Spatio-temporal statistics (MATH4341) Michaelmas term, 2024

Lecture notes part 3: Aerial unit data / spatial data on lattices

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Aim. To introduce Aerial unit data modeling: the basic building models.

Reading list & references:

[1] Cressie, N. (2015; Part II). Statistics for spatial data. John Wiley & Sons.
[2] Kent, J. T., & Mardia, K. V. (2022). Spatial analysis (Vol. 72). John Wiley & Sons.
[3] Gaetan, C., & Guyon, X. (2010; Ch 3). Spatial statistics and modeling (Vol. 90). New York: Springer.

Part 1. Basic stochastic models & related concepts for model building

Note 1. Recall from Section 2.2 of "Lecture notes part 1: Types of spatial data" that modeling aerial unit / lattice data types involves the use of random field models with a discrete index set. Such data are collected over areal units such as pixels, census districts or tomographic bins. Often, there is a natural neighborhood relation or neighborhood structure.

Note 2. This means we need to introduce suitable basic building models able to represent the characteristics of the underline data generating mechanisms. These as the "Discrete Random Fields".

1. Discrete Random Fields

Note 3. We re-introduce the definition of the random field with regards to the aerial unit data framework.

Definition 4. A random field $Z = (Z_s; s \in \mathcal{S})$ on a set of indexes $\mathcal{S}$ taking values in $\mathcal{Z}^{\mathcal{S}}$ is a family of random variables $\{Z_s := Z_s(\omega); s \in \mathcal{S}, \omega \in \Omega\}$ where each $Z_s(\omega)$ is defined on the same probability space $(\Omega, \mathfrak{F}, \mathrm{pr})$ and taking values in $\mathcal{Z}$.

Note 5. In aerial unite data modeling, the (spatial) set of sites $\mathcal{S}$, at which the process is defined, is discrete, it can be finite or infinite (e.g. $\mathcal{S} \subseteq \mathbb{Z}^d$), regular (e.g. pixels of an image) or irregular (states of a country).

Note 6. The general state space $\mathcal{Z}$ of the random field can be quantitative, qualitative or mixed. E.g., $\mathcal{Z} = \mathbb{R}_+$ in a Gamma random field, $\mathcal{Z} = \mathbb{N}$ in a Poisson random field, $\mathcal{Z} = \{0, 1\}$ in a binary random field.

Note 7. If $\mathcal{Z}$ is finite or countably infinite, the (joint)distribution of Z has a PMF

$$\mathrm{pr}_Z(z) = \mathrm{pr}(Z = z) = \mathrm{pr}(\{Z_s = z_s;\ s \in \mathcal{S}\}),\ \forall z \in \mathcal{Z}^{\mathcal{S}}$$

otherwise if $\mathcal{Z} \subseteq \mathbb{R}^d$ and Z continuous we will use the joint PDF.

Definition 8. The discrete set of sites $\mathcal{S} = \{s_i; i = 1, ..., n\}$ is often called lattice of sites.

Notation 9. Often we will use the notation Z_s instead of $Z(s)$ or Z_i instead of $Z(s_i)$. Hence, since $\mathcal{S} = \{s_i; i = 1, ..., n\}$, we can consider a more convenient notation

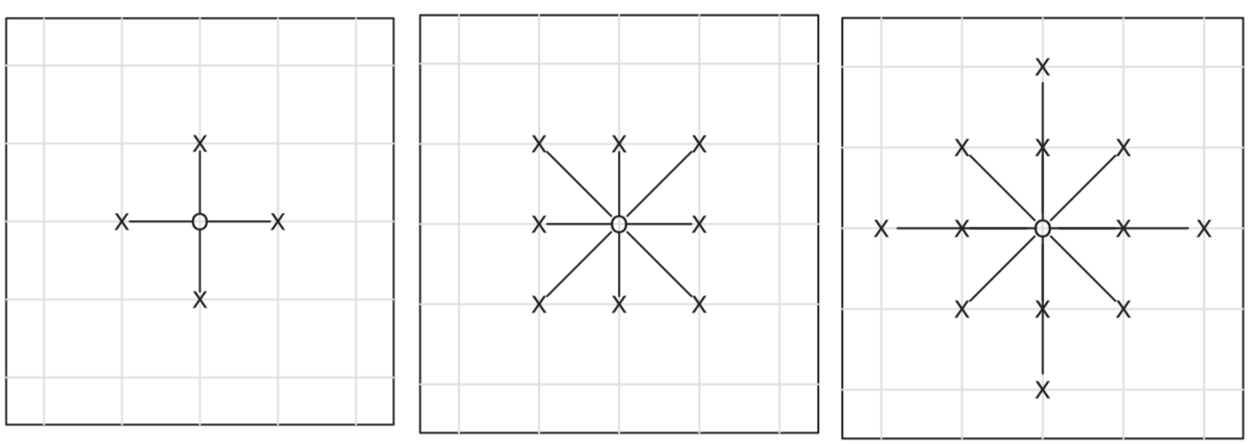
$$Z = (Z_s; s \in \mathcal{S})^\top = (Z_i = Z(s_i); i = 1, ...n)^\top.$$

Note 10. Modeling aerial unit data often requires the specification of a neighborhood relation or neighborhood structure.

Notation 11. The notation $i \sim j$ between two sites $i, j \in \mathcal{S}$ means that "sites i and j are neighboring" according to a "neighborhood relation" $\sim$.

Definition 12. Given a lattice of sites $\mathcal{S}$ and "neighborhood relation" $\sim$, we can define the neighborhood $\mathcal{N}_s$ of $s \in \mathcal{S}$ as

$$\mathcal{N}_s = \{s' \in \mathcal{S} : s \sim s'\}$$



Definition 13. Proximity matrix W is called a matrix W which aims at spatially connecting unites i and j in some fashion given some symmetric neighborhood relation $\sim$ on $\mathcal{S}$. Usually $[W]_{i,i} = 0$.

Note 14. Proximity matrix W may be such that it represents the neighborhood relation $\sim$ in a binary fashion e.g.

$$[W]_{i,j} = \begin{cases} 1 & \text{if } i \sim j \text{ and } i \neq j \\ 0 & \text{if } i \nsim j \text{ or } i = j \end{cases}$$

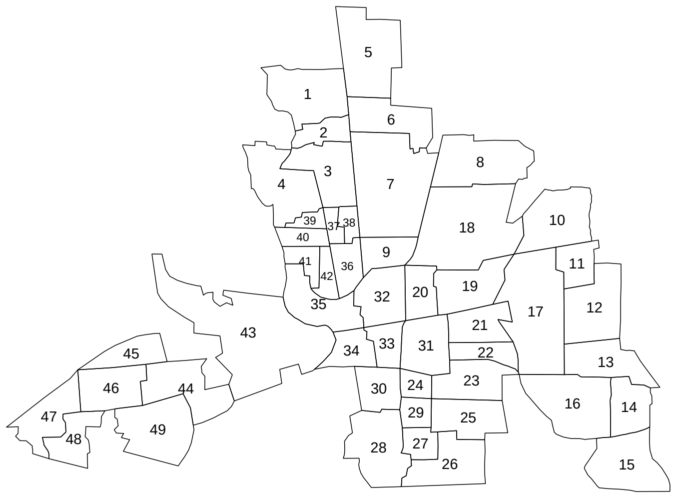


FIGURE 1.1. Lattice of spatial sites for Columbus dataset. Each neighborhood is a site. Each site is labeled. The collection of sites is the lattice of sites.

or how close site i is to site j based on some distance $d(i,j)$, e.g.

$$[W]_{i,j} = \begin{cases} d(i,j) & \text{if } i \sim j \text{ and } i \neq j \\ 0 & \text{if } i \nsim j \text{ or } i = j \end{cases}$$

Note 15. Proximity matrix W does not necessarily need to be symmetric, some times it is standardized as $[W]_{i,j} \leftarrow [W]_{i,j} / \sum_j [W]_{i,j}$.

Example 16. Consider the Columbus OH dataset which concerns spatially correlated count data arising from 49 districts/neighborhood in Columbus, OH in 1980. This is the R dataset `columbus{spdep}`. Figure 1.1 presents the sites and the lattice of sites. Each neighborhood is a site. Each site is label. The collection of sites is the lattice of sites coded with a unique labeled according to some order. One may define the "neighborhood relation $i \sim j$ considering counties that share common boarders (adjacent). Then for site $i = 43$, $i \sim j$ involves any $j \in \{44, 35, 34\}$ and for site $i = 20$, $i \sim j$ involves any $j \in \{32, 9, 18, 19, 31, 33\}$. Here $\mathcal{N}_{43} = \{44, 35, 34\}$ and $\mathcal{N}_{20} = \{32, 9, 18, 19, 31, 33\}$. The proximity matrix based on binary scheme will contain elements $W_{43,35} = 1$, $W_{43,43} = 0$, and $W_{43,33} = 0$.

Example 17. (Logistic/Ising model) Let variable Z_i denote the presence of a characteristic as $Z_i = 1$ or absence of it as $Z_i = 0$ on a site labeled by $i \in \mathcal{S}$. Then $\mathcal{Z} = \{0, 1\}$. The Ising model is defined by the (joint) PMF

$$\text{pr}_Z(z) \propto \exp\left(\alpha \sum_{i \in \mathcal{S}} z_i + \beta \sum_{\{i,j\}: i \sim j} z_i z_j\right), \ \forall z \in \mathcal{Z}^{\mathcal{S}} \tag{1.1}$$

E.g., it can model a black & white noisy image, where $\mathcal{S}$ denotes the labels of the image pixels, and Z_i denotes the presence of a black pixel ($Z_i = 1$) or its absence ($Z_i = 0$). Under Ising model (1.1), the characteristic is observed with probability $\mathrm{pr}_{Z_i}(z_i = 1) = \frac{\exp(\alpha)}{1+\exp(\alpha)}$ when $\beta = 0$. The characteristic's presence is encouraged in neighboring sites when $\beta > 0$, and discouraged when $\beta < 0$.

Notation 18. We use notation, for $\mathcal{A} \subset \mathcal{S}$

$$\mathrm{pr}_{\mathcal{A}}\left(z_{\mathcal{A}}|z_{\mathcal{S}\backslash\mathcal{A}}\right) = \mathrm{pr}\left(Z_{\mathcal{A}} = z_{\mathcal{A}}|Z_{\mathcal{S}\backslash\mathcal{A}} = z_{\mathcal{S}\backslash\mathcal{A}}\right)$$

Definition 19. Local characteristics of a random field Z on $\mathcal{S}$ with values in $\mathcal{Z}$ are the conditionals

$$\mathrm{pr}_i\left(z_i|z_{\mathcal{S}-i}\right) = \mathrm{pr}_{\{i\}}\left(z_{\{i\}}|z_{\mathcal{S}\backslash\{i\}}\right),\ i \in \mathcal{S},\ z \in \mathcal{Z}^{\mathcal{S}}$$

Example 20. (Cont. Example 17) The local characteristics of the Ising model in (1.1) are

$$\mathrm{pr}_i\left(z_i = 1|z_{\mathcal{S}-i}\right) = \frac{\exp\left(\alpha + \beta \sum_{\{i,j\}:i\sim j} z_j\right)}{1 + \exp\left(\alpha + \beta \sum_{\{i,j\}:i\sim j} z_j\right)}$$

2. Compatibility of conditional distributions

Note 21. Here, we discuss how to represent a joint probability distribution via its full conditionals. We need this for model building purposes.

Definition 22. Let random vector $Z = (Z_1, ..., Z_n)$ with joint distribution $\pi(Z_1, ..., Z_n)$. The set of distributions $\{\pi_i(\cdot|Z_{-i}); i = 1, ...n\}$ is called compatible to the joint distribution $\pi(Z_1, ..., Z_n)$ if the joint distribution $\pi(Z_1, ..., Z_n)$ has conditionals $\{\pi_i(Z_i|Z_{-i}); i = 1, .., n\}$.

Note 23. To specify suitable building models representing spatial dependency of a random field $(Z_i)_{i\in\mathcal{S}}$, it is often easier to visualize the joint distribution pr_z in terms of conditional distributions $\{\pi_i(Z_i|Z_{\mathcal{S}-i}); i \in \mathcal{S}\}$ rather than directly.

Note 24. Thus, instead of specifying a joint model for $(Z_i)_{i\in\mathcal{S}}$, a researcher may propose putative families of conditional distributions $\{\pi_i(Z_i|Z_{\mathcal{S}-i}); i \in \mathcal{S}\}$. However, an arbitrary chosen set of conditional distributions $\{\pi_i(\cdot|\cdot); i \in \mathcal{S}\}$ is not generally compatible, in the sense that there exists a joint distribution for $(Z_i)_{i\in\mathcal{S}}$, and hence we need to impose conditions.

Note 25. In what follows, we discuss necessary and sufficient conditions regarding compatibility.

Proposition 26. *(Compatibility condition) Let F be a joint distribution with $dF(x,y) = f(x,y)\,d(x,y)$ on $\mathcal{S}_x \times \mathcal{S}_y$. Let candidate condition distributions*

$$G \text{ with } dG(x|y) = g(x|y)\,dx, \text{ on } x \in \mathcal{S}_x$$
$$Q \text{ with } dQ(y|x) = q(y|x)\,dy, \text{ on } y \in \mathcal{S}_y$$

and let $N_g = \{(x,y) : g(x|y) > 0\}$ and $N_q = \{(x,y) : q(y|x) > 0\}$. A distribution F with conditionals exists iff

(1) $N_g = N_q = N$

(2) there exist functions u and $v$where $g(x|y)/q(y|x) = u(x)\,v(y)$ for all $(x,y) \in N$ and $\int u(x)\,dx < \infty$

Proof. Omitted[1]. □

Note 27. Essentially the above conditions guarantee that

$$k(y)\,g(x|y) = f(x,y) = h(x)\,q(y|x)$$

where k, g, h, q are densities.

Example 28. The conditionals $x|y \sim \mathrm{N}(a + by, \sigma^2 + \tau^2 y^2)$ and $y|x \sim \mathrm{N}(c + dx, \tilde{\sigma}^2 + \tilde{\tau}^2 x^2)$ are compatible if $\tau^2 = \tilde{\tau}^2 = 0$, $d/\tilde{\sigma}^2 = b/\sigma^2$, and $|db| < 1$.

Solution. See Exercise 29 in the Exercise sheet.

Note 29. Proposition 26 can be extended to more dimensions. For more info see (Arnold, B. C., & Press, S. J. (1989). in footnote 1)

Note 30. The following theorem shows that local characteristics can determine the entire distribution in certain cases.

Theorem 31. *(Besag's factorization theorem; Brook's Lemma) Let Z be a $\mathcal{Z}$ valued random field taking values in $\mathcal{Z}^{\mathcal{S}}$ where $\mathcal{S} = \{1, ..., n\}$ with $n \in \mathbb{N}$, and such as $pr_Z(z) > 0$, $\forall z \in \mathcal{Z}^{\mathcal{S}}$. Then for all*

$$\frac{pr_Z(z)}{pr_Z(z^*)} = \prod_{i=1}^{n} \frac{pr_i\left(z_i|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}{pr_i\left(z_i^*|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}, \quad \forall z, z^* \in \mathcal{Z}^{\mathcal{S}} \tag{2.1}$$

Proof. I will show that

$$\mathrm{pr}_Z(z) = \prod_{i=1}^{n} \frac{\mathrm{pr}_i\left(z_i|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}{\mathrm{pr}_i\left(z_i^*|z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)} \mathrm{pr}_Z(z^*)$$

[1] See Arnold, B. C., & Press, S. J. (1989). Compatible conditional distributions. Journal of the American Statistical Association, 84(405), 152-156.

It is

$$\mathrm{pr}_Z\left(z_1,...,z_n\right)=\frac{\mathrm{pr}_n\left(z_n|z_1,...,z_{n-2},z_{n-1}\right)}{\mathrm{pr}_n\left(z_n^*|z_1,...,z_{n-2},z_{n-1}\right)}\mathrm{pr}_Z\left(z_1,...,z_{n-1},z_n^*\right)$$

Let proposition P_j be

$$\mathrm{pr}_Z\left(z\right)=\prod_{i=n-j}^{n}\frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}{\mathrm{pr}_i\left(z_i^*|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}\mathrm{pr}_Z\left(z_1,...,z_{n-j-1},z_{n-j}^*,...,z_n^*\right)$$

Proposition P_0 is true

$$\mathrm{pr}_Z\left(z\right)=\frac{\mathrm{pr}_n\left(z_n|z_1,...,z_{n-1}\right)}{\mathrm{pr}_n\left(z_n^*|z_1,...,z_{n-1}\right)}\mathrm{pr}_Z\left(z_1,...,z_{n-1},z_n^*\right) \tag{2.2}$$

Proposition P_1 is true

$$\mathrm{pr}_Z\left(z_1,...,z_{n-1},z_n^*\right)=\frac{\mathrm{pr}_{n-1}\left(z_{n-1}|z_1,...,z_{n-2},z_n^*\right)}{\mathrm{pr}_{n-1}\left(z_{n-1}^*|z_1,...,z_{n-2},z_n^*\right)}\mathrm{pr}_Z\left(z_1,...,z_{n-2},z_{n-1}^*,z_n^*\right)$$

Assume that P_j is true. Then proposition $\mathrm{P}_{j+1}$is true as well, because

$$\begin{aligned}\mathrm{pr}_Z\left(z\right)&=\prod_{i=n-j}^{n}\frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}{\mathrm{pr}_i\left(z_i^*|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}\mathrm{pr}_Z\left(z_1,...,z_{n-j-1},z_{n-j}^*,...,z_n^*\right)\\&=\prod_{i=n-j}^{n}\frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}{\mathrm{pr}_i\left(z_i^*|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}\\&\quad\times\frac{\mathrm{pr}_{n-j-1}\left(z_{n-j-1}|z_1,...,z_{n-j-2},z_{n-j}^*,...,z_n^*\right)}{\mathrm{pr}_{n-j-1}\left(z_{n-j-1}^*|z_1,...,z_{n-j-2},z_{n-j}^*,...,z_n^*\right)}\mathrm{pr}_Z\left(z_1,...,z_{n-j-2},z_{n-j-1}^*,...,z_n^*\right)\\&=\prod_{i=n-(j+1)}^{n}\frac{\mathrm{pr}_i\left(z_i|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}{\mathrm{pr}_i\left(z_i^*|z_1,...,z_{i-1},z_{i+1}^*,...,z_n^*\right)}\mathrm{pr}_Z\left(z_1,...,z_{n-(j+1)-1},z_{n-(j+1)}^*,...,z_n^*\right)\end{aligned}$$

Then (2.1) is correct according to the induction principle. □

Note 32. Theorem 31 shows that the joint $\mathrm{pr}_Z\left(\cdot\right)$ can be constructed from its conditionals $\{\mathrm{pr}_i\left(\cdot|\cdot\right)\}$ if distributions $\{\mathrm{pr}_i\left(\cdot|\cdot\right)\}$ are compatible for $\mathrm{pr}_Z\left(\cdot\right)$, under the requirement that this construction is invariant wrt the coordinate permutation $\{1,...,n\}$ and the reference state z^*– these invariances correspond to the conditions in Proposition 26.