

STK4150/9150 - Environmental and Spatial Statistics

Q & A

Thordis Thorarinsdottir

Spring 2026

Dimensions

For continuous spatial random fields, when we say we only look at two-dimensional fields, how does the elevation fit into models of, for example, temperature?

When we say we only look at two-dimensional fields, we mean that we consider fields with a two dimensional index $X(\mathbf{s})$ where $\mathbf{s} \in \mathbb{R}^2$. The field usually takes any real value, $X(\mathbf{s}) \in \mathbb{R}$.

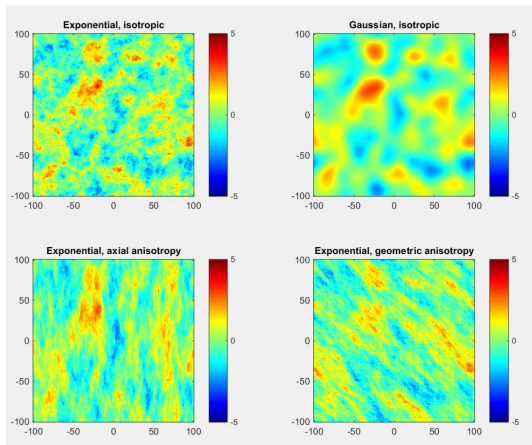
Then, we can have that, for example, the mean function depends on many fixed covariates such as elevation, distance from ocean, pressure, etc.:

$$\mu(\mathbf{s}) = \beta_0 + \beta_1 x_1(\mathbf{s}) + \beta_2 x_2(\mathbf{s}), \dots, \beta_p x_p(\mathbf{s}).$$

We don't consider, for example, the joint distribution of temperature and pressure, or the three dimensional distribution of temperature (with the third dimension being altitude), that is, where all of these are random.

Inferring model type from a plot

Can we go through isotropic/stationary examples on plots.



Stationary covariance:

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C_Y(\mathbf{h})$$

Isotropic covariance:

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C_Y(\|\mathbf{h}\|)$$

Geometric anisotropy in covariance:

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C_Y(\|\mathbf{A}\mathbf{h}\|)$$

Spectral representation

What did we use spectral representation for?

We only use it to characterise which functions can be covariance functions of stationary processes:

Theorem [Bochner, 1955] If $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |C_Y(\mathbf{h})| d\mathbf{h} < \infty$, then a valid real-valued covariance function can be written as

$$C_Y(\mathbf{h}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \cos(\boldsymbol{\omega}^T \mathbf{h}) f_Y(\boldsymbol{\omega}) d\boldsymbol{\omega}$$

where the *spectral density* of $C_Y(\mathbf{h})$, $f_Y(\boldsymbol{\omega}) \geq 0$ is symmetric about $\boldsymbol{\omega} = 0$.

More generally, it is sometimes simpler to work with spatial modelling in the frequency domain rather than the location domain. However, the discussion of this is outside of the scope of the class.

Plug-in rule for estimation

Plug in rule for estimation at new locations (exercise week 7), is this something we need when we use kriging?

This is one (simple) way to deal with unknown parameters in kriging (see also V26_03_ch4_geostat.pdf, pages 49++).

Type of continuous models

Can we focus on only the case where $Z = Y + \epsilon$ and $X = X\beta + \delta$ or are there other continuous models that are important?

It is good to focus here, and this seems like a good choice of models to focus on.

Estimation

When you ask about estimation is that theoretically or can it just be “we use this package in R”?

It is largely meant conceptually: is it straight forward, such as easy to use maximum likelihood, or is it hard? If it is hard, why and how solved in general terms.

Hammersley-Clifford

How can we mention Besag's Lemma and Hammersley-Clifford theorem without having to go into detail because there is not a lot of time?

The lemma is just a lemma needed for the proof of the theorem.

Hammersley-Clifford: For strictly positive distributions it holds that

Markov property $\Leftrightarrow$ Gibbs factorisation over cliques

INLA

When we talk about estimation in spatial lattice processes, is that mostly explaining INLA? And can we go through INLA quickly?

For models with latent GRFs and GMRFs, INLA is an example of a common inference approach. INLA is an approximative Bayesian inference approach that replaces MCMC with fast Laplace approximations to posterior marginals. (And be able to explain briefly what that means if asked)

CAR models

Can we go through the Besag-York-Mollie CAR model and how it is connected to the definition of a CAR model in the slides?

Define $\mathcal{N}(\mathbf{s}_i) = \{j; W_{ij} = 1\}$, the neighbor set of site i with $\mathbf{s}_i \in D_S$.

A conditional autoregressive (CAR) process is a process with conditional Gaussian distributions of the form

$$\mathbb{E}[Y(s_i) | \mathbf{Y}_{-i}] = \sum_{j \in \mathcal{N}(\mathbf{s}_i)} c_{ij} Y(\mathbf{s}_j)$$

$$\text{Var}[Y(s_i) | \mathbf{Y}_{-i}] = \sigma_i^2$$

$$Z_k | \lambda_k \sim \text{Poisson}(\lambda_k)$$

$$\ln(\lambda_k) = \mathbf{x}_k^\top \boldsymbol{\beta} + O_k + \psi_k$$

$$\boldsymbol{\beta} \sim \text{Gau}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$$

$$\psi_k = \varphi_k + \theta_k$$

$$\varphi_k | \boldsymbol{\varphi}_{-k}, \mathbf{W}, \tau^2 \sim \text{Gau}\left(\frac{\sum_{i=1}^K w_{ki} \varphi_i}{\sum_{i=1}^K w_{ki}}, \frac{\tau^2}{\sum_{i=1}^K w_{ki}}\right)$$

$$\theta_k \sim \text{Gau}(0, \sigma^2)$$

$$\tau^2, \sigma^2 \sim \text{Inverse-Gamma}(a, b)$$

de Wij

Should we mention the de Wij process?

It is historically important, but mainly of theoretical interest today.

Second-order intensity

Could you explain the interpretation of the second-order intensity function from point processes with an example?

First-order intensity $\lambda(x)$: how many points do we expect near location x per unit volume?

Second-order intensity $\rho^{(2)}(x, y)$: how many pairs of points do we expect simultaneously near x and y per unit area in (x, y) -space?

- If no interaction (Poisson): $\rho^{(2)}(x, y) = \lambda(x)\lambda(y)$.
- If there is clustering between x and y : $\rho^{(2)}(x, y) > \lambda(x)\lambda(y)$
- If there is repulsion between x and y : $\rho^{(2)}(x, y) < \lambda(x)\lambda(y)$

J-function

Same for the J-function? Do we assume the process to be homogenous to be able to use the J-function as a summary statistic?

For a stationary point process in $\mathbb{R}^2$:

- $F(r) = \mathbb{P}(\text{Distance from an arbitrary location to the nearest point} \leq r)$
- $G(r) = \mathbb{P}(\text{Distance from a typical point to its nearest neighbor} \leq r)$

The *J*-function is

$$J(r) = \frac{1 - G(r)}{1 - F(r)}, \quad r \geq 0.$$

For a homogeneous Poisson process: $F(r) = G(r) \Rightarrow J(r) \equiv 1$ for all r

J -function

- $F(r) = \mathbb{P}(\text{Distance from an arbitrary location to the nearest point} \leq r)$
- $G(r) = \mathbb{P}(\text{Distance from a typical point to its nearest neighbor} \leq r)$
- $J(r) = \frac{1-G(r)}{1-F(r)}, \quad r \geq 0$

Trees in a forest:

- Forest with seedlings around parent trees: Typical tree has close neighbors
 - $G(r)$ rises faster than under Poisson for small r , while $F(r)$ changes less $\Rightarrow$
 $1 - G(r)$ is small and $J(r) < 1$
- Plantation with trees planted on a grid: Typical tree has no close neighbors
 - $G(r)$ rises slower than under Poisson for small r , while $F(r)$ changes less $\Rightarrow$
 $1 - G(r)$ is larger and $J(r) > 1$

J -function is primarily used under homogeneity, similar to $K/F/G$, while inhomogeneous versions exist.

Janossy density

Is Janossy density something worth mentioning, if so can we go through it better?

The *Jánossy density* is defined in terms of $(p_n, \pi_n)_{n \in \mathbb{N}}$ as

$$j_n(\mathbf{x}_1, \dots, \mathbf{x}_n) = n! p_n \pi_n(\mathbf{x}_1, \dots, \mathbf{x}_n), \quad n \in \mathbb{N}_0.$$

where p_n is the probability that there are exactly n points, and $\pi_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is the joint probability density function of the n points, given that there are exactly n points in the region.

That is, $j_n(\mathbf{x}_1, \dots, \mathbf{x}_n) d\nu(\mathbf{x}_1) \cdots d\nu(\mathbf{x}_n)$ is the probability of finding exactly n points, one at each of infinitesimal regions centered at $\mathbf{x}_1, \dots, \mathbf{x}_n$.

→ not worth mentioning, since it is never used in actual modelling.

POT

Is the Poisson process in the POT method a stationary process?

In classic POT, the underlying data generating process is assumed to be stationary. This yields a homogeneous Poisson point process of exceedances, which is stationary as a point process.

Block maxima

The definition of block maxima

(See V26_extremes_univ.pdf)

Writing on the board

"Do you have tips for writing down formulas on the blackboard during the presentation? Do we just say "x is defined by ..." and write that down? Do we name every component? Do we talk about it generally while writing down?"

Notation

"When defining models, when do you use square brackets, and when normal ones?"

A senior academic to a PhD student: "Alex, an important part of your early career will be to establish your own notation for concepts that is completely defined by you. But if you ever refer to the cone of positive definite matrices according to a Gaussian Graphical Model please make sure you use my $P(G)$ and not that awful $M(G)$ of West and Carvalho."