

Old Dominion University

**ODU Digital Commons**

---

Mathematics & Statistics Faculty Publications

Mathematics & Statistics

---

3-2018

## **Density Estimation of Spatio-Temporal Point Patterns Using Moran's Statistics**

Jennifer L. Lorio

Norou Diawara

Lance A. Waller

Follow this and additional works at: [https://digitalcommons.odu.edu/mathstat\\_fac\\_pubs](https://digitalcommons.odu.edu/mathstat_fac_pubs)



Part of the [Applied Statistics Commons](#), and the [Probability Commons](#)

---

# Density Estimation of Spatio-temporal Point Patterns Using Moran's Statistic

Jennifer L. Lorio<sup>1</sup>, Norou Diawara<sup>1</sup> & Lance A. Waller<sup>2</sup>

<sup>1</sup> Department of Mathematics & Statistics, Old Dominion University, Norfolk, VA USA

<sup>2</sup> Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA USA

Correspondence: Norou Diawara, Department of Mathematics & Statistics, Old Dominion University, 2300 Engineering & Computational Sciences Building, Norfolk, VA 23529. E-mail: [ndiawara@odu.edu](mailto:ndiawara@odu.edu)

Received: November 12, 2017 Accepted: December 5, 2017 Online Published: February 22, 2018

doi:10.5539/ijsp.v7n2p80

URL: <https://doi.org/10.5539/ijsp.v7n2p80>

## Abstract

Moran's Index is a statistic that measures spatial autocorrelation, quantifying the degree of dispersion (or spread) of objects in space. When investigating data in an area, a single Moran statistic may not give a sufficient summary of the autocorrelation spread. However, by partitioning the area and taking the Moran statistic of each subarea, we discover patterns of the local neighbors not otherwise apparent. In this paper, we consider the model of the spread of an infectious disease, incorporate time factor, and simulate a multilevel Poisson process where the dependence among the levels is captured by the rate of increase of the disease spread over time, steered by a common factor in the scale. The main consequence of our results is that our Moran statistic is calculated from an explicit algorithm in a Monte Carlo simulation setting. Results are compared to Geary's statistic and estimates of parameters under Poisson process are given.

**Keywords:** Moran's index, spatio-temporal models, simulated processes

## 1. Introduction

Moran's Index is a measure of spatial autocorrelation within a domain area; it has and continues to be applied in many fields. As described by Moran (1950), when given a set of spatial variates such as  $(x, y)$  (defined on a two-dimensional discrete area), we may want to investigate whether there is any evidence that spatial autocorrelation is present overall or in neighboring clusters based on selected features. For example, in epidemiology, we may be interested in how a disease spreads based on the number of people infected in a given area. By adding a time component to the model, Moran's indices can lead to better understanding of the spread.

Indeed, modeling disease spatio-temporal spread using ordinary least squares or weighted least squares will lead to much bias in the estimates and violation of model assumptions (Pace & LeSage, 2009). A simple global Moran's value cannot explain differences, area spread, or the relationship of some natural ubiquitous phenomena. This has led many authors such as Anselin (1995) to partition the global Moran value into a sum of local indices of spatial association (LISA), because data show local trends that are not shared globally. However, the impact of time was not incorporated. Spatio-temporal autocorrelation was first introduced by Cliff and Ord (1975) and the concept has been explored by many others over the years (see for example, Martin & Oeppen, 1975; Wang & He, 2007; and Lee & Li, 2017). We propose to evaluate spatio-temporal association, extending Moran's index spatial autocorrelation models over iterative time in dynamic subareas under the Poisson process. A full description of the model is presented as a way to measure spatio-temporal non-stationary association.

In this paper, we design a multilevel Poisson process that is time dependent, and the dependence among the levels is captured by a metric of closeness, the rates of increase of the disease spread over time, steered by a common factor in the scale at each time step. Meddens and Hicke (2014) describe spatial and temporal patterns of tree mortality from beetle ecology and dynamics in parts of Colorado and Wyoming. The spatial point process experiment in Vaillant et al. (2011) is revisited the same number of time intervals. In our case, we partition the domain area into subareas and propose incremental time rates of increase under critical scales from each subarea. The scale parameter effectively creates a region of finer detail around disease occurrence. Using Monte Carlo techniques, simulated data are used to investigate model fit based on Moran's statistic under discrete time Markov chain. Diagnostics and comparison statistics are performed. Calibration of clusters for predictive locations is performed to enhance understanding of Moran's autocorrelation, trend in time, and functionality of model analysis.

## 2. Method

### 2.1 Temporal Version of Moran Spatial Autocorrelation

For a measurable space  $S = \mathbb{R}^d$ , a domain  $D \subset S$ , and a collection of random points  $x_1, x_2, \dots$ , let  $N(D)$  represent the count or number of points  $x_i$  that capture location occurrence of some disease/event in the domain  $D$ . Then  $N(D)$  can be seen as a measure and is called a point process. For nonoverlapping areas  $D_1, D_2, \dots$ ,  $N(D_1), N(D_2), \dots$  are mutually independent and for each area  $D$ ,  $N(D)$  follows a Poisson distribution with mean  $\lambda|D|$ , where  $|D|$  is the area (Lebesgue) measure of  $D$  and  $\lambda$  is the intensity of the process. As a process, a counting may be associated to it and is defined as:

$$N(D) = \sum_{i=1}^{\infty} \mathbf{1}_D(x_i), \text{ where } \mathbf{1}_D(x_i) = \begin{cases} 1 & \text{if } x_i \in D, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

To every domain, we consider an associated intensity measure  $\lambda$  with  $\lambda(D) = E(N(D))$ , the expected number of points that are obtained in  $D$ . Such measure can be assumed to be finite if  $N(D) < \infty$ , hence the name finite measure.

Here, we consider the *nonhomogeneous Poisson point process* with non-constant intensity function  $\lambda$  in time say,  $\lambda(t)$  with  $t \in [0, T]$  with the main assumption that points are independent of each other. The nonhomogeneous Poisson process has the following three important properties as described in Baddeley et al. (2016):

- **conditional property:** given exactly  $n$  points in a region  $D$ , these points are mutually independent and each point has the same probability distribution over  $D$ , with probability density  $f(t) = \lambda(t)/\mu$ , where  $\mu = \int_D \lambda(t)dt$ .
- **superposition property:** If  $Y_1, Y_2, \dots, Y_n$  are independent Poisson random variables with means  $E(Y_i) = \lambda_i$ , then  $\sum_i Y_i \sim \text{Poisson}(\sum \lambda_i)$ .

We consider the conditional distribution of the occurrence of disease in the partitioned subareas defined by Voronoi cells induced by a set of points (sites) as described in Kallenberg (2001). These cells are defined as:

$$V_\mu = \{s \in \mathbb{R}^d; \mu(D + s) < \infty\}, \quad \mu \in \mathcal{N}(\mathbb{R}^d),$$

where  $\mathcal{N}(\mathbb{R}^d)$  is the class of locally finite measures on  $\mathbb{R}_+$  and  $D_r$  denotes the ball of radius  $r$  around the point  $s$ . Doing so, we have a local precise measurement of the density and neighborhoods. We set a partition of  $D$  into subareas  $D_1, D_2, \dots$  where a finite measure or “locally finite” measure is such that  $\mu D_k < \infty$  for all  $k \geq 1$ .

Note that  $\lambda D = \lambda(\cdot, D)$  is a random variable in  $[0, +\infty]$  for every  $D \subseteq S = \mathbb{R}^d$ . Such an idea is also described in Thäle and Yukich (2016) with properties of Poisson-Voronoi generated sets.

Using Equation (1), Vaillant et al. (2011) modeled the propagation of sugar cane yellow leaf virus with a focus on the spatial spread of disease over six time periods (weeks 6, 10, 14, 19, 23, and 30 in the growing season). For each pair of consecutive observation dates  $(t_{i-1}, t_i)$ ,  $i = 1, 2, \dots, 6$ , they defined a Moran-type index based on a nearest neighbor scheme as follows:

$$M_i = \sum_{(x,y) \in D} w_{x,y} \mathbf{1}_{[0,t_{i-1}]}(T_x) \mathbf{1}_{(t_{i-1},t_i]}(T_y),$$

where  $D$  denotes the discrete set of plant locations,  $T_x$  denotes the date (time variable) of virus detection for plant  $x$ ,  $\mathbf{1}_{[0,t_{i-1}]}(T_x)$  is an indicator whether time  $T_x$  falls in the interval  $[0, t_{i-1}]$ , and  $w_{x,y}$  denotes the weights, which are non-zero only if  $x$  and  $y$  are neighbors, located within the same subarea. There are several possible weight functions as we will describe later in this section.

We will utilize a similar space-time definition of Moran’s index. Since the partitioned areas vary from one time to the next, we define  $D_k^t$  as subarea  $k \geq 1$  at time  $t$  and then we will define a Moran-type autocorrelation index on each measurable subset  $D_k^t$  as:

$$M_k^t = \sum_{u,u' \in D_k^t} w_{u,u'} \mathbf{1}_{[t-1,t)}(T_u, T_{u'}), \quad k \geq 1, \quad t \geq 1, \quad (2)$$

where  $w_{u,u'}$  denotes the spatial weight between points  $u$  and  $u'$  of disease/event occurrence,  $T_u$  and  $T_{u'}$  denote the time of detection of  $u$  and  $u'$ , and  $\mathbf{1}_{[t-1,t)}(T_u, T_{u'})$  is an indicator of whether times  $T_u$  and  $T_{u'}$  both fall in the interval  $[t - 1, t)$ .

In the definition of Moran’s autocorrelation index,  $w_{u,u'}$  represents a spatial weight between any two distinct event locations generated within disc  $D_k^t$ , which could be a function, e.g.,

- (i) the inverse distance between two points;
- (ii) the inverse distance squared between two points;
- (iii) an estimate of the autocorrelation/semivariance statistic;
- (iv) the “geographical” weights defined as  $w_{ij} = \begin{cases} e^{(-d_{ij}/r)} & j \neq i, \\ 0 & \text{otherwise,} \end{cases}$   
 where  $r$  is the maximum distance in the minimum tree that spans all points and does not have any nodes that link back to itself as in Murakami et al. (2017).

As stated in Chen (2012), selection of the weight function objectively is an open question. Different wight functions have different spatial effects. Later in Section 4, we give reasons why we choose a particular weight function.

Also, while values of the Moran-type statistic may not always be between -1 and 1 (the full range depends on the weights), they are essentially in the same spirit as larger values indicate larger autocorrelation measures.

We will use a modified version of Geary’s  $C$  autocorrelation index (Geary, 1954; Sokal, Oden, & Thomson, 1998) as a comparison to determine and evaluate the trend between the two measures. We will define it as  $C_k^t$ :

$$C_k^t = \frac{n-1}{2W_0} \sum_{u,u' \in D_k^t} w_{u,u'} \mathbf{1}_{(0,t-1]}(T_u) \mathbf{1}_{(t-1,t]}(T_{u'}),$$

with  $w_{u,u'}$  as in Equation (2) and  $W_0 = \sum_{u,u' \in D^t} w_{u,u'}$ , the sum of weights for all points generated across all discs at time  $t$ .

### 2.2 Spatio-temporal Process

For a time period of length  $T$  partitioned into  $m$  time subintervals  $t_0 = 0 < t_1 < t_2 < \dots < t_r < \dots < t_m$ , consider  $[t_{r-1}, t_r)$ ,  $r = 0, 1, \dots, m$ ,  $0 \leq t_r \leq T$ . The number of random occurrences or points generated within  $(t_{r-1}, t_r)$  in state space  $D$  are tied to locations which are sequences of nested counts of total number of points generated at time  $t_r$ , denoted  $n(t_r)$ , and sequence of points  $k(t_1), k(t_2), \dots, k(t_m)$ , where  $1 \leq k(t_1) \leq n(t_1); 1 \leq k(t_2) \leq n(t_2); \dots; 1 \leq k(t_r) \leq n(t_r); \dots; 1 \leq k(t_m) \leq n(t_m)$ , and  $n(t_r)$  represents the total number of points generated at time  $t_r$  within specific locations. The mathematics then requires notation capable of representing outcomes across multiple time point steps and subareas. The number of occurrences is defined as  $X^{t_r}$  and the location of event occurrences is defined as  $X^{t_r}$ .

The process follows a discrete time Markov chain as defined in Resnick (2002):

$$P(X^{t_r} = j | X^{t_1} = i_1, \dots, X^{t_{r-2}} = i_{r-2}, X^{t_{r-1}} = i_{r-1}) = P(X^{t_r} = j | X^{t_{r-1}} = i_{r-1})$$

for some function of propagation from location  $i$  to location  $j$  and subarea containing points  $i$  and  $j$  in time interval  $(t_{r-1}, t_r)$ . The associated count within  $D_k^t$  defines a sequence  $\{X^{t_r}\}$  of number of occurrences within  $[t_{r-1}, t_r)$  in subarea  $D_k^t$  and follows the Markov property; i.e. given the current state of the system, we can make predictions about the future state without regard for previous states. This is a discrete time finite Markov chain.

The transition probability from location  $i$  to location  $j$  in interval time  $(t, t+s)$ , denoted by  $P_{ij}(s)$ ,  $n_i$  and  $n_j$ . The conditional probability is:

$$P(X(t+s) = j | X(t) = i) =: P_{ij}(s), \quad i, j \in S, \quad s, t \geq 0$$

To minimize complexity, we define the count process  $\{X^t, t \geq 0\}$  of occurrences between consecutive times  $t_{r-1}$  and  $t_r$  to be

$$P_{ij}(t_r - t_{r-1}) \sim P(\lambda(t)|D),$$

where  $\lambda(t)$  denotes the local intensity for any  $t$  within  $[t_{r-1}, t_r)$ . Such a process has stationary transition probabilities and transition matrix of the counts  $Q = (Q_{ij})$  constructed from the transition probabilities,  $P_{ij}(s)$ , where  $Q_{ij}$  is the  $(i, j)$  count from  $P_{ij}(s)$  from consecutive time periods  $t_1, t_2, \dots, t_m$  adjusted with subarea containing locations  $i$  and  $j$ , and  $Q_{ii} = 0, \forall i \in S$ , at each time period.

We extend  $D_k^t$  such that they form a nested sequence  $D_{k(t_r)}^{t_r} \subset D_{k(t_{r-1})}^{t_{r-1}}$ , the subarea within time  $(t_{r-1}, t_r)$  and count  $X_{k(t)}^{t_r}$ ,  $1 \leq k(t) \leq n(t)$ . Following Resnick (2002), we can define the nested subareas  $D_{k(t)}^t$  and each fixed time  $t$ , a sequence  $\{E_{k(t)}^t\}_{k \geq 0}$  of independent and identically distributed exponential random variables with unit mean such that  $E_{k(t)}^t$  is independent of  $X(t)$  on  $D_{k(t)}^t$ ,  $1 \leq k(t) \leq n(t)$  and set  $X_{k(t)}^t$  equal to the count within  $D_{k(t)}^t$  for interval time  $[t_{r-1}, t_r)$ . Now define

$W(D_{k(t)}^t) = \frac{E_{k(t)}^t}{\lambda |D_{k(t)}^t|}$ , where  $|D_{k(t)}^t|$  is the area (Lebesgue) measure within  $D_{k(t)}^t$ . Then

$$P(W(D_{k(t)}^t) > u | X_{k(t)}^t = i \in D_{k(t)}^t) = e^{-\lambda |D_{k(t)}^t| u},$$

for  $u > 0$ ,  $1 \leq k(t) \leq n(t)$ , and at location  $i \in D_{k(t)}^t$ .

By discretizing time, we then define two finite sequences  $\{(X^{t_r}), (t_r)\}$ , where  $X^t = \sum_{k(t)=1}^{n(t)} X_{k(t)}^t$  with locations  $X_{k(t)}^t$  for  $t_n < t < t_{n+1}$  and  $\{t_r\}$ ,  $r = 1, \dots, m$  is the sequence of times when the process is observed. The sequence  $\{X^t\}$  is the cumulative count of occurrence within interval time  $[t_{r-1}, t_r]$  in a subset of space  $D$  such that  $t_r - t_{r-1}$  is conditionally independent and exponential given  $X^{t_{r-1}}$ . More precisely, the sequence  $\{X_{k(t)}^t\}$  is defined as follows:

$$\begin{aligned} X_1^{t_1} &= X^{t_0} \mathbf{1}_{D_1^{t_1}} + W(D_1^{t_1}), \\ X_2^{t_1} &= X^{t_0} \mathbf{1}_{D_2^{t_1}} + W(D_2^{t_1}), \\ &\vdots \\ X_{k(t_1)}^{t_1} &= X^{t_0} \mathbf{1}_{D_{k(t_1)}^{t_1}} + W(D_{k(t_1)}^{t_1}), \\ &\vdots \\ X_{n(t_1)}^{t_1} &= X^{t_0} \mathbf{1}_{D_{n(t_1)}^{t_1}} + W(D_{n(t_1)}^{t_1}), \end{aligned}$$

and  $X^{t_1} = \sum_{k(t_1)=1}^{n(t_1)} X_{k(t_1)}^{t_1}$  ;

$$\begin{aligned} &\vdots \\ X_1^{t_r} &= X^{t_{r-1}} \mathbf{1}_{D_1^{t_r}} + W(D_1^{t_r}), \\ &\vdots \\ X_{n(t_r)}^{t_r} &= X^{t_{r-1}} \mathbf{1}_{D_{n(t_r)}^{t_r}} + W(D_{n(t_r)}^{t_r}), \end{aligned}$$

and  $X^{t_r} = \sum_{k(t_r)=1}^{n(t_r)} X_{k(t_r)}^{t_r}$  ;

$$\begin{aligned} &\vdots \\ X_1^{t_m} &= X^{t_{m-1}} \mathbf{1}_{D_1^{t_m}} + W(D_1^{t_m}), \\ &\vdots \\ X_{n(t_m)}^{t_m} &= X^{t_{m-1}} \mathbf{1}_{D_{n(t_m)}^{t_m}} + W(D_{n(t_m)}^{t_m}), \end{aligned}$$

and  $X^{t_m} = \sum_{k(t_m)=1}^{n(t_m)} X_{k(t_m)}^{t_m}$ ,

where  $n(t_r)$  is the total number of points generated at time  $t_r$ ,  $W(D_{k(t)}^t) = \frac{E_{k(t)}^t}{\lambda |D_{k(t)}^t|}$ , and  $X(t)$  is the total count for  $t_r \leq t < t_{r+1}$ ,  $1 \leq r \leq m$ .

Setting  $t_\infty = \lim_{r \rightarrow \infty} t_r$  the process  $\{X^t\}$  becomes explosive. However,

$$X^{t_r} = \sum_{k(t_r)=1}^{n(t_r)} X_{k(t_r)}^{t_r} \mathbf{1}_{[t_{r-1}, t_r)}$$

or  $X^t = X^{t-1} + \sum_{k(t)=1}^{n(t)} W(D_{k(t)}^t)$ ,  $t \in [t_{r-1}, t_r]$ ,  $1 \leq r \leq m$ .

The pair sequence  $\{X^t, t_r\}$  has the following two properties.

- (i) The sequence  $\{t_r - t_{r-1}, 1 \leq r \leq m\}$  are conditionally independent and exponentially distributed given  $\{X^n\}$ ; i.e. memoryless in time: for  $u_m > 0, m = 1, \dots, n$ ,

$$\begin{aligned}
 &P(X^{t_r} - X^{t_{r-1}} > u_r, 1 \leq r \leq m | X_0^t = i_0, \dots, X^{t_{r-1}} = i_{r-1}) = \\
 &P\left(\sum_{k(t_r)=1}^{t_r} W(D_{k(t_r)}^{t_r}) > u_r | X^{t_{r-1}} = i_{r-1}\right) = \prod_{k(t_r)=1}^{n(t_r)} P\left(\frac{E_{k(t_r)}^{t_r}}{|\lambda D_{k(t_r)}^{t_r}|} > u_r\right) \\
 &= \prod_{k(t_r)=1}^{n(t_r)} e^{-\lambda |D_{k(t_r)}^{t_r}| u_r} = \exp\left\{-\sum_{k(t_r)=1}^{t_r} \lambda |D_{k(t_r)}^{t_r}| \cdot u_r\right\}.
 \end{aligned}$$

- (ii) The distributional structure of  $\{X^t, (t_r)\}$  is a function of the transition matrix and the limiting function: For  $u > 0, i \in S$ ,

$$P\left(X_{k(t_r)}^{t_r} = j; t_r - t_{r-1} > u | X^{t_{r-1}} = i\right) = Q^t \cdot \exp\left\{-\sum_{k(t)=1}^{n(t_r)} \lambda |D_{k(t)}^t| u\right\}, \tag{3}$$

$$t \in [t_{r-1}, t_r), Q^t = (Q_{ij}^t).$$

The nested sequence is shown in Figure 1, where  $E_{k(t)}^t$  equals the count at time  $t$  for subarea  $D_{k(t)}^t$ , with  $1 \leq k(t) \leq n(t)$ , and  $n(t)$  represents the number of points generated at time  $t$  and  $Q$  is the  $(i, j)$  transition probability of count from locations  $i$  to  $j$  between times  $t_{r-1}$  and  $t_r$ .

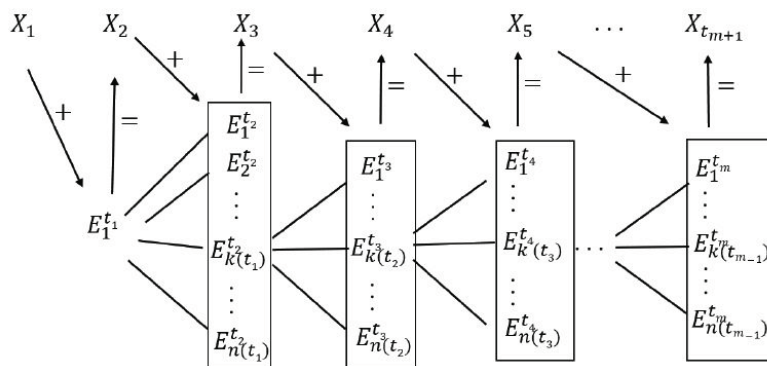


Figure 1. Markov process

The special case of transition probabilities following a Poisson process and where the  $\lambda(t)|D_{k(t)}^t|$  are proportional to time plays an important role as it naturally occurs in many statistical phenomena. If we assume that the number of points generated in subarea  $D_k^t$  has a Poisson density, the density function is defined as:

$$P_{ij}(t) = f(n_j(t)|\lambda(t)) = \frac{e^{-\lambda(t)|D_k^t|} (\lambda(t)|D_k^t|)^{n_j(t)}}{n_j(t)!}, \quad n_j(t) \in \mathbb{N},$$

whereas in Equation (3),  $Q_{ij}^t = \frac{(\lambda|D_k^t|)^{n_j(t)}}{n_j(t)!}$ . We propose a comparable implementation and compute the successive Moran's spatio-temporal autocorrelations under such setting.

### 2.3 Simulation Study

Here we provide an illustration of the theory built in the previous section. We conduct the simulation using R with the spatstat package and represent the weights of the Moran's index as the inverse distance between two points. The

weight function was chosen as described in Chen (2012) since we are modeling spread of disease which is a large scale complex system and we intend to capture transitions globally to locally. We begin with an observed unit area ( $1 \times 1$  unit) and generate a Poisson point processes within that area. The Poisson process is used as general setup since it is widely applicable and is applicable to many real world situations such as disease cases, radioactive decay, and plant propagation. Other processes, for example negative binomial and Cox, may be utilized to explain other space-time point phenomena.

Next, we define the Voronoi cells around each point (e.g. the locus of locations nearer to that point than any other). Each location site is then used as the center point to generate a disc with diameter equal to the minimum of the distance to the Voronoi edge and the distance to the observed area edge (edge of  $D$ ). This effectively means that an infected site can only subsequently infect sites that are closer to it than any other infected site. These Voronoi discs define the subareas  $D_{k(t)}^i$  for the point process generated at the next time interval (step). In application, subareas can be tied to locations such as country, state, county, and zip code. The choice of subareas could be used to determine whether more affluent regions are able to deal more effectively with disease outbreak, for example. The Moran statistic is calculated at each step as the sum of the inverse distances between an offspring site and another offspring site. This format will continue through time  $t = 4$  adhering to the Markov property ( $t = 4$  chosen to follow results from a forestry example in Meddens and Hicke, 2014; or as in Vaillant et al., 2011). Each subarea will be magnified by a scale parameter,  $\alpha > 0$ , such that the new radius is  $\alpha$  times the previous radius with intensity function  $\lambda * t$ , following the spatial resolution of aggregated grid cells that experienced a 1% tree mortality as in Meddens and Hicke (2014).

**Algorithm:** Iterative local Moran statistics

**Procedure:**

- (i) Define the partition and generate a Poisson process in the area with intensity  $\lambda$ .
- (ii) Calculate the Moran statistic.
- (iii) Define new subareas with point from previous process as center and radius equal to the minimum of distance to the edge of the window or distance to the edge of the Voronoi cell.
- (iv) Rescale subarea and iterate local points from a Poisson process with intensity  $\lambda * t$ .
- (v) Compute the local Moran statistic for each subarea at that time step.

**Repeat** steps (iii) through (v).

**Stop** when all times are reached.

**End**

Figure 2 shows the original observed area with the seven points generated at  $t = 1$  alongside the discs generated from these points and the points generated at  $t = 2$ .

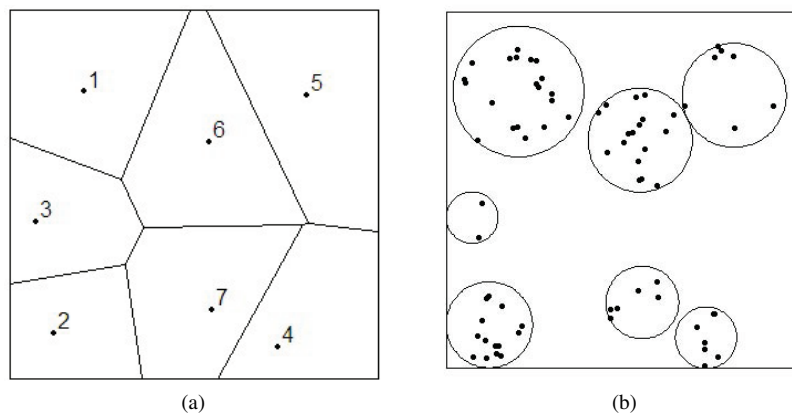


Figure 2. Poisson point process plots,  $\lambda = 4$  (a)  $t = 1$  and (b)  $t = 2$

### 3. Results

The Moran value for  $t = 1$  equals 42.57. The number of points generated in each Voronoi subarea, the rescaled area, and Moran values generated at time  $t = 2$  are displayed in Table 1 with  $\lambda = 4$  and  $\alpha = 5$ . For example, at point 1 in Figure 2(a), we generate a disc of maximum radius within the Voronoi cell centered at the point generated at the previous time with 20 points as shown in Figure 2(b). The number of points generated in other Voronoi cells is displayed in Table 1. The Geary's values are also computed. The ordering of Moran and Geary's values display resemblance as they "rank" the strength of correlations in a similar way.

Table 1. Moran and Geary's  $c$  values for  $\lambda = 4$  and  $\alpha = 5$  at  $t = 2$

| disc | Points | Area  | Moran  | Geary's $C$ |
|------|--------|-------|--------|-------------|
| 1    | 20     | 2.650 | 409.00 | 6.773       |
| 2    | 14     | 1.148 | 315.07 | 5.217       |
| 3    | 2      | 0.411 | 2.09   | 0.035       |
| 4    | 7      | 0.585 | 120.95 | 2.003       |
| 5    | 7      | 1.676 | 55.61  | 0.921       |
| 6    | 17     | 1.676 | 350.45 | 5.803       |
| 7    | 6      | 0.817 | 51.21  | 0.848       |

The process is described at time  $t = 3$  with disc 1 having 20 subareas (see Figure 3), disc 2 having 14 subareas, etc. The non-zero Moran values for disc 1 at time  $t = 3$  are listed in Table 2.

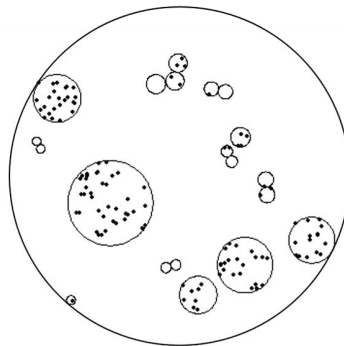


Figure 3. Disc 1 at  $t=3$

Table 2. Non-zero Moran and Geary's  $c$  values for disc 1 at  $t = 3$

| disc | Points | Area  | Moran   | Geary's $C$ |
|------|--------|-------|---------|-------------|
| 1    | 24     | 1.269 | 676.94  | 8.941       |
| 4    | 2      | 0.135 | 2.93    | 0.039       |
| 5    | 39     | 4.037 | 1176.21 | 15.536      |
| 6    | 22     | 1.723 | 628.83  | 8.306       |
| 7    | 8      | 0.785 | 84.38   | 1.115       |
| 8    | 2      | 0.072 | 7.90    | 0.104       |
| 9    | 15     | 1.183 | 262.56  | 3.468       |
| 12   | 4      | 0.212 | 29.42   | 0.389       |
| 14   | 3      | 0.185 | 13.75   | 0.182       |

After four time intervals, the clustering of points generated is displayed in Figure 4, with red shades representing the most dense areas and white shades representing areas with no instances of disease.



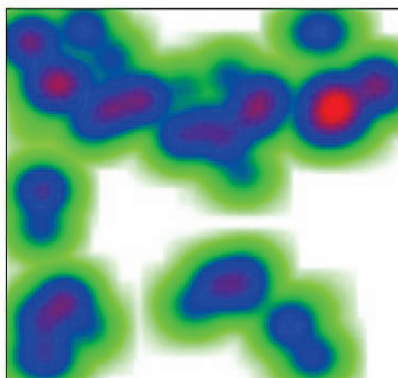


Figure 4. Apparent clustering through  $t = 4$

Such a representation would not have been clearly observed if the area was not partitioned and if the rate of spread was not tabulated in space and time. This framework will allow us to detect clusters and estimate the density.

#### 4. Discussion

##### 4.1 Model Analysis

In this section estimation of the parameter,  $\lambda$ , for the Poisson distribution given the data generated in the simulation is proposed using the maximum likelihood method as presented in Baddeley et al. (2016). A scale parameter is incorporated into the Poisson distribution through the use of the area of the Voronoi subarea  $D_k^t$ ,  $t = 1, 2, 3, 4$ ,  $1 \leq k \leq n(t)$ .

The parameter estimate becomes  $\hat{\lambda}(t) = \frac{\sum n_j}{\sum |D_k^t|}$  with a standard error of  $SE = \sqrt{\frac{\hat{\lambda}}{n}}$ , where  $n = n(t) =$  number of subareas  $D_k^t$ .

Using the data generated in the simulation above, we tabulate the estimates of  $\lambda$  at  $t = 2, 3$ , and 4 with different scale parameters,  $\alpha$ . Table 3 shows these parameter estimates of  $\lambda$  with standard error in parenthesis. For  $\alpha = 5$  at time  $t = 2$ ,  $\hat{\lambda} = 8.14$  with associated  $SE=1.07$  and the true parameter is  $\lambda = 8$ . The data appear to follow a Poisson distribution with intensity  $\lambda * t$ ; however, the estimated parameter values for  $\alpha = 2.5$  seem less reliable due to the sparse number of points generated at each time step.

Table 3.  $\hat{\lambda}$  values at times 2, 3, & 4 with SE in parenthesis

| time | $\alpha$     |              |              |              |
|------|--------------|--------------|--------------|--------------|
|      | 2.5          | 5            | 7.5          | 10           |
| 2    | 10.26 (1.21) | 8.14 (1.08)  | 8.08 (1.07)  | 8.23 (1.08)  |
| 3    | 10.76 (0.68) | 12.45 (0.41) | 11.69 (0.27) | 12.08 (0.20) |
| 4    | 10.63 (0.67) | 16.24 (0.18) | 16.01 (0.08) | 16.04 (0.04) |

We also investigated increasing  $\alpha$  as function of time,  $\alpha(t) = \alpha^{t-1}$ . Using an initial value of  $\alpha = 2.5$ , the parameter estimates at  $t = 2, 3, 4$  were 10.26 (1.21), 13.44 (0.76), and 15.98 (0.31), respectively. By making the scale parameter a function of time, the parameter estimates appear more precise for small starting values, but under this construct, using  $\alpha(t) = 10^{t-1}$ , generates up to 30,000 points per subarea at  $t = 4$ . As a result, we will use  $\alpha = 5$  for the scale parameter in the following analysis.

Next we investigate any trend in our Moran index versus number of points generated and area of the disc. Figure 5 shows Moran's values plotted versus area and versus number of points. The plots of Moran values for time  $t = 2$  and for disc 1 at  $t = 3$  show an approximate linear trend between Moran and number of points generated, even though there is no apparent relationship between area and Moran.

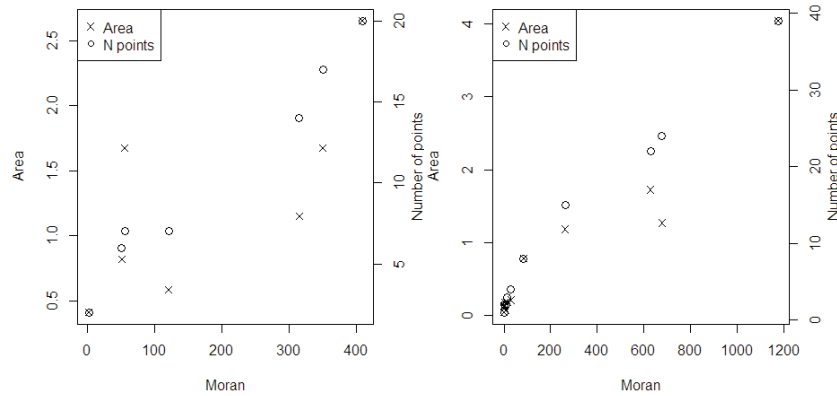


Figure 5. Trend analysis (a)  $t = 2$  and (b)  $t = 3$  for disc 1

The 3-D plot in Figure 6 shows that at each successive time point the number of points generated versus the Moran gets progressively larger. This is due to the number of points generated increasing as time increases, but the area increase is moderated as time increases. This results in points that are closer together, thus increasing the sum of inverse distances and the Moran value.

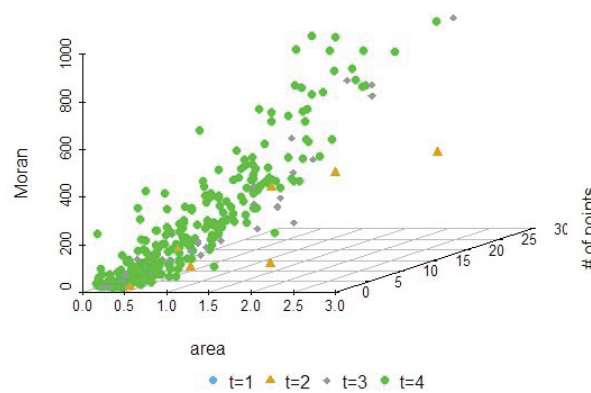


Figure 6. Plot of Moran values

Focusing on the seven points generated at  $t = 1$  and the associated discs at  $t = 2$ , we investigate the Markov property of the simulation results. Table 4 shows the number of points generated within each of the original seven discs at times  $t = 2, 3, 4$  and also includes the area of each disc.

Table 4.  $\hat{\lambda}$  values at times 2, 3, & 4 with SE in parenthesis

| disc | area  | time |     |     |
|------|-------|------|-----|-----|
|      |       | 2    | 3   | 4   |
| 1    | 2.650 | 20   | 124 | 806 |
| 2    | 1.148 | 14   | 66  | 391 |
| 3    | 0.411 | 2    | 16  | 138 |
| 4    | 0.585 | 7    | 25  | 143 |
| 5    | 1.676 | 7    | 113 | 745 |
| 6    | 1.676 | 17   | 89  | 565 |
| 7    | 0.817 | 6    | 58  | 226 |

The transition matrix for each disc  $Q_k^t$  is developed using the ratio of points generated at each time to the total number of points generated over all times. This matrix is upper triangular since we cannot transition backward in time or stay in the

same time.

$$Q'_k = \begin{bmatrix} 0 & \frac{n_{12}}{\sum} & \frac{n_{13}}{\sum} & \frac{n_{14}}{\sum} \\ 0 & 0 & \frac{n_{23}}{\sum} & \frac{n_{24}}{\sum} \\ 0 & 0 & 0 & \frac{n_{34}}{\sum} \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad Q'_1 = \begin{bmatrix} 0 & 0.0211 & 0.131 & 0.848 \\ 0 & 0 & 0.133 & 0.867 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where  $\sum = \sum_{r=1}^4 n(t_r)$  for disc  $k$ .

We then conduct cluster analysis to determine similarities between discs. Using the Canberra distance measure,  $\sum \frac{|X_k - X'_k|}{|X_k + X'_k|}$ , to create a distance matrix using the counts in Table 4, the dendrogram tree in Figure 7 was generated to look at similarities in clusters. It can be inferred that disc association is related to area and abundance of occurrence of events. The separation of clusters two and three to govern the tree shows that even though they are spatial neighbors, they are perceived to reflect distinct evolutionary paths. Such finding could add to the conclusion given in Meddens and Hick (2014) when they state that spectral responses of lodgepole pine, ponderosa pine, and mountain pine beetles lead to different stages of tree maturity.

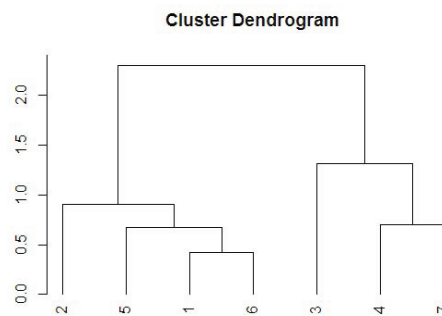


Figure 7. Cluster analysis

#### 4.2 Summary

Our designed model has shown that global properties of spatio-temporal disease spread can be tuned in by modifying the space and adding time efficiently. The results show that long range, spread feature and variability yielded spatial clustering of occurrence events. To obtain a quantitative analysis of the performance of the Moran’s values, we compared them with adjusted Geary’s  $C$  statistics. We controlled for targeted subareas and such an approach provides a framework for understanding “disorganized” or “disordered” evolution of some natural organisms in the form of cluster analysis. The distribution of subareas can be found based on our density estimation or concentration under the modified Moran’s Index in space and time.

Extension to different dynamics other than the Poisson will be considered in future studies and widen the scope of the spatio-temporal model incorporating covariates.

#### Acknowledgements

The research for this paper was made possible by financial support from the U.S. Naval Academy. Material contained herein is solely the responsibility of the authors and is made available for the purpose of peer review and discussion. Its contents do not necessarily reflect the views of the Department of the Navy or the Department of Defense.

#### References

Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, 27(2), 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>

Baddeley, A., Rubak, E., & Turner, R. (2016). *Spatial Point Patterns: Methodology and Applications with R*, CRC Press, Boca Raton.

Chen, Y. (2012). On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5, 65–72. <https://doi.org/10.1007/s12076-011-0076-6>

Cliff, A., & Ord, J. (1975). Space-time modelling with an application to regional forecasting. *Transactions of the Institute of British Geographers*, 64, 119–128. <https://doi.org/10.2307/621469>

- Geary, R. C. (1954). The contiguity ratio & statistical mapping. *The Incorporated Statistician*, 5(3), 115–145. <https://doi.org/10.2307/2986645>
- Kallenberg, O. (2001). *Foundations of Modern Probability*, 2nd ed., Springer, New York.
- Lee, J., & Li, S. (2017). Extending Moran's index for measuring spatiotemporal clustering of geographic events. *Geographical Analysis*, 49, 36–57. <https://doi.org/10.1111/gean.12106>
- Martin, R. L., & Oeppen, J. E. (1975). The identification of regional forecasting models using space:time correlation functions. *Transactions of the Institute of British Geographers*, 66, 95–118. <https://doi.org/10.2307/621623>
- Meddens, A. J. H., & Hicke, J. A. (2014). Spatial and temporal patterns of Landsat-based detection of tree mortality caused by a mountain pine beetle outbreak in Colorado, USA. *Forest Ecology and Management*, 322-15, 78–88. <https://doi.org/10.1016/j.foreco.2014.02.037>
- Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1-2), 17–23. <https://doi.org/10.1093/biomet/37.1-2.17>
- Murakami, D., Yoshida, T., Seay, H., Griffith, D. A., & Yamagata, Y. (2017). A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*, 19, 68–69. <https://doi.org/10.1016/j.spasta.2016.12.001>
- Pace, R. K., & LeSage, J. P. (2009). Omitted variable and spatially dependent variables. In *Progress in Spatial Analysis: Advances in Spatial Science*, ed. A. Páez et al., Springer-Verlag, Berlin Heidelberg, 17–28.
- Resnick, S. I. (2002). *Adventures in Stochastic Processes*, Birkhäuser, Boston. <https://doi.org/10.1007/978-1-4612-0387-2>
- Sokal, R. R., Oden, N. L., & Thomson, B. A. (1998). Local spatial autocorrelation in a biological model. *Geographical Analysis*, 30(4), 331–354. <https://doi.org/10.1111/j.1538-4632.1998.tb00406.x>
- Thäle, C., & Yukich, J. E. (2016). Asymptotic theory for statistics of Poisson-Voronoi approximation. *Bernoulli*, 22(4), 2372–2400. <https://doi.org/10.3150/15-BEJ732>
- Vaillant, J., Puggioni, G., Waller, L. A., & Daugrois, J. (2011). A spatio-temporal analysis of the spread of sugarcane yellow leaf virus. *Journal of Time Series Analysis*, 32, 392–406. <https://doi.org/10.1111/j.1467-9892.2011.00730.x>
- Wang, Y. F., & He, H. L. (2007). *Spatial Data Analysis Method*, Science Press, Beijing.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).