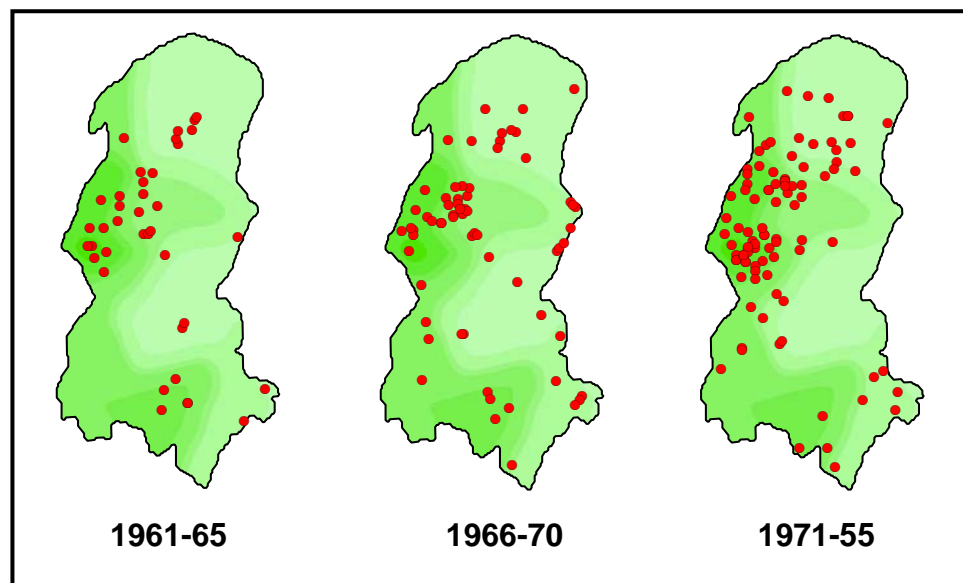


Figure 6.1 Lymphoma Cases in each Five-Year Period

The simple regression procedure used in Assignment 1 related times of cases to those of their nearest-neighbors. But since population density is ignored in this approach, the “clustering” result obtained by this procedure is questionable at best. Hence, one

¹ This is *Paper 1* in the Reference Materials on the class web page.

² These cases differ slightly from those in Figure 5 of [W]. The present approximation is based on the counting convention stated in [BG, p.81] that time is “measured in days elapsed since January 1st, 1960”. This rule does not quite agree with the actual dates in the Appendix of [W], but the difference is very slight.

objective of the present section is to develop an alternative “random labeling” test that is more appropriate. But before doing so, we shall consider the general question of space-time clustering more closely.

6.1 Space-Time Clustering

Event sequences exhibit *space-time clustering* if events that are close in space tend to be closer in time than would be expected by chance alone. The most classic examples of space-time clustering are *spatial diffusion processes* in which point events are propagated from locations to neighbors through some form of local interactions. Obvious examples include the spread of forest fires (where new trees are ignited by the heat from trees burning nearby), or the spread of contagious diseases (where individuals in direct contact with infected individuals also become infected). Here it is worth noting that cancers such as Burkitt’s Lymphoma are *not* directly contagious. However, as observed in [W,p.116], malaria infections may be a contributing factor leading to Burkitt’s Lymphoma, and the spread of malaria itself involves a diffusion process in which mosquitoes transmit this disease from existing victims to new victims.

But even with genuine diffusion processes one must be careful in analyzing space-time clustering. Consider the onset of a new flu epidemic introduced into a region, R , by a single carrier, c , and suppose that the cases occurring during the first few days are those shown in Figure 6.2 below.

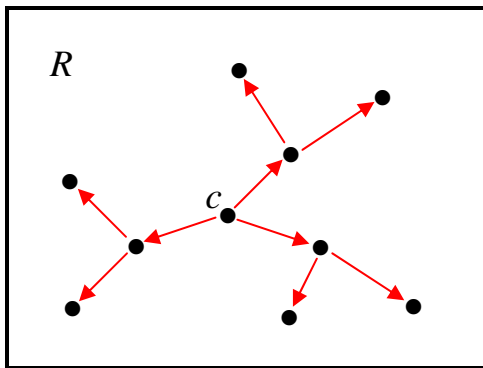


Figure 6.2. Early Epidemic

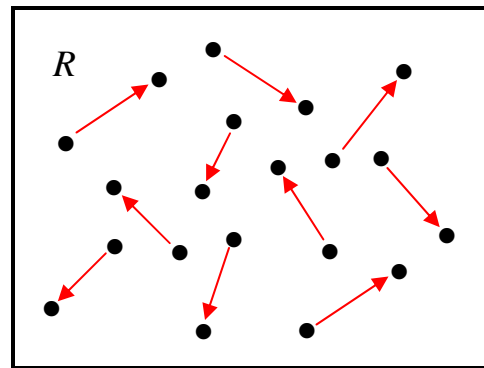


Figure 6.3. Late Epidemic

Here there is a clear diffusion effect in which the initial cases involve contacts with c , and are in turn propagated to others by secondary contacts. But notice that even though the initial three cases shown are all close to c , this process spreads out quickly. So while the six “second round” cases shown in the figure may all occur at roughly the same time, they are already quite dispersed in space. This example shows that cases occurring close in time need *not* occur close in space. However, this figure also suggests that cases occurring close in space may indeed have a tendency to occur close in time.³ So there

³ Here we assume that most contacts involve individuals living in close spatial proximity – which may not be the case. For example, some individuals have significant contact with co-workers at distant job sites.

appears to be some degree of asymmetry between space and time in such processes. We shall return to this issue below.

While the early stages of this epidemic show clear propagation effects, this is not true at later stages. After the first few weeks, such an epidemic may well have spread throughout the region, so that the pattern of new cases occurring on each day may be very dispersed, as shown in Figure 6.3. More importantly, this pattern will most likely be quite similar from day to day. At this stage, the diffusion process is said to have reached a *steady state* (or *stationary state*). In such a steady state it is clearly much harder to detect any space-time clustering whatsoever. Diffusion is still at work, but the event pattern is no longer changing in detectable ways.⁴ However, it may still be possible to detect such space-time effects indirectly. For example, if one were to examine the distribution of cases on day t , and to identify the new cases on day $t+1$, then it might still be possible to test whether these new cases are “significantly closer” to the population of cases on day t than would be expected by chance alone. We shall not pursue such questions here. Rather the intent of this illustration is to show that space-time clustering can be subtle in even the clearest examples of spatial diffusion.

6.2 Space-Time K-Functions

With this preliminary discussion we turn now to the measurement of space-time clustering. Here we follow approach of [BG, Section 4.3] by constructing a space-time version of K-functions. Consider a *space-time pattern* of events, $\{e_i = (s_i, t_i) : i = 1, \dots, n\}$, in region, R , where s_i again denotes the *location* of event e_i in R , and t_i denotes the *time* at which event e_i occurs. If for a given event e_i we are interested in the numbers of events that are “close” to e_i in both space and time, then for each *spatial distance*, h , and *time increment*, Δ , it is natural to define the corresponding *space-time neighborhood* of event, $e_i = (s_i, t_i)$, by the Cartesian product:

$$(6.2.1) \quad C_{(h,\Delta)}(e_i) = \{(s, t) : \|s_i - s\| \leq h, |t_i - t| \leq \Delta\} \\ = \{s : \|s_i - s\| \leq h\} \times \{t : |t_i - t| \leq \Delta\}$$

Hence the circular neighborhoods, $C_h(s_i)$, in two dimensions are now replaced by *cylindrical neighborhoods*, $C_{(h,\Delta)}(e_i)$, in three dimensions, as shown in Figure 6.4 below.

⁴ A more extreme example is provided by change in temperature distribution within a room after someone has lit a match. While the match is burning, there is very sharp peak in the temperature distribution that spreads out from this point source of heat. After the match has gone out, this heat is not lost. Rather it continues to diffuse throughout the room until a new steady state is reached in which the temperature is everywhere slightly higher than before.

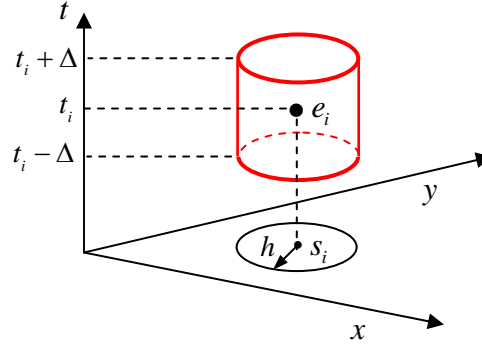


Figure 6.4 Space-Time Neighborhoods

As in two dimensions, one can define a relevant *space-time region* as the Cartesian product, $R \times T$ of the given spatial region, R , and a relevant time interval, T . For a given pattern of events, $\{e_i = (s_i, t_i) : i = 1, \dots, n\}$, the default time interval, T , for purposes of analysis is usually taken to be the *smallest* time interval containing all event times, i.e.,

$$(6.2.2) \quad T = t_{\max} - t_{\min} = \max\{t_i : i = 1, \dots, n\} - \min\{t_i : i = 1, \dots, n\}$$

as illustrated in Figure 6.5 below:⁵

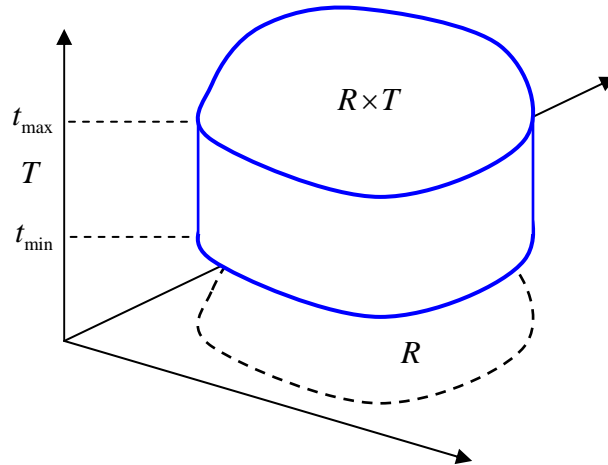


Figure 6.5. Space-Time Region for Analysis

In this context, the desired space-time extension of K-functions is completely straightforward. First, if for any two space-time events, $e_i = (s_i, t_i)$ and $e_j = (s_j, t_j)$ we

⁵ At this point it should be noted that, as with two dimensions, the cylindrical neighborhoods in (6.2.1) are subject to “edge effects” in $R \times T$, so that in general, one must replace $C_{(h,\Delta)}(e_i)$ by $C_{(h,\Delta)}(e_i) \cap (R \times T)$.

now let $t_{ij} = |t_i - t_j|$ (and again let $d_{ij} = \|s_i - s_j\|$) then as an extension of (4.3.2), we now have the following space-time *indicator functions*:

$$(6.2.3) \quad I_{(h,\Delta)}(d_{ij}, t_{ij}) = \begin{cases} 1, & (d_{ij} \leq h) \text{ and } (t_{ij} \leq \Delta) \\ 0, & \text{otherwise} \end{cases}$$

If for a given space-time point process we let λ_{st} denote the *space-time (st) intensity* of events, i.e., the expected number of events per unit of space-time volume, then the *desired space-time K-function* is again defined for each $h \geq 0$ and $\Delta \geq 0$ to be the expected number of additional events within space-time distance (h, Δ) of a randomly selected event, e_i , i.e.,

$$(6.2.4) \quad K(h, \Delta) = \frac{1}{\lambda_{st}} \sum_{j \neq i} E[I_{(h,\Delta)}(d_{ij}, t_{ij})]$$

So as in (4.3.4), for any given pattern size, n , the *pooled* form of this function

$$(6.2.5) \quad K(h, \Delta) = \frac{1}{n \cdot \lambda_{st}} \sum_{i=1}^n \sum_{j \neq i} E[I_{(h,\Delta)}(d_{ij}, t_{ij})]$$

implies that the natural estimator of $K(h, \Delta)$ is given by *sample space-time K-function*:

$$(6.2.6) \quad \hat{K}(h, \Delta) = \frac{1}{n \cdot \hat{\lambda}_{st}} \sum_{i=1}^n \sum_{j \neq i} I_{(h,\Delta)}(d_{ij}, t_{ij})$$

Here the *sample estimate*, $\hat{\lambda}_{st}$, of the space-time intensity is given by

$$(6.2.7) \quad \hat{\lambda}_{st} = \frac{n}{a(R) \cdot (t_{\max} - t_{\min})}$$

where the denominator is now seen to be the *volume* of the space-time region, $R \times T$, in Figure 6.5 above.

6.3 Temporal Indistinguishability Hypothesis

To test for the presence of space-time clustering, one requires the specification an appropriate null hypothesis representing the complete absence of space-time clustering. Here the natural null hypothesis to adopt is simply that there is no relation between the locations and timing of events. Hence in a manner completely paralleling the treatment of marked point processes in (5.6.1) it is convenient to separate space and time, and write the joint probability of space-time events as,

$$(6.3.1) \quad \Pr[(s_i, t_i) : i = 1, \dots, n] = \Pr[(s_1, \dots, s_n), (t_1, \dots, t_n)]$$

$$= \Pr[(t_1, \dots, t_n) | (s_1, \dots, s_n)] \cdot \Pr(s_1, \dots, s_n)$$

where $\Pr(s_1, \dots, s_n)$ again denotes the *marginal* distribution of event locations, and where $\Pr[(t_1, \dots, t_n) | (s_1, \dots, s_n)]$ denotes the *conditional* distribution of event times given their locations.⁶ In this context, If the marginal distribution of event times is denoted by $\Pr(t_1, \dots, t_n)$, then a parallel to (5.6.2), the relevant hypothesis of *space-time independence* for our present purposes can be stated as follows:

$$(6.3.2) \quad \Pr[(t_1, \dots, t_n) | (s_1, \dots, s_n)] \equiv \Pr(t_1, \dots, t_n)$$

Here it should be noted (as in footnote 5 of Section 5) that from a formal viewpoint, this independence condition could equally well be stated by switching the roles of locations, (s_1, \dots, s_n) , and times, (t_1, \dots, t_n) , in (6.3.2). But as noted in Section 6.1 above, there is a subtle asymmetry between space and time that needs to be considered here. In particular, recall that event sequences are said to exhibit *space-time clustering* if events that are close in space tend to be closer in time than would be expected by chance alone. Hence it is somewhat more natural to *condition* on the spatial locations of events and look for time similarities among those events that are close in space.

Note also that as with marked point processes, the indexing of events, e_i , is completely arbitrary. Here it might be argued that the ordering of indices i should reflect the ordering of event occurrences. But this is precisely why event times have been listed as distinct attributes of space-time events. Hence in the present formulation, it is again most appropriate to treat space-time pairs, (s_i, t_i) and (s_j, t_j) as exchangeable events. In a manner paralleling condition (5.6.3), this implies that for all permutations (π_1, \dots, π_n) of the subscripts $(1, \dots, n)$ the marginal distribution of event times should satisfy the *exchangeability* condition:

$$(6.3.3) \quad \Pr(t_{\pi_1}, \dots, t_{\pi_n}) = \Pr(t_1, \dots, t_n)$$

These two conditions together constitute our null hypothesis that spatial events are completely indistinguishable in terms of their occurrence times. Hence we now designate the combination of conditions, (5.6.2) and (5.6.3) as the *temporal indistinguishability hypothesis*.

6.4 Random Labeling Test

We now extend the argument in Section 5.6.2 to obtain an exact sampling distribution for testing this temporal indistinguishability hypothesis. To do so, observe first that the

⁶ Again for simplicity we take the number of space-time events, n , to be fixed. Alternatively, the distributions in (6.3.1) can all be conditioned on n .

argument in (5.6.4) now shows that conditional distribution in (6.3.2) inherits exchangeability from (6.3.3), i.e., that for all permutations (π_1, \dots, π_n) of $(1, \dots, n)$,

$$(6.4.1) \quad \Pr[(t_{\pi_1}, \dots, t_{\pi_n}) | (s_1, \dots, s_n)] = \Pr(t_{\pi_1}, \dots, t_{\pi_n}) \\ = \Pr(t_1, \dots, t_n) = \Pr[(t_1, \dots, t_n) | (s_1, \dots, s_n)]$$

Hence the only question is how to condition these permutations to obtain a well-defined probability distribution. Recall that the appropriate conditional information shared by all permutations of population labels, (m_1, \dots, m_n) , was precisely the number of instances of each label, “1” and “2”, i.e., the population sizes, n_1 and n_2 . Here the set of label frequencies, $\{n_1, n_2\}$, is now replaced by the set of *time frequencies*, $\{n_t : t \in T\}$, where n_t is the number of times that t occurs in the given set of event times, (t_1, \dots, t_n) , i.e.,⁷

$$(6.4.2) \quad n_t = |\{i : t = t_i, i = 1, \dots, n\}|$$

It is precisely this frequency distribution which is shared by all permutations, $(t_{\pi_1}, \dots, t_{\pi_n})$ in (6.4.1). Indeed, it follows [as a parallel to (5.6.5)] that for every list of times (t'_1, \dots, t'_n) consistent with this distribution, there is some permutation $(t_{\pi_1}, \dots, t_{\pi_n})$ of (t_1, \dots, t_n) with:

$$(6.4.3) \quad (t'_1, \dots, t'_n) = (t_{\pi_1}, \dots, t_{\pi_n})$$

Hence if the conditional distribution of such times given both (s_1, \dots, s_n) and $\{n_t : t \in T\}$ is denoted by $\Pr[\cdot | (s_1, \dots, s_n), \{n_t : t \in T\}]$, then the same arguments in (5.6.6) through (5.6.8) now yield the following exact conditional distribution for all permutations (π_1, \dots, π_n) of these occurrence times under the *temporal indistinguishability hypothesis*:

$$(6.4.4) \quad \Pr[(t_{\pi_1}, \dots, t_{\pi_n}) | (s_1, \dots, s_n), \{n_t : t \in T\}] \equiv \frac{1}{n!}$$

As in Section 5.6.2, this sampling distribution again leads directly to a *random-labeling test* of this hypothesis. For completeness, we list the steps of this test, which closely parallels the random-labeling test of Section 5.6.2:

- (i) Given *observed* locations, (s_1, \dots, s_n) , and occurrence times, (t_1, \dots, t_n) , simulate N random permutations $[\pi_1(\tau), \dots, \pi_n(\tau)]$, $\tau = 1, \dots, N$ of $(1, \dots, n)$, and form the permuted labels $(t_{\pi_1(\tau)}, \dots, t_{\pi_n(\tau)})$, $\tau = 1, \dots, N$ [which is now equivalent to taking a sample of size N from the distribution in (6.4.4)].

⁷ Note that in most cases these frequencies will either be zero or one. But the present general formulation allows for the possibility of simultaneous events, as for example Lymphoma cases reported on the same day (or even instantaneous events, such as multiple casualties in the same auto accident).

(ii) If $\hat{K}^\tau(h, \Delta)$ denotes the *sample space-time K-function* resulting from joint realization, $[(s_1, \dots, s_n), (t_{\pi_1(\tau)}, \dots, t_{\pi_n(\tau)})]$, then choose relevant sets of distance radii, $\{h_w : w = 1, \dots, W_R\}$, for R , and time intervals, $\{\Delta_v : v = 1, \dots, V_T\}$ for T , and calculate the sample space-time K-function values, $\{\hat{K}^\tau(h_w, \Delta_v) : w = 1, \dots, W_R, v = 1, \dots, V_T\}$ for each $\tau = 1, \dots, N$.

(iii) Finally, if the *observed sample space-time K-function*, $\hat{K}^0(h, \Delta)$, is constructed from the observed event sequence, $[(s_1, \dots, s_n), (t_1, \dots, t_n)]$, then under the *temporal indistinguishability hypothesis* each observed value, $\hat{K}^0(h_w, \Delta_v)$, should be a “typical” sample from the list of values $[\hat{K}^\tau(h_w, \Delta_v) : \tau = 0, 1, \dots, N]$. Hence if M_+^0 denotes the number of simulated random relabelings, $\tau = 1, \dots, N$, with $\hat{K}^\tau(h_w, \Delta_v) \geq \hat{K}^0(h_w, \Delta_v)$, then the probability of obtaining a value *as large as* $\hat{K}^0(h_w, \Delta_v)$ under this hypothesis is estimated by the *space-time clustering p-value*:

$$(6.4.5) \quad \hat{P}_{st\text{-clustered}}(h_w, \Delta_v) = \frac{M_+^0 + 1}{N + 1}$$

(iv) Similarly, if M_-^0 denotes the number of simulated random relabelings, $\tau = 1, \dots, N$, with $\hat{K}^\tau(h_w, \Delta_v) \leq \hat{K}^0(h_w, \Delta_v)$, then the estimated probability of obtaining a value *as small as* $\hat{K}^0(h_w, \Delta_v)$ under this hypothesis is again given by the *space-time dispersion p-value*:

$$(6.4.6) \quad \hat{P}_{st\text{-dispersed}}(h_w, \Delta_v) = \frac{M_-^0 + 1}{N + 1}$$

Our primary interest here is of course in space-time clustering for relatively small values of h and Δ . But it is clear that a range of other questions could in principle be addressed within the more general framework outlined above.

6.5 Application to the Lymphoma Example

This testing procedure is implemented in the MATLAB program, **space_time_plot.m**, and can be applied to the Lymphoma example above as follows. In the MATLAB workspace, **lymphoma.mat**, the (188 x 3) matrix, **LT**, contains space-time data for the **n=188** lymphoma cases, with rows (x_i, y_i, t_i) denoting the location, (x_i, y_i) , and onset time, t_i , of each case i . In this program, the maximum distance again set to $h_{\max}/2$ as in (4.5.1) above, and similarly, the maximum temporal interval is set to half the maximum

time interval, $\Delta_{\max}/2$, where $\Delta_{\max} = t_{\max} - t_{\min}$ in Figure 6.5 above. Given these maximum values, the user has the option of choosing subdivisions of $h_{\max}/2$ into s equal increments, $h_i = (i/s)(h_{\max}/2)$, $i=1, \dots, s$, and subdivisions of $\Delta_{\max}/2$ into t equal increments, $\Delta_j = (j/t)(\Delta_{\max}/2)$, $j=1, \dots, t$. So for example the following command uses 999 random relabelings of times to test for space-time clustering of the Lymphoma data, **LT**, at each point on a grid of space-time neighborhoods (h_i, Δ_j) with $s = t = 20$:

```
>> results = space_time_plot(LT,999,20,20);
```

The results of these $s \times t = 400$ tests is plotted on a grid and then interpolated in MATLAB to obtain a p-value contour map such as the one shown in Figure 6.6 below:

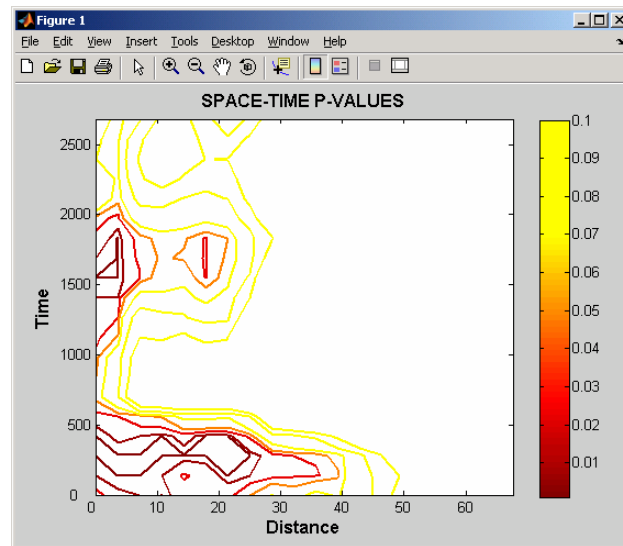


Figure 6.6. P-value Map for Lymphoma Data

Note first that each location in this region corresponds to the *size* of a space-time neighborhood. Hence those areas with darker contours indicate space-time scales at which there are significantly *more* cases in neighborhoods of this size (about randomly selected cases) than would be expected under the temporal indistinguishability hypothesis. In particular, the dark contours in the lower left corner show that there is very significant concentration in small space-time neighborhoods, and hence significant *space-time clustering*. This not only confirms the findings of the simple regression analysis done in Assignment 1, but also conveys a great deal more information. In fact the darkest contours show significance at the .001 level (which is the maximum significance achievable with 999 simulations).⁸

⁸ Note also that these p-values can be retrieved in numerical form from the output structure, **results**, in the command above.

Before discussing these results further, it is of interest to observe that while the direct plot in MATLAB above is useful for obtaining visual results quickly, these p-values can also be exported to ARCMAP and displayed in sharper and more vivid formats. For example, the above results were exported to ARCMAP and smoothed by ordinary kriging to obtain the sharper representation shown in Figure 6.7 below:

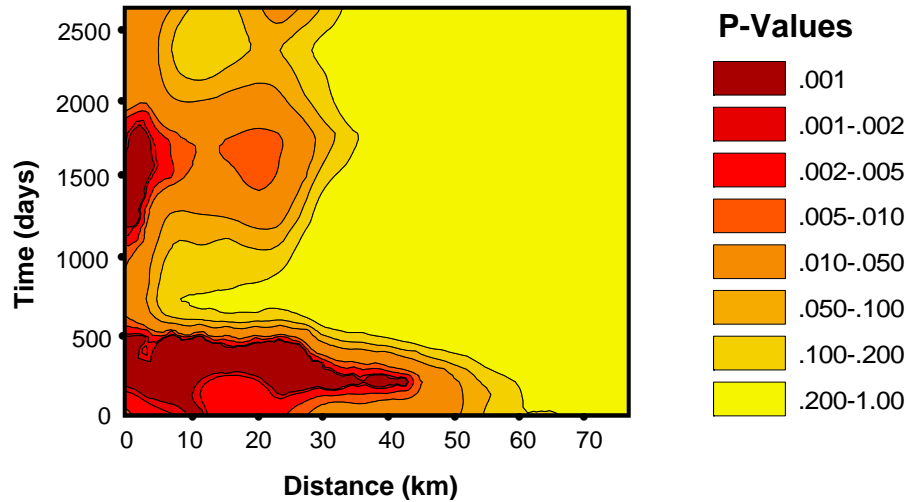


Figure 6.7 Smoothed P-Value Map in ARCMAP

Using this sharper image, notice first that the horizontal band of significance at the bottom of the figure indicates significant clustering of cases within 500 days of each other (≈ 1.4 years) over a wide range of distances. This suggests the presence of short periods (about 1.4 years) with unusually high numbers of cases over a wide region, i.e., local peaks in the frequency of cases over time. This can be confirmed by Figure 6.8 below, where a number of local peaks are seen, such as in years 7, 11, 13 and 15 (with year 1 corresponding to 1961)

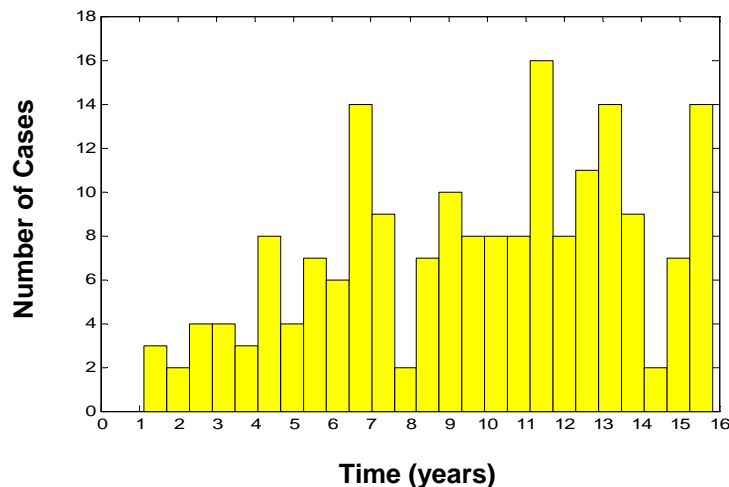


Figure 6.8 Time Frequency of Lymphoma Cases

Next observe that there is a secondary mode of significance at about 1500 days (≈ 4 years) on the left edge of Figure 6.7. This indicates that many cases occurred close to one another over a time lag of about 4 years. Note in particular that the peak years 7, 11, and 15 are spaced at 4 years. This suggests that such peaks may represent new outbreaks of Lymphoma cases in the same areas at intervals of about 4 years. Hence the p-value plots in Figures 6.6 and 6.7 above do indeed yield more information than simple space-time clustering of events.