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, <u>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)</u>.

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

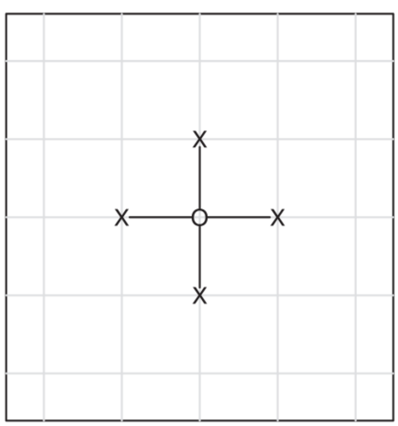
$$Z = (Z_s; s \in \mathcal{S})^\top = (Z_i = Z(s_i); i = 1, ...n)^\top.$$

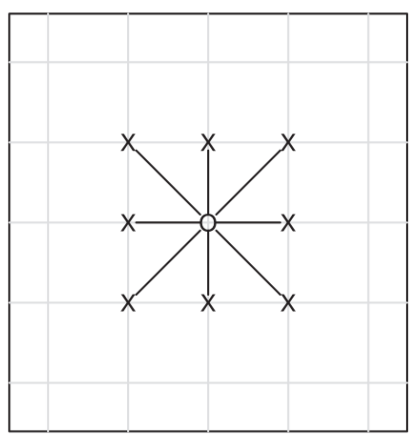
Note 10. Modeling aerial unit data often requires the specification of a neighborhood relation or neighborhood structure.

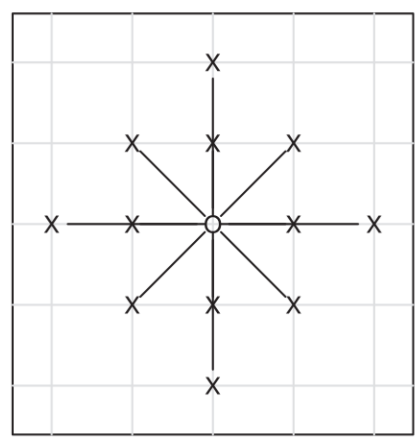
Notation 11. The notation $i \sim j$ between two sites $i, j \in \mathcal{S}$ means that "sites i and j are neighboring" according to a "neighborhood relation" $\sim$.

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.

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

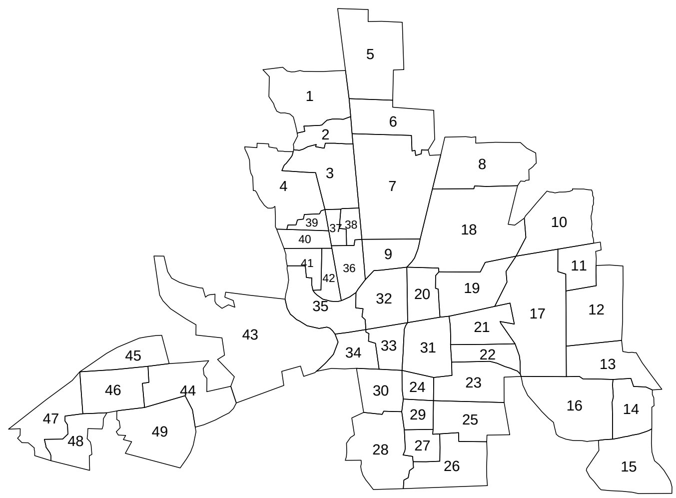


FIGURE 1.1. Lattice of spatial sites for Columbus dataset. Each neighborhood is a site. Each site is labeled. The collection of sites is the lattice of sites.

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

$$[W]_{i,j} = \begin{cases} 1/d(i,j) & \text{if } i \sim j \text{ and } i \neq j \\ 0 & \text{if } i \not\sim j \text{ or } i = j \end{cases}$$

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

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

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 $\text{pr}_{Z_i}(z_i = 1) = \frac{\exp(\alpha)}{1+\exp(\alpha)}$ 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 $\mathcal{A} \subset \mathcal{S}$

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

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

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

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

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

2. Lattice Random Fields (background)

Note 21. (Recall basic properties Fourier transform) Let $\{\beta_s : s \in \mathcal{S}\}$, $\mathcal{S} \subseteq \mathbb{Z}^d$ be a set of real coefficients

- The Fourier transform

$$\tilde{\beta}(\omega) = \sum_{s\in\mathcal{S}} \beta_s e^{is^\top\omega},\ \omega \in (-\pi, \pi]^d.$$

- The inverse Fourier transform is

$$\beta_s = \int_{(-\pi,\pi]^d} e^{-is^\top\omega}\tilde{\beta}(\omega)\,\mathrm{d}\omega$$

- Regularity conditions for the Fourier transform to be a well-defined function are
  1. If $\{\beta_s\}$ are summable, $\sum_{s\in\mathcal{S}}|\beta_s| < \infty$, then $\tilde{\beta}(\omega)$ is bounded and continuous function of ω.
  2. If $\{\beta_s\}$ are square summable, $\sum_{s\in\mathcal{S}}|\beta_s|^2 < \infty$ then $\tilde{\beta}(\omega)$ is square-integrable over $(-\pi, \pi]^d$ and (visa versa).

Note 22. Let $(Z_s : s \in \mathcal{S})$ be a (weakly) stationary random field where $\mathcal{S} \subseteq \mathbb{Z}^d$. The following is a tool (similar to Bochner's theorem) for specifying covariance function of stationary lattice random field. It results from Herglotz's theorem.

Proposition 23. *(Herglotz's theorem) Let $c : \mathbb{Z}^d \to \mathbb{R}$ be a real-valued function on integers for $d \geq 1$. Then $c(\cdot)$ is positive semidefinite (stationary covariance function) if and only if* *it can be represented as*

$$c(h) = \int_{(-\pi,\pi]^d} \exp\left(i\omega^\top h\right) dF(\omega)$$

where F is a *symmetric positive bounded finite measure on* $(-\pi,\pi]^d$ *and* $F(-\pi) = 0$. F *is called spectral measure of* $c(h)$. f *is called spectral density of* $c(h)$ *if*

$$dF(\omega) = f(\omega)\, d\omega$$

Proposition 24. *If* $c(\cdot)$ *is integrable, the spectral density* $f(\cdot)$ *can be computed by inverse Fourier transformation*

$$f(\omega) = \left(\frac{1}{2\pi}\right)^d \sum_h \exp\left(-i\omega^\top h\right) c(h)$$

Note 25. Let $\left(Y(s) : s \in \mathbb{R}^d\right)$ be a stationary random field with spectral measure $F_Y(\omega)$, $\omega \in \mathbb{R}^d$ and let $\left(Z_s : s \in \mathbb{Z}^d\right)$ with $Z_s = Y(s)$ for $s \in \mathbb{Z}^d$, then Z_s has spectral measure

$$F_Z(\omega) = \sum_{k\in\mathbb{Z}^d} F_Y(2\pi k + d\omega),\ \omega \in (-\pi,\pi]^d$$

where frequencies separated by a lag $2\pi k$, $k \in \mathbb{Z}^d$, are aliased together in the construction of$F_Z(\omega)$. Hence, there are infinitely many ways to interpolate a stationary process on $\mathbb{Z}^d$ to give a stationary random field $\mathbb{R}^d$.

Note 26. Let $(U_s : s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$, be a stationary random field covariance function

$$c_U(h) = \int_{(-\pi,\pi]^d} e^{i\omega^\top h} f(\omega)\, \mathrm{d}\omega$$

and spectral density $f(\omega)$ over $(-\pi,\pi]^d$. Let $(V_s : s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$, be a random field such as

$$V_s = \sum_{h\in\mathbb{Z}^d} U_{s+h}\beta_h,\ s \in \mathcal{S} \subseteq \mathbb{Z}^d$$

where $\left\{\beta_h : h \in \mathbb{Z}^d\right\}$ are summable functions, i.e. $\sum_h |\beta_h| < \infty$. The covariance function of $(V_s : s \in \mathcal{S})$ is

$$c_V(h) = \mathrm{Cov}(V_s, V_{s+h}) = \sum_{t\in\mathbb{Z}^d} \sum_{t'\in\mathbb{Z}^d} \beta_t \beta_{t'} c_U(h + t - t')$$

with spectral measure

$$\mathrm{d}F_V(\omega) = \left|\tilde{\beta}(\omega)\right|^2 \mathrm{d}F_U(\omega),\ \omega \in (-\pi,\pi]^d \tag{2.1}$$

where

$$\tilde{\beta}(\omega) = \sum_{h\in\mathbb{Z}^d} \beta_h e^{ih^\top \omega},\ \omega \in (-\pi,\pi]^d \tag{2.2}$$

is the Fourier transform of β_h.

Proof. This is straightforward from Proposition 23 (Herglotz's theorem) $\square$

3. Compatibility of conditional distributions

Note 27. Here, we discuss how to represent a joint probability distribution via its full conditionals. We need this for model building purposes.

Definition 28. 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 29. 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 30. 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 31. In what follows, we discuss necessary and sufficient conditions regarding compatibility.

Proposition 32. *(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]. $\square$

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

Note 33. 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 34. The conditionals $x|y \sim \mathrm{N}\left(a + by, \sigma^2 + \tau^2 y^2\right)$ and $y|x \sim \mathrm{N}\left(c + dx, \tilde{\sigma}^2 + \tilde{\tau}^2 x^2\right)$ 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 35. Proposition 32 can be extended to more dimensions. For more info see (Arnold, B. C., & Press, S. J. (1989). in footnote 1)

Note 36. The following theorem shows that local characteristics can determine the entire distribution in certain cases.

Theorem 37. *(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*

$$\frac{pr_Z(z)}{pr_Z(z^*)} = \prod_{i=1}^{n} \frac{pr_i\left(z_i | z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}{pr_i\left(z_i^* | z_1, ..., z_{i-1}, z_{i+1}^*, ..., z_n^*\right)}, \quad \forall z, z^* \in \mathcal{Z}^{\mathcal{S}} \tag{3.1}$$

Proof. I will show that

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

It is

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

Let proposition P_j be

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

Proposition P_0 is true

$$\mathrm{pr}_Z(z) = \frac{\mathrm{pr}_n(z_n | z_1, ..., z_{n-1})}{\mathrm{pr}_n(z_n^* | z_1, ..., z_{n-1})} \mathrm{pr}_Z(z_1, ..., z_{n-1}, z_n^*) \tag{3.2}$$

Proposition P_1 is true

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

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

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

Then (3.1) is correct according to the induction principle. □

Note 38. Theorem 37 shows that the joint $\mathrm{pr}_Z(\cdot)$ can be constructed from its conditionals $\{\mathrm{pr}_i(\cdot|\cdot)\}$ if distributions $\{\mathrm{pr}_i(\cdot|\cdot)\}$ are compatible for $\mathrm{pr}_Z(\cdot)$, 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 32.

4. Spatial autoregressive models

We present two basic spatial Autoregressive models, the SAR and CAR, able to represent spatial dependency.

4.1. **Simultaneous AutoRegressive (SAR) models.**

Definition 39. "Autoregressive" representation for lattice random field $(Z_s; s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$ is called

$$\sum_{h \in \mathcal{H} \cup \{0\}} a_h \left(Z_{s+h} - \mu_{s+h}\right) = \varepsilon_s,\ \ s \in \mathcal{S} \tag{4.1}$$

where $\{a_h; h \in \mathcal{H}\}$, $\mathcal{H} = \left\{h \in \mathbb{Z}^d - \{0\} : s + h \in \mathcal{S}\right\}$, are coefficients and $\{\varepsilon_s; s \in \mathcal{S}\}$ is a discrete white noise random field with variance $\lambda_s = \mathrm{Var}(\varepsilon_s)$.

4.1.1. *Assuming stationarity.*

Note 40. We will focus our study on stationary CAR random field $(Z_s; s \in \mathcal{S})$, i.e. $\mu_s = \mu$ and $\lambda_s = \sigma^2_\varepsilon$ for $s \in \mathcal{S} \subseteq \mathbb{Z}^d$.

Note 41. $\{\varepsilon_s; s \in \mathcal{S}\}$ has c.f. $c_\varepsilon(h) = \sigma_\varepsilon^2 \delta_{\{0\}}(h)$ and hence spectral density $f_\varepsilon(\omega) = \sigma_\varepsilon^2 / (2\pi)^d$. The spectral density f for $(Z_s; s \in \mathcal{S})$ is

$$f(\omega) = \frac{1}{|\tilde{a}(\omega)|^2} f_\varepsilon(\omega) = \frac{1}{|\tilde{a}(\omega)|^2} \left(\frac{1}{2\pi}\right)^d \sum_h e^{-i\omega^\top h} c_\varepsilon(h)\,\mathrm{d}h$$
$$= \frac{1}{|\tilde{a}(\omega)|^2} \left(\frac{1}{2\pi}\right)^d \sum_{h \in \mathbb{Z}^d} e^{-i\omega^\top h} \sigma_\varepsilon^2 \delta_{\{0\}}(h)\,\mathrm{d}h = \frac{1}{|\tilde{a}(\omega)|^2} \frac{\sigma_\varepsilon^2}{(2\pi)^d} \tag{4.2}$$

where $\tilde{a}(\omega) = \sum_h a_h e^{ih^\top \omega}$ since (26).

Note 42. For random field $(Z_s; s \in \mathcal{S})$ to be stationary, $f(\omega)$ in (4.2) must be integrable function and bounded function on $\omega \in (-\pi, \pi]^d$. Hence, we can set restrictions on coefficients

$$a_0 > 0, \text{ and } \sum_h |a_h| < a_0 \tag{4.3}$$

satisfying regularity conditions in (21).

Note 43. To make model (4.1) identifiable and use it for inference, we can introduce further restrictions $a_h = a_{-h}$.

Definition 44. Lattice random field $(Z_s; s \in \mathcal{S})$ is called Simultaneous AutoRegressive (SAR) if it is given in an "Autoregression" representation (4.1)

$$\sum_{h \in \mathcal{N}} a_h Z_{s+h} = \varepsilon_s,\ s \in \mathcal{S}$$

whose coefficients $\{a_h; h \in \mathcal{H}\}$ satisfy the symmetry condition

$$a_h = a_{-h},\ \forall h$$

and $f_Z(\omega)$ in (4.2) is an integrable function over $\omega \in (-\pi, \pi]^d$.

4.1.2. *Assuming Gaussian distribution.*

Note 45. Following we provide the matrix form of the definition used in software implementations.

Definition 46. Consider discrete set of sites $\mathcal{S} = \{s_i; i = 1, ..., n\}$ and a lattice random field $(Z_s; s \in \mathcal{S})$. Vectorize $Z = (Z_1, ..., Z_n)^\top$ with $Z_i = Z(s_i)$ and set

$$Z = \mu + A(Z - \mu) + E \Longleftrightarrow E = (I - A)(Z - \mu)$$

Assume that matrix A is such that $[A]_{i,i} = 0$ and $(I - A)^{-1}$ exists. Assume n-dimensional Gaussian random vector $E \sim \mathrm{N}_n(0, \Lambda)$ with $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$. We say that Z follows a "Gaussian" Simultaneous Autoregressive model.

Note 47. The joint distribution of Z following the SAR model in Definition 46 is

$$Z \sim \mathrm{N}\left(\mu, (I-A)^{-1} \Lambda \left(I - A^{\top}\right)^{-1}\right) \tag{4.4}$$

Proof. Z is a linear combination of Gaussian random vectors, hence if follows a Gaussian distribution. It's mean and variance are

$$\mathrm{E}(Z) = \mathrm{E}\left((I-A)^{-1} E + \mu\right) = \mu,$$

$$\mathrm{Var}(Z) = \mathrm{Var}\left((I-A)^{-1} E + \mu\right) = (I-A)^{-1} \mathrm{Var}(E) \left(I - A^{\top}\right)^{-1} = (I-A)^{-1} \Lambda \left(I - A^{\top}\right)^{-1}$$

□

4.2. Conditional autoregressive models (CAR).

Definition 48. A lattice random field $(Z_s; s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$ is called Conditional AutoRegressive (CAR) model if

$$\mathrm{E}(Z_s | Z_{S-s}) = \mu_s + \sum_{h \in \mathcal{H}} b_h (Z_{s+h} - \mu_{s+h})$$

$$\mathrm{Var}(Z_s | Z_{S-s}) = \kappa_s$$

where $b_h = -b_h$ and $\mathcal{H} = \left\{h \in \mathbb{Z}^d - \{0\} : s + h \in \mathcal{S}\right\}$.

Note 49. Alternatively CAR lattice random field $(Z_s; s \in \mathcal{S})$, $\mathcal{S} \subseteq \mathbb{Z}^d$can be written as

$$Z_s = \mu_s + \sum_{h \in \mathcal{H}} b_h (Z_{s+h} - \mu_{s+h}) + \epsilon_s \tag{4.5}$$

where $(\epsilon_s : s \in \mathcal{S})$ is the residual random field with mean $\mathrm{E}(\epsilon_s) = 0$ and variance $\mathrm{Var}(\epsilon_s) = \kappa_s$. Also $b_h = -b_h$ and $\mathcal{H} = \left\{h \in \mathbb{Z}^d - \{0\} : s + h \in \mathcal{S}\right\}$.

4.2.1. *Assuming stationary.*

Note 50. We will focus our study on stationary CAR random field $(Z_s; s \in \mathcal{S})$, and hence we set $\mu_s = \mu$ and $\kappa_s = \kappa$ for $s \in \mathcal{S} \subseteq \mathbb{Z}^d$.

Note 51. It is

$$\begin{aligned}
\mathrm{E}(Z_s \epsilon_s) &= \mathrm{E}\left(\left(\epsilon_s - \mu - \sum_{h \in \mathcal{H}} b_h (Z_{s+h} - \mu)\right) \epsilon_s\right) \\
&= \mathrm{E}\left(\epsilon_s^2\right) - \mu \mathrm{E}(\epsilon_s) - \sum_{h \in \mathcal{H}} b_h \mathrm{E}((Z_{s+h} - \mu) \epsilon_s) \\
&= \mathrm{E}\left(\epsilon_s^2\right) + 0 + 0 = \kappa
\end{aligned}$$

Note 52. The covariance function $c(\cdot)$ of stationary CAR random field $(Z_s; s \in \mathcal{S})$ is

$$c(h) = \sum_{h' \in \mathcal{H}} b_{h'} c(h - h') + \kappa \delta_{\{0\}}(h)$$

as, for $h \neq 0$ it is

$$\begin{aligned}
c(h) = \mathrm{E}\left((Z_s-\mu)(Z_{s+h}-\mu)\right) &= \mathrm{E}\left(\left(\sum_{h'\in\mathcal{H}} b_{h'}(Z_{s+h'}-\mu)+\epsilon_s\right)(Z_{s+h}-\mu)\right)\\
&= \mathrm{E}\left(\left(\sum_{h'\in\mathcal{H}} b_{h'}(Z_{s+h'}-\mu)+\epsilon_s\right)(Z_{s+h}-\mu)\right)\\
&= \sum_{h'\in\mathcal{H}} b_{h'}\mathrm{E}(Z_{s+h'}Z_{s+h}) - \mu^2\sum_{h'\in\mathcal{H}} b_{h'} + \mathrm{E}(\epsilon_s Z_{s+h})\\
&= \sum_{h'\in\mathcal{H}} b_{h'}b_{h'}c(h-h') - 0 + 0
\end{aligned}$$

and for $h=0$

$$\begin{aligned}
c(0) = \mathrm{E}\left((Z_s-\mu)(Z_s-\mu)\right) &= \\
&= \mathrm{E}\left(\left(\sum_{h\in\mathcal{H}} b_h(Z_{s+h}-\mu)+\epsilon_s\right)\left(\sum_{h'\in\mathcal{H}} b_{h'}(Z_{s+h'}-\mu)+\epsilon_s\right)\right)\\
&= \sum_{h\in\mathcal{H}}\sum_{h'\in\mathcal{H}} b_{h'}b_{h'}c(h-h') + \kappa
\end{aligned}$$

Note 53. The spectral density of the covariance function $c(h)$ of stationary CAR random field $(Z_s; s\in\mathcal{S})$ is computed by inverse Fourier transform as

$$\begin{aligned}
f(\omega) &= \left(\frac{1}{2\pi}\right)^d \sum_{\mathrm{h}} e^{-i\omega^\top h} c(h)\\
&= \left(\frac{1}{2\pi}\right)^d \sum_{\mathrm{h}} e^{-i\omega^\top h}\left(\sum_{h'\in\mathcal{H}} b_{h'}c(h-h') + \kappa\delta_{\{0\}}(h)\right)\\
&= \tilde{b}(\omega) f(\omega) + \kappa\frac{1}{(2\pi)^d} \implies
\end{aligned}$$

$$f(\omega) = \frac{\kappa}{(2\pi)^d}\frac{1}{1-\tilde{b}(\omega)},\ \tilde{b}(\omega) = \sum_{h\in\mathcal{H}} b_h e^{ih^\top\omega} = \sum_{h\in\mathcal{H}} b_h\cos\left(\omega^\top h\right) \tag{4.6}$$

Hence sufficient conditions for CAR random field $(Z_s; s\in\mathcal{S})$ in (4.5) to be stationary is the spectral density (4.6) to be bounded. This is true if $\tilde{b}(\omega) < 1$ which is implied by

$$\sum_{h\in\mathcal{H}} |b_h| < \infty$$

satisfying regularity conditions in (21).

4.2.2. *Assuming Gaussian distribution.*

Note 54. Following we provide the matrix form of the definition used in software implementations.

Definition 55. "Gaussian" Conditional autoregressive model, CAR, assumes that the local characteristics $\{\mathrm{pr}_i (z_i|z_{\mathcal{S}-i})\}$ are Gaussian distributions

$$Z_i|z_{\mathcal{S}-i} \sim \mathrm{N}\left(\underbrace{\mu_i + \sum_{j\neq i} b_{i,j} (Z_j - \mu_j)}_{=\mathrm{E}(Z_i|Z_{\mathcal{S}-i})}, \underbrace{\kappa_i}_{=\mathrm{Var}(Z_i|Z_{\mathcal{S}-i})}\right), \quad \forall i \in \mathcal{S} \tag{4.7}$$

Proposition 56. *Let $K = diag(\{\kappa_i\})$ with $\kappa_i > 0$, matrix B with $B_{i,i} := [B]_{i,i} = 0$, and real vector μ with suitable dimensions. If Z follows a Gaussian CAR (Definition 55), $I - B$ is non-singular, and $(I - B)^{-1} K > 0$, then the joint distribution of Z is*

$$Z \sim N\big(\mu, (I - B)^{-1} K\big) . \tag{4.8}$$

Proof. Without lose of generality, consider zero mean $\mu = 0$ (or equivalently set $Z := Z - \mu$). The full conditionals $Z_i|z_{\mathcal{S}-i}$ in (4.7) are compatible with the joint distribution $\mathrm{pr}_Z (z)$. By using Besag's factorization theorem (Theorem 37) with reference state/configuration $z^* = 0$ we get

$$\begin{aligned}
\mathrm{pr}_Z (z) &= \prod_{i=1}^{n} \frac{\mathrm{pr}_i \left(z_i|z_1, ..., z_{i-1}, z_{i+1}^* = 0, ..., z_n^* = 0\right)}{\mathrm{pr}_i \left(z_i^* = 0|z_1, ..., z_{i-1}, z_{i+1}^* = 0, ..., z_n^* = 0\right)} \mathrm{pr}_Z (z^* = 0) \\
&= \prod_{i=1}^{n} \frac{\mathrm{N}\left(z_i| \sum_{j<i} b_{i,j} z_j + 0, \kappa_i\right)}{\mathrm{N}\left(0| \sum_{j<i} b_{i,j} z_j + 0, \kappa_i\right)} \mathrm{pr}_Z (z^* = 0) \\
&\propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\kappa_i}\left(z_i - \sum_{j<i} b_{i,j} z_j\right)^2 + \frac{1}{2\kappa_i}\left(0 - \sum_{j<i} b_{i,j} z_j\right)^2\right) \\
&= \prod_{i=1}^{n} \exp\left(-\frac{1}{2\kappa_i}\left(z_i^2 - 2z_i \sum_{j<i} b_{i,j} z_j\right)\right) \mathrm{pr}_Z (z^* = 0) \\
&= \exp\left(-\sum_{i} \frac{z_i^2}{2\kappa_i} + \frac{1}{2} 2 \sum_{i} \sum_{j<i} \frac{b_{i,j}}{\kappa_i} z_i z_j\right) \mathrm{pr}_Z (z^* = 0) \\
&= \exp\left(-\frac{1}{2} z^\top K^{-1} z + \frac{1}{2} z^\top K^{-1} B z\right) \mathrm{pr}_Z (z^* = 0) = \exp\left(-\frac{1}{2} z^\top \left[K^{-1} (I - B)\right] z\right) \mathrm{pr}_Z (z^* = 0) \\
&= \mathrm{N}\left(z|0, (I - B)^{-1} K\right)
\end{aligned} \tag{4.9}$$

Recovering the mean from (4.9), it is

$$\mathrm{pr}_Z(z) = \mathrm{N}\left(z-\mu|0, (I-B)^{-1}K\right) = \mathrm{N}\left(z|\mu, (I-B)^{-1}K\right)$$

□

Note 57. When CAR is used for modeling, B is often specified to be sparse either due to some natural problem specific property, or for our computational convenience as is may allow the use of sparse solvers. To achieve this, one way is to specify $B = \phi W$ where $\phi > 0$ and W is an adjacency/proximity matrix; that is $[B]_{i,j} = \phi 1(i \sim j) 1(i \neq j)$ will be non-zero only for adjacent pairs i and j.

Note 58. The system in (4.8) can be rewritten as

$$Z = \mu + B(Z-\mu) + E \Longleftrightarrow E = (I-B)(Z-\mu) \tag{4.10}$$

by setting $E = (I-B)(Z-\mu)$. The distribution of Z in (4.8) induces a distribution on E as $E \sim \mathrm{N}\left(0, K(I-B)^{\top}\right)$ because

$$\mathrm{E}(E) = \mathrm{E}((I-B)(Z-\mu)) = (I-B)\,\mathrm{E}(Z-\mu) = 0$$
$$\mathrm{Var}(E) = \mathrm{Var}((I-B)Z) = (I-B)\,\mathrm{Var}(Z)\,(I-B)^{\top} = (I-B)(I-B)^{-1}K(I-B)^{\top}$$

Part 2. Model building for aerial data & related inference