Geostatistical Methods

Moreno Bevilacqua
Universidad Adolfo Ibañez
moreno.bevilacqua@uai.cl

2021

Contents

A. , 3

Books

- P. Abrahamsen *A review of Gaussian Random Fields and correlation functions.* Norwegian Computing Center, 1997. (you can find it on internet).
- S. Banerjee, B.P. Carlin and A.F. Gelfand *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall/CRC, Boca Raton, 2015 (2nd Edition).
- J.-P. Chilès and P. Delfiner *Geostatistics.* Wiley, New York, 2012.
- N. Cressie *Statistics for Spatial Data (revised edition).* Wiley, New York, 1993.
- A.F. Gelfand, P.J. Diggle, M. Fuentes and P. Guttorp *Handbook of Spatial Statistics.* Chapman & Hall/CRC, Boca Raton, 2010
- P.J. Diggle, P.J. Ribeiro *Model based geostatistics.* Springer, New York, 2007.
- C. Gaetan, X. Guyon *Spatial Statistics and Modeling.* Springer, New York, 2010.
- O. Schabenberger, C.A. Gotway *Statistical Methods for Spatial Data Analysis.* Chapman and Hall, 2005.
- M. Stein *Interpolation of Spatial Data: Some Theory for Kriging.* Wiley, New York, 1999.

Unit A

Types of geostatistical data

In spatial problems, observations come from a multivariate spatio temporal stochastic process (**random field**)

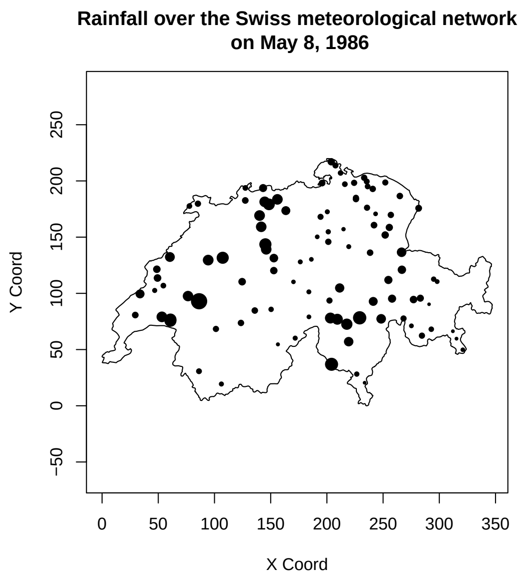
$$X = \{X_p(s,t), \qquad (s,t) \in S \times T\} \qquad p = 1, \ldots, m.$$

The positions of observation sites $s \in S$, $t \in T$ are either fixed in advance. Classically, S is a 2-dimensional subset, $S \subseteq \mathbb{R}^2$ and $T \subseteq \mathbb{R}^1$ while p is an integer denoting the number of variable considered.

Type of data:

- Spatial data: one realization from a spatial random field (t is fixed, $p = 1$).
- Spatio-temporal data: one realization from a space time random field ($p = 1$)
- Multivariate ($p > 1$) spatial or spatio-temporal data: (one) realizations from a multivariate space (time) random field.

1. Spatial Point referenced data

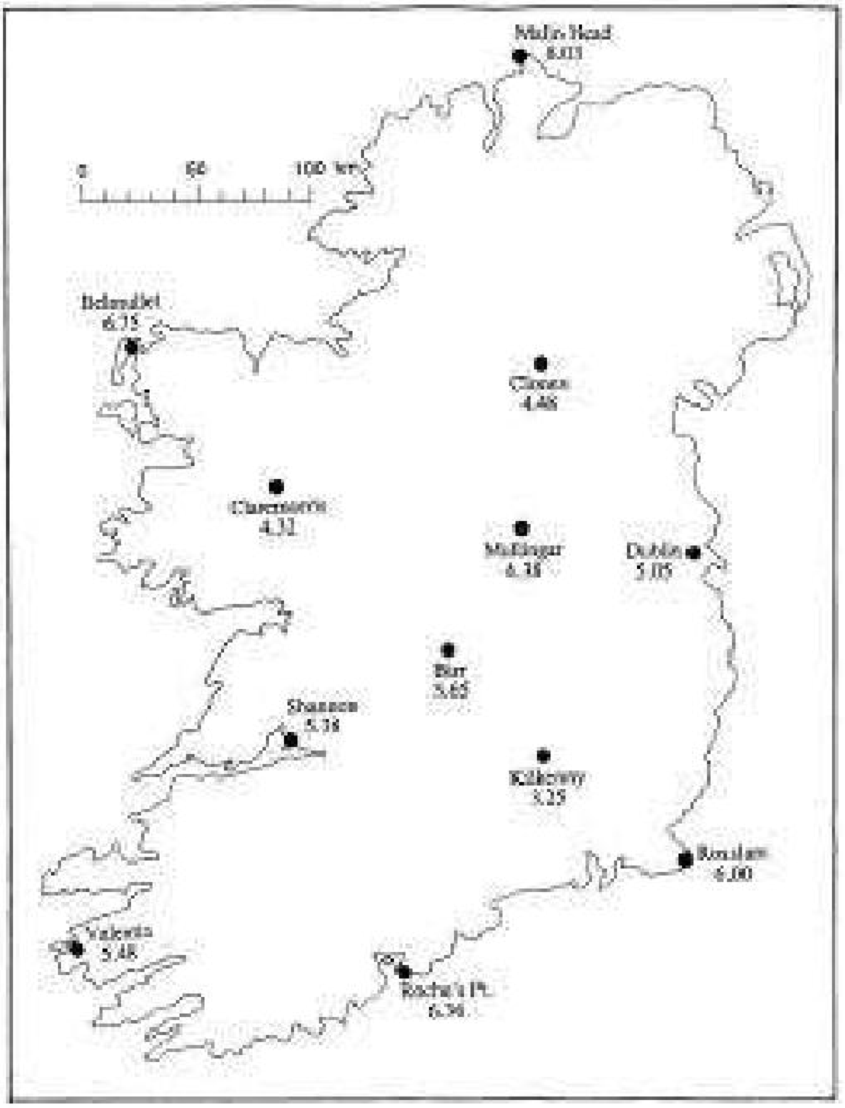


Here, S is a *continuous* subspace of $\mathbb{R}^d$ and the random field $X = \{X(s),\ s \in S\}$ is observed at n fixed sites $\{s_1, \ldots, s_n\}$.

Goals:

- **modeling**, jointly modeling of spatial trend and dependence
- **estimation**, efficient methods of estimation.
- **prediction** (or kriging) at unobserved sites and reconstruction of X across the whole space S.

2. Spatio temporal Point referenced data



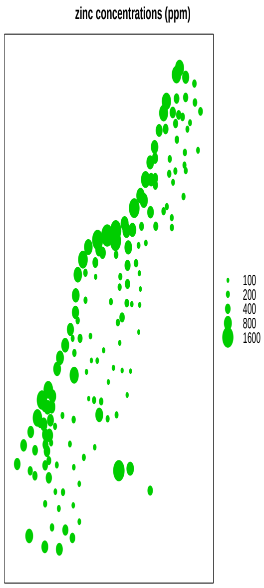
Daily average wind speed in m/s in Ireland from 11 meteorological stations (1961-1978) (Haslett and Raftery , 1989, 72270 observations). Here, typically $S \subseteq \mathbb{R}^d$ and $T \subseteq \mathbb{R}$.

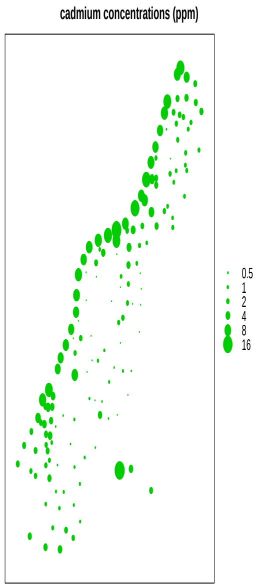
The random field $X = \{X(s,t), \, (s,t) \in S \times T\}$ is observed at n_s fixed sites $\{s_1, \ldots, s_{n_s}\}$ and n_t temporal instants $\{t_1, \ldots, t_{n_t}\}$.

Goals:

- **modeling**, jointly modeling of spatio-temporal trend and dependence
- **estimation**, efficient methods of estimation.
- **prediction** (or kriging) at unobserved spatial site and/or unknown temporal instant.

3. Multivariate Point referenced data





Data-set *meuse* in *gstat* R package (zinc and cadmium concentration in Netherland).

Here, S is a *continuous* subspace of $\mathbb{R}^d$ and p is the number of the variables considered.

The p-variate random field $X = \{(X_1(s), \ldots, X_p(s))^T, \quad s \in S\}$, is observed at possibly n fixed sites $\{s_1, \ldots, s_n\}$. Goals:

- **modeling**, jointly modeling of multivariate spatial trend and (cross) dependence
- **estimation**, efficient methods of estimation.
- **prediction** (or cokriging) at unobserved spatial site of one or more variables.

Distances

Let us consider a sphere with radius $R > 0$, $R = 6371Km$ **a reasonable approximation of the planet earth**:

$$\mathbb{S}^2 = \{x \in \mathbb{R}^3 : ||x|| = R\}.$$

- The **great circle (GC) or geodesic distance** corresponds to the **shortest path joining two points on the spherical surface** and it is the natural distance to be used when analyzing point reference data.
- Given two points $x_i, x_j \ \in \mathbb{S}^2$ GC distance **is given by**:

$$d_{GC}(x_i, x_j) = R[arccos(x_i^T x_j / R^2)].$$

  However, points on the sphere are **usually given in spherical coordinates**

$$x = (x_1, x_2, x_3) \iff (lon, lat, R)$$

  .

- Specifically, a point $x \in \mathbb{S}^2$ can be parametrized trough spherical coordinates using **radians** (lon^*, lat^*, R) where $lon^* \in [-\pi, \pi]$ is the **longitude** and $lat^* \in [-\pi/2, \pi/2]$ is the **latitude**.

- Very often points on the sphere are expressed in **decimal degrees**

$$(lon, lat, R) = (180lon^*/\pi, 180lat^*/\pi, R)$$

where $lon \in [-180, 180]$ is the **longitude** and $lat \in [-90, 90]$ is the **latitude**.

- The connection between euclidean and spherical coordinates is given by $x = (x_1, x_2, x_3)^T$ with

$x_1 = Rcos(lon^*)cos(lat^*), x_2 = Rsin(lon^*)cos(lat^*), x_3 = Rsin(lat^*)$

.

- Using the geometric definition of the dot product

$$x_i^T x_j = ||x_i|| ||x_j|| cos(\theta_{ij}) = R^2 cos(\theta_{ij})$$

. **This implies that GC distance can be computed as**:

$$d_{GC}(x_i, x_j) = R\theta_{ij}$$

where $0 \leq \theta_{ij} \leq \pi$ is the angle between x_i,x_j.

- To find the angle θ_{ij} between x_i,x_j, we use the relation

$$\theta_{ij} = arccos(x_i^T x_j / R^2)$$

and using spherical coordinates we obtain:

$$\theta_{ij} = \arccos\{\sin lat_i^* \sin lat_j^* + \cos lat_i^* \cos lat_j^* \cos(lon_i^* - lon_j^*)\}$$

- Finally **given two-points given in lon/lat format (decimal degree)** $P_i = (lon_i, lat_i)$, $P_j = (lon_j, lat_j)$ on the sphere of radius R **the GC distance can be computed as**

$$d_{GC}(P_i, P_j) = R\theta_{ij}$$

and $0 \leq d_{GC}(P_i, P_j) \leq R\pi$.

- Alternatively, one can compute **the chordal (CH) distance, that is the segment below the arc joining any pair of points located over the spherical shell**. It is given by:

$$d_{CH}(P_i, P_j) = 2Rsin(\theta_{ij}/2). \qquad \text{(A.1)}$$

  CH obviously **underestimates the GC distance, and the approximation error increases with the size of the considered portion of the planet. It is a euclidean distance on** $\mathbb{R}^3$.

- **Planar projections of the the sphere allows the use of the Euclidean distances in** $\mathbb{R}^2$**.** Most of the (covariance) models used in geostatistics in the last 15 years are valid in $\mathbb{R}^d$ ($d = 2$).

- There is no best projection because none of the available projections is free of distortion.

- Map projection is based on considering points on the sphere $P_j = (lon_j, lat_j, R)$ and transformation P such that $P(P_j) = (x_j, y_j)$ so that $x_j = f(lon_j, lat_j, R)$, $y_j = g(lon_j, lat_j, R)$, where f and g are appropriate functions. Examples:

  – Sinusoidal projection

$$x_j = Rlon_j^* cos(lat_j^*), \quad y_j = Rlat_j^*.$$

– Mercator projection

$$x_j = Rlon_j^*, \quad y_j = Rlog(tan(0.25\pi + 0.5lat_j^*)).$$

- Practice on R 1.

– **Suggested: Banerjee, S. (2005), "On Geodetic Distance Computations in Spatial Modeling," Biometrics, 61, 617–625.**

Random fields

- **Random field** (or stochastic process):
  Let (Ω, F, P) a probability space. A random field is a real valued function $X(s, \omega)$ which:

  - for every $s \in S \subseteq \mathbb{R}^d$ is a measurable function of $\omega \in \Omega$ i.e. (a random variable)
  - for a fixed $\omega \in \Omega$, $X(s, \omega)$ is a non random function of s. It is called sample path or trajectory.

  The dependency on the underlying probability space will usually be suppressed and we denote $X(s, \omega)$ with

  $$X = \{X(s), s \in S \subset \mathbb{R}^d\}$$

  .

- The **finite-dimensional distribution** of the random field X are the distributions of the finite-dimensional vectors

  $$X_n = [X(s_1), \ldots X(s_n)]^T, \qquad \{s_1, \ldots, s_n\} \in S,$$

  for all possible choices of $s_1, \ldots, s_n$ and every n, i.e.

  $$F_{s_1,\ldots,s_n}(k_1, \ldots, k_n) = Pr(X(s_1) \leq k_1, \ldots, X(s_n) \leq k_n).$$

- Given a family of finite-dimensional distributions, Kolmogorov's existence theorem, establish the existence of a stochastic process associated with such family. In particular only multivariate distributions with specific features can be the finite dimensional distribution of a random field (see for instance Abrahmasen 1997).

- **Second-order random field** :
  X is a second-order (or square integrable or L^2) random field if for all $s \in S$, $E(X(s)^2) < \infty$.

  The **mean** of X is the function $m : S \longrightarrow R$ defined by $m(s) = E(X(s))$.

  The **covariance** of X is the function $C : S \times S \longrightarrow R$ defined for all $s_i, s_j \in S$ by

$$C(s_i, s_j) = Cov(X(s_i), X(s_j))$$

.

- Covariances are **positive semidefinite** functions:

$$\sum_i^m \sum_j^m a_i C(s_i, s_j) a_j \geq 0$$

  for each choice of $a_l \in \mathbb{R}^m$ and $\{s_1, \ldots, s_m\}$ and integer m.
  This property is a consequence of non-negativity of the variance of linear combinations:

$$Var(\sum_i^m a_i X(s_i)) = \sum_i^m \sum_j^m a_i C(s_i, s_j) a_j \geq 0$$

- This implies that the matrix $[C(s_i, s_j)]_{i=j=1}^{n}$ is **positive semidefinite**.

- Since covariance functions **must be definite positive** (flexible) **parametric covariance models** are used for the covariance functions

Gaussian random fields

Gaussian random fields are an important class of second order random fields.

- A real-valued random field

$$X = \{X(s),\, s \in S \subset \mathbb{R}^d\}$$

is a Gaussian random field if all the finite-dimensional distributions have a multivariate normal distribution. That is the random vector $X_n = (X(s_1), \ldots, X(s_n))^T$ has a **multivariate normal distribution with mean vector** m **and covariance matrix** Σ which will be denoted by

$$X_n \sim N(m, \Sigma)$$

with

$$m = [m(s_1), \ldots, m(s_n)]^T, \quad \Sigma = [Cov(X(s_i), X(s_j))]_{i=j=1}^n$$

and probability density function

$$f_{X_n}(x_n, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{(x_n - m)^T \Sigma^{-1} (x_n - m)}{2}}$$

- Do Gaussian random fields actually exist? For a multivariate normal distribution it is easy to check that the conditions of the Kolmogorov existence theorem are satisfied.
- **Multivariate normal distribution is completely characterized by its mean and covariance matrix**. Then for a given positive semidefinite function always exist a Gaussian random field with such covariance function.

Stationarity

- Let X a second order random field with mean $m(s)$ and covariance $Cov(X(s), X(s+h))$.

- X is **strictly stationary** if:

$$F_{s_1,s_2,..s_n}(k_1, k_2, \ldots, k_n) = F_{s_1+h,s_2+h,..s_n+h}(k_1, k_2, \ldots, k_n)$$

  for each h, i.e all the finite dimensional distribution are translation invariant.

- X is **weakly stationary** if:

$$m(s) = m, \quad Cov(X(s), X(s+h)) = C(h)$$

  .

- Stationarity in the strict sense implies weak stationarity whereas the opposite is not necessarily true. The two conditions are equivalent for Gaussian random field.

- If $C(h) = C(||h||)$ then the covariance is **isotropic**.

- **Suggested: Chapter 2 in Handbook of spatial statistics, Chapman and Hall 2010.**

- **Suggested: Chapter 1 in 'A Review of Gaussian Random Fields and Correlation Functions' Abrahamsen (1997).**

Covariance functions: properties

Let X a second order random field weakly stationary with covariance C.
Let $C_1, C_2, \ldots C_n$ are stationary covariances. Then (see for instance Gaetan and Guyon 2011):

1. $C(h)$ is a positive semidefinite function.
2. $C(0) = Var(X(s))$, $\forall s$.
3. $|C(h)| \leq C(0), \quad \forall h$.
4. $C(h) = C(-h)$ i.e $C(h)$ is a even function.
5. if A is a full rank matrix, the random field $X(As)$ is stationary with covariance function $C(Ah)$
6. $\sum_{i=1}^{n} a_i C_i(h)$ if $a_i > 0$ is a stationary covariance.
7. $\prod_{i=1}^{n} C_i(h)$ is a stationary covariance.
8. More generally, if $C(\cdot; u)$, $u \in U \subseteq \mathbb{R}^k$ is a stationary covariance for each u and if μ is a positive measure on $\mathbb{R}^k$ such that $C_\mu(h) = \int_U C(h; u)\mu(du)$ exists for all h, then C_μ is a stationary covariance.
9. $\lim_{n\to\infty} C_n(h)$, is a stationary covariance provided that the limit exists for all h.

Random field decomposition

A **classical additive decomposition** for (Gaussian) random fields defined in $\mathbb{R}$ is given by:

$$Y(s) = m(s) + X(s) + \varepsilon(s)$$

where

- $m(s)$ is a **deterministic spatial trend function**.
  The most commonly used parametric mean model is a linear function, given by

  $$m(s) = M(s)^T \beta$$

  where $M(s)$ is a vector of covariates (explanatory variables) observed at s, and β is an unrestricted parameter vector. Alternative choices include nonlinear mean functions.

- X is a weakly stationary (not necessarily) **isotropic** random field with zero mean and covariance function

  $$Cov(X(s), X(s+h)) = C(||h||) = \sigma^2 \rho(||h||)$$

  .

- ε is a zero mean process independent of $X(s)$ and such that

  $$Cov(\varepsilon(s), \varepsilon(s+h)) = C_\varepsilon(0) = \tau^2$$

if $h = 0$ and 0 otherwise.

Covariance parameters

- For the random field Y with isotropic correlation function $\rho(||h||, \alpha)$: The covariance function is given by:

$$C(h; \sigma^2, \tau^2, \alpha) = \begin{cases} \sigma^2 + \tau^2 & ||h|| = 0 \\ \sigma^2 \rho(||h||, \alpha) & ||h|| > 0 \end{cases}$$

where α includes all the correlation parameters (range and other parameters)

- An **alternative parametrization (useful in the non-Gaussian case)** is to include the nugget effect in X that is by considering

$$Y(s) = m(s) + X(s)$$

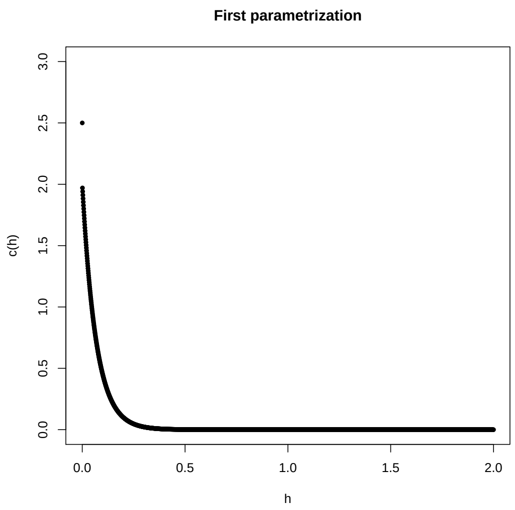
with

$$C(h; \sigma^2, \tau^2, \alpha) = \begin{cases} \sigma^2 & ||h|| = 0 \\ \sigma^2 (1 - \tau^2) \rho(||h||, \alpha) & ||h|| > 0 \end{cases}$$

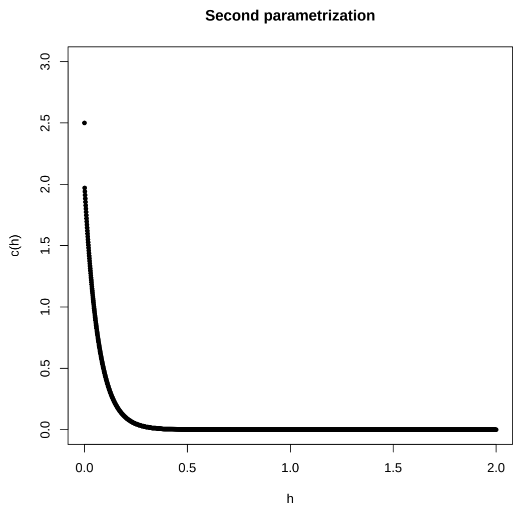
with $0 \leq \tau^2 \leq 1$.

- An example $\rho(||h||, a) = e^{-\frac{||h||}{a}}, \qquad a = 0.2/3$

$\sigma^2 = 2, \tau^2 = 0.5$



$\sigma^2 = 2.5, \tau^2 = 0.2$



Covariance parameters

In the first parametrization

- The **nugget** i.e. $\tau^2 > 0$ ($\sigma^2\tau^2$ in the second parametrization) represents the variability at distances smaller than the typical sample spacing and it can be interpreted as an error measure. It is a discontinuity of the covariance function at the origin.
- The **total variance** (or sill) given by $\sigma^2 + \tau^2$ (σ^2 in the second parametrization).

Correlation parameters :

- The **range** or **spatial scale** parameter
- **Other** parameters associated to the covariance model (smoothing parameters for instance)
- **Practice R 2**

Spectral representation of covariances

- Fourier theory and Bochner's theorem together **imply a bijection between stationary covariances** C **and their spectral densities** f. It is thus equivalent to characterize a stationary model in L^2 by its stationary covariance C or its spectral fensity f.

- Bochner theorem: $C(h)$ **is a covariance function** of a stationary second order random field defined on $\mathbb{R}^d$ **if and only** it can be represented as

$$C(h) = \int_{\mathbb{R}^d} e^{ih^T u} F(du) = \int_{\mathbb{R}^d} e^{ih^T u} f(u) du$$

  where F is a (absolute continuos) non negative bounded measure, called the **spectral measure** and f is the **spectral density**

- The inverse Fourier transform lets us express f in terms of C:

$$f(u) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-ih^t u} C(h) dh$$

- Bochner's theorem **can be stated in terms of characteristic function**, that is **a covariance function is the characteristic function of some** d**-multivariate random vector** and conversely **the characteristic function of some multivariate random vector is a covariance function**.

- For **isotropic covariance function spectral representation simplifies**.

  A real function $\rho(r)$ where $r = ||h||$ with $h \in \mathbb{R}^d$ is an isotropic covariance function **if and only if** it can be represented in the form:

$$\rho(r) = 2^{(d-2)/2}\Gamma(d/2) \int_0^{\infty} \frac{J_{(d-2)/2}(kr)}{(kr)^{(d-2)/2}} f(k)dk$$

  where $f(.)$ is a **positive spectral density** and J_ν is the Bessel function of the first kind of order ν. As an example the case $d = 2$ is:

$$\rho(r) = \int_0^{\infty} J_0(kr)f(k)dk$$

  with spectral density:

$$f(k) = \frac{1}{2\pi} \int_0^{\infty} r\rho(r)J_0(kr)dr$$

  This give a criteria to show a candidate correlation function $\rho(r)$ to be a correlation function. It is sufficient to prove that $f(k) \geq 0, \quad \forall k$.
  Unfortunately, it is quite difficult in general to prove the nonnegativity of the previous integral.

  How to prove in general that a function is semi positive definite on $\mathbb{R}^d$ (i.e. a covariance)??

- Verify that the associate spectral density is positive

- By construction: showing that it can be obtained using properties (sum, product and scale mixture) of covariance functions

- Sufficient conditions?
  Criteria of the Pòlya type: Suppose that $\rho(\cdot)$ is a continuous function with $\rho(0) = 1$ and $\lim_{r \to +\infty} \rho(r) = 0$. The celebrated criterion of Pòlya type (Chilles and Delfiner (1999), Stein (1999)) states that if $\rho(.)$ is convex then it is a covariance function in R^1. In R^2 criteria of the Pòlya type can be found in Gneiting (1999a, 1999b,2001).

- Other criteria: Using completely monotone function: A function $\rho : [0; \infty) \longrightarrow R$ which is continous and which satisfies

  $$(-1)^l \rho(r)^{(l)} \geq 0, \quad r > 0, l = 1, 2, ..$$

  is called completely monotone.
  The link among completely monotone functions and positive semidefinite function is established in the following results:

  A function $\rho(\cdot)$ is completely monotone on $[0, \infty)$ if and only if $\rho(||.||^2)$ is positive definite and isotropic on $\mathbb{R}^n, \forall n$

- **Suggested: Chapter 1.2 in Gaetan and Guyon (2010)**

- **Suggested: Chapter 3,4 in Abrahamsen (1997).**

Intrinsic stationarity

- The weakly stationarity property **may not be sometimes satisfied**. A way to weaken the weakly stationarity hypothesis is to consider **the increments of the random field**.
- X is an **intrinsic random field** if for each $h \in S$, the random field

$$\{X(s+h) - X(s) : \ s \in S\}$$

is weakly stationary or equivalently if the random field has constant mean and the variance of the random field depends on h.

- Let us define the **variogram** of X the function $\gamma : S \to \mathbb{R}$ defined by:

$$2\gamma(h) = Var(X(s+h) - X(s)).$$

- The function $\gamma(h)$ is called the **semi-variogram**
- Every second order stationary random field with covariance C is clearly an intrinsic random field.
- In a stationary random field, given a covariance function C **the relation between the covariance and the semi-variogram is**:

$$\gamma(h) = C(0) - C(h).$$

- **However, the converse is not true.** Variograms of stationary random fields are bounded, because $|C(h)| \leq C(0)$, **This provides a way of checking for weakly stationarity**

- $\gamma(h) = \gamma(-h)$ and $\gamma(0) = 0$

- (Semi)Variograms are conditionally negative definite that is $\forall a \in R^n : \sum_{i=1}^{n} a_i = 0$ and $\{s_1, \ldots, s_n\} \subset S$ then

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma(s_i - s_j) \leq 0$$

- (Semi)Variograms are closed under positive sums and scale mixtures.

- A link between covariances and variograms: If $\gamma(h)$ is a variogram then $C(h) = e^{-a\gamma(h)}$ is a covariance for each $a > 0$ (Schoenberg 1938).

Variogram parameters

- Resuming for a second order stationary random field with isotropic correlation model $\rho(||h||, \alpha)$ the semi-variogram function is given by:

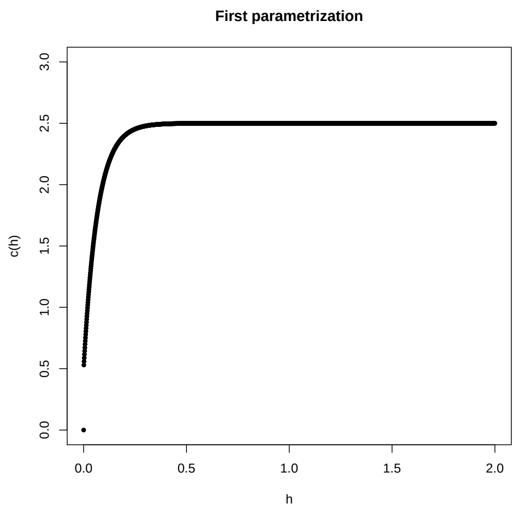
$$\gamma(h, \sigma^2, \tau^2, \alpha) = \begin{cases} 0 & ||h|| = 0 \\ \tau^2 + \sigma^2(1 - \rho(||h||, \alpha)) & ||h|| > 0 \end{cases}$$

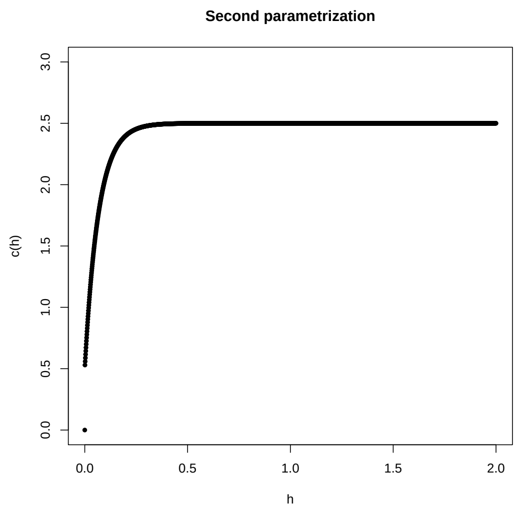
- Using the alternative parametrization

$$\gamma(h; \sigma^2, \tau^2, \alpha) = \begin{cases} 0 & ||h|| = 0 \\ \sigma^2(1 - (1 - \tau^2)\rho(||h||, \alpha)) & ||h|| > 0 \end{cases}$$

- An example $\rho(||h||, \alpha) = e^{-\frac{||h||}{\alpha}}, \qquad \alpha = 0.2/3$

$\sigma^2 = 2, \tau^2 = 0.5$ $\qquad$ $\sigma^2 = 2.5, \tau^2 = 0.2$





- Practice R 3.

Geometric properties: continuity, differentiability

When considering random functions **we are interested in characterizing (modeling) their geometrical properties that is continuity and differentiability** .

These are key properties that **have to be taken into account when choosing a parametric correlation model**.

The correlation functions are responsible for the continuity and differentiability of (Gaussian) random fields.

When studying a function being a random field, continuity is related to the convergence of sequences of random variables. Different types of continuity definition exists for a random field $X = \{X(s), s \in S \subset \mathbb{R}^d\}$(Abrahamsen 1997).

Mean square continuity is the most tractable one:

- A random field X is mean square continous in S if for every sequence s_n for which $||s_n - s|| \to 0$ as $n \to \infty$ then

$$\mathsf{E}(|X(s_n) - X(s)|^2) \to 0.$$

The following theorem provides the **link between the behavior of the covariance function and the mean square continuity of a random field**.

Theorem: **A random field X is mean square continuous at $s \in S$ if and only if its covariance function $C(h)$ is continuous at 0.**

Another type of continuity is related **to the sample paths**

- A random field X has continous sample paths with prob. 1 in S if for every sequence s_n for which $||s_n - s|| \to 0$ as $n \to \infty$ then

$$Pr(|X(s_n) - X(s)| \to 0 \quad) = 1, \quad \forall s \in S \tag{A.2}$$

Key result: Mean square continuity is a necessary and sufficient condition for sample path continuity for Gaussian random fields.

A consequence of this result is that **the presence of a nugget effect implies not mean square continuity and viceversa.**

- **Means square differentiability is also important do describe the smoothness** of the random field.

- X is mean square differentiable if there exists a random field $X^{'}$ such that:

$$E\left[\left(\frac{X(s_n) - X(s)}{||s_n - s||} - X^{'}(s)\right)^2\right] \longrightarrow 0 \quad , as \quad ||s_n - s|| \longrightarrow 0$$

- Higher-order mean-square differentiability is then defined sequentially in the obvious way (that is X is twice mean-square differentiable if $X^{'}$ is mean-square differentiable, and so on.)

- **Key result: The mean-square differentiability of X is directly linked to the differentiability of its covariance function,** according to the following result:

- Let X be a stationary Gaussian process with correlation function $\rho(h)$. Then: X **is** k **times mean-square differentiable if and only if** $\rho(h)$ **is at least** $2k$ **times differentiable at** $h = 0$**.**

- Mean square k differentiability implies mean square continuity

- A Gaussian random field **has** $k-$**differentiable sample paths if the process is** k **times mean-square differentiable**.

- **Suggested: Chapter 1.4 in Gaetan and Guyon (2010)**
- **Suggested: Chapter 2 in Abrahamsen (1997).**

Modeling geometrical properties

- **Continuity and mean square differentiability** are connected with **the behaviour of the correlation function at the origin**
- This implies that the choice of the correlation model **is crucial for modeling these geometrical properties**
- First, **any zero nugget correlation model implies mean square continuity of the sample paths**
- Second, we would like to have a correlation model with **a parameter describing the $2k$ differentiability at the origin** that is **the k differentiability of the sample paths**.
- Two flexible correlation models with these features: Matern and Generalized Wendland models.

Flexible correlation models

The Matérn family of isotropic correlation functions is defined as follows:

$$\mathcal{M}_{\nu,\alpha}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\alpha}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{r}{\alpha}\right), \qquad r \geq 0.$$

where $\mathcal{K}_{\nu}(x)$ is the K_{ν} is the modified Bessel function of the second kind of order ν.

- The model is valid in any dimension d if $\alpha > 0$ and $\nu > 0$
- The parameter ν **indexes the mean squared differentiability of a Gaussian RF having a Matérn correlation function and its associated sample paths.**
- In particular, **for a positive integer k, the sample paths are k times mean-square differentiable, if and only if $\nu > k$**
- Closed form for the special cases for $\nu = 0.5, 1.5, 2.5...$
- Convergence to the Gaussian correlation model

$$\mathcal{M}_{\nu,\alpha/(2\sqrt{\nu})}(r) \xrightarrow[\nu\to\infty]{} \exp(-r^2/\alpha^2)$$

Flexible correlation models

The Generalized Wendland family of isotropic correlation functions is defined as follows:

$$\mathcal{GW}_{\nu,\mu,\beta}(r) = \begin{cases} K\left(1-\left(\frac{r}{\beta}\right)^2\right)^{\nu+\mu} {}_2F_1\left(\frac{\mu}{2}, \frac{\mu+1}{2}; \nu+\mu+1; 1-\left(\frac{r}{\beta}\right)^2\right) & 0 \le r \le \beta \\ 0 & r > \beta, \end{cases}$$

with $K = \frac{\Gamma(\nu)\Gamma(2\nu+\mu+1)}{\Gamma(2\nu)\Gamma(\nu+\mu+1)2^{\mu+1}}$ and ${}_2F_1(a,b,c;x)$ is the Gaussian hypergeometric function.

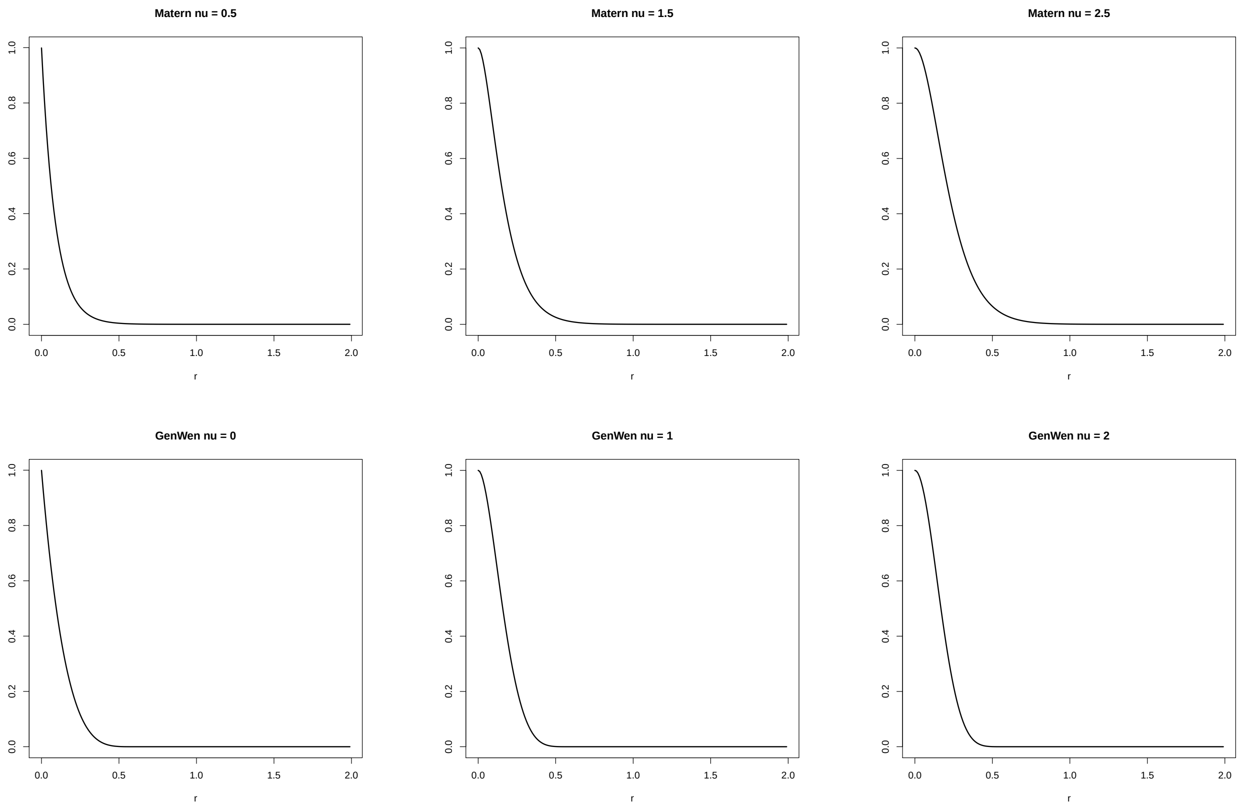
- The model is valid if $\nu \ge 0$, $\beta > 0$ and $\mu \ge (d+1)/2 + \nu$
- The generalized Wendland model **allows for parameterization in a continuous fashion of the mean squared differentiability of the underlying Gaussian RF and its associated sample paths, as in the Matérn case**.
- In particular, for a **positive integer k, the sample paths of the generalized-Wendland model are k times mean-square differentiable**, in any direction, if and only if $\nu > k - 0.5$.
- It is **compactly supported**. The associated correlation matrix is sparse. **Computational gains**.
- Closed form for the special cases for $\nu = 0, 1, 2, ...$

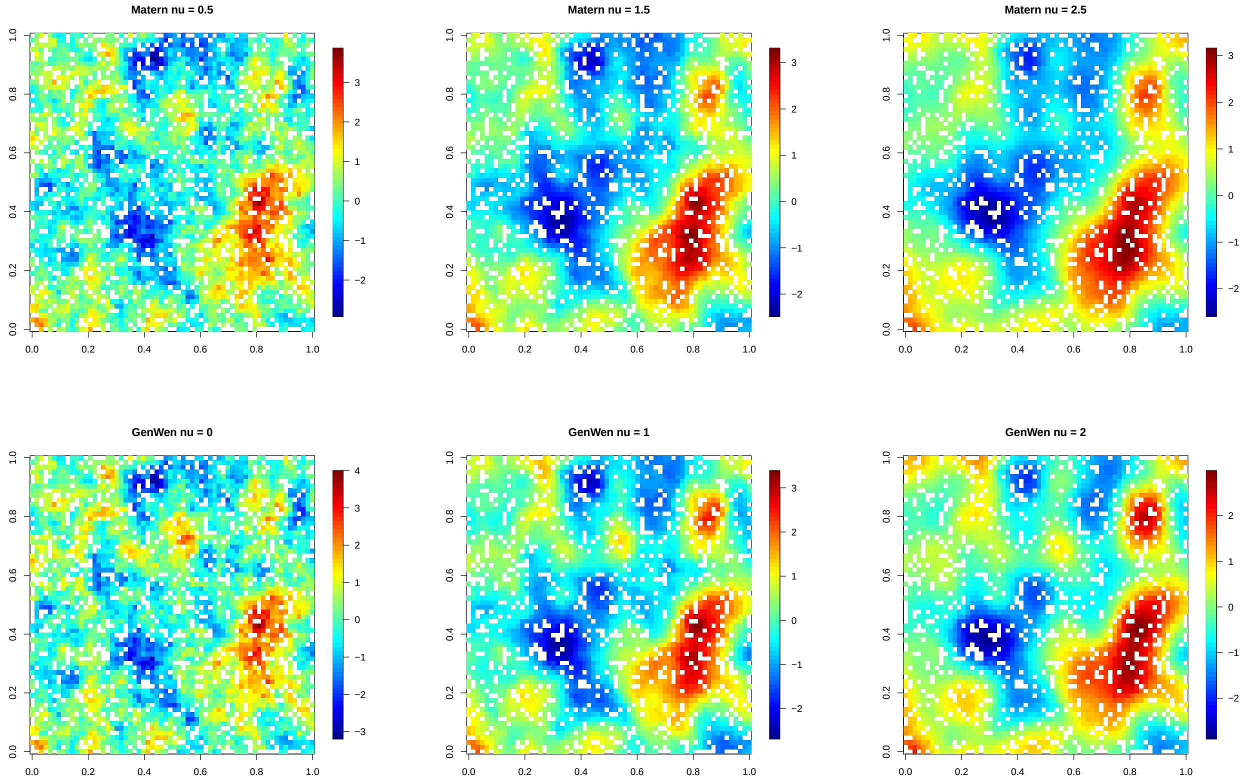
Flexible correlation models

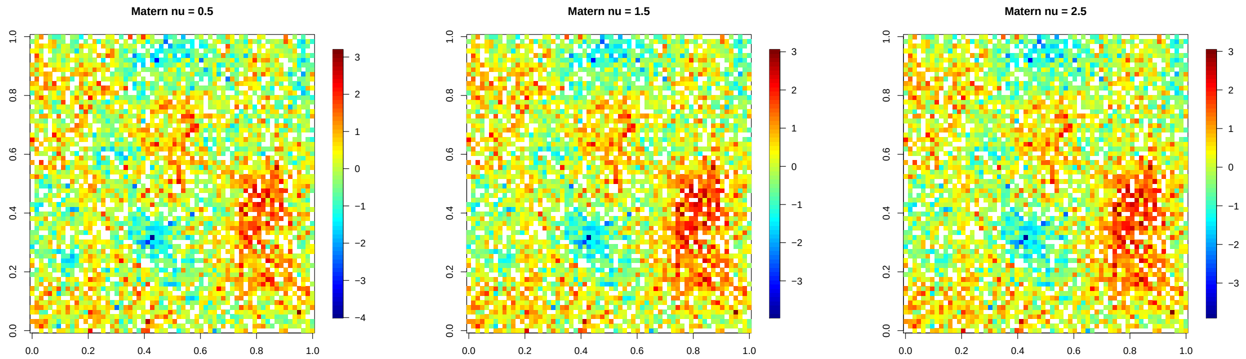
Important special cases:

ν	$\mathcal{GW}_{\nu,\mu,\beta}(r)$	$\mathcal{M}_{\nu+1/2,\alpha}(r)$	m.s.d.
0	$\left(1-\frac{r}{\beta}\right)_+^{\mu}$	$e^{-\frac{r}{\alpha}}$	0
1	$\left(1-\frac{r}{\beta}\right)_+^{\mu+1}\left(1+\frac{r}{\beta}(\mu+1)\right)$	$e^{-\frac{r}{\alpha}}(1+\frac{r}{\alpha})$	1
2	$\left(1-\frac{r}{\beta}\right)_+^{\mu+2}\left(1+\frac{r}{\beta}(\mu+2)+\left(\frac{r}{\beta}\right)^2(\mu^2+4\mu+3)\frac{1}{3}\right)$	$e^{-\frac{r}{\alpha}}(1+\frac{r}{\alpha}+\frac{r^2}{3\alpha^2})$	2

- **R practice** 4, 5.







– **Suggested: Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2019), "Estimation and Prediction using generalized Wendland functions under fixed domain asymptotics," The Annals of Statistics, 47, 828–856.**

Flexible correlation models

A stronger relation:

- A **reparametrized version of the Generalized Wendland models is a generalization of the Matern model**

- Specifically let $\delta_{\nu,\mu,\beta} = \beta \left(\frac{\Gamma(\mu+2\nu+1)}{\Gamma(\mu)} \right)^{\frac{1}{1+2\nu}}$.

- Then

$$\lim_{\mu \to \infty} \mathcal{GW}_{\nu,\mu,\delta_{\nu,\mu,\beta}}(r) = \mathcal{M}_{\nu+1/2,\beta}(r), \quad \nu \geq 0$$

- **Suggested: Bevilacqua, M., Caamano-Carrillo, C., and Porcu, E. (2021), "Unifying Compactly Supported and Matern Covariance Functions in Spatial Statistics,". arXiv:2008.02904v2**

Flexible correlation models

The Generalized Cauchy family of isotropic correlation functions is defined as follows:

$$\mathcal{C}_{\delta,\lambda,\gamma}(r) = \left(1 + (r/\gamma)^{\delta}\right)^{-\lambda/\delta}, \qquad r \geq 0,$$

- The conditions $\delta \in (0, 2]$ and $\lambda > 0, \gamma > 0$, are necessary and sufficient for the validity of the models.
- The γ is the spatial scale parameter.
- The parameter δ is crucial for the differentiability at the origin. Specifically, for $\delta = 2$, they are infinitely times differentiable and they are not differentiable for $\delta \in (0, 2)$.
- If $\lambda \in (0, 1]$ it has long memory that is
  $\int_{\mathbb{R}_+} \mathcal{C}_{\delta,\lambda,\gamma}(r) dr = \infty$
  This basically implies a "very low" decay of the correlation function.
- **R practice**

- **Suggested: Gneiting, T. and Schlather, M. (2004). Stochastic models that separate fractal dimension and the hurst effects. SIAM Rev., 46:269–282**
- **Lim, S. and Teo, L. (2009) Gaussian fields and gaussian sheets with generalized cauchy covariance structure. Stochastic random fields and their Applications, 119(4):1325–1356**

Spatio temporal random field

- Let
$$X = \{X(s,t), (s,t) \in D \subset \mathbb{R}^d \times \mathbb{R}\}$$
a **space time random field**.

- Time could be considered an additional coordinate and, thus, from a mathematical point of view, any spatio-temporal random field can be considered a random field on $\mathbb{R}^{d+1}$.

- From a physical perspective, this view is insufficient. Time differs intrinsically from space, in that time moves only forward, while there may not be a preferred direction in space. Any realistic statistical model will make an attempt to take into account this issue.

- The main interest, as in the spatial case, is on modeling the space time covariance function between $X(s_i, t_i)$ and $X(s_j, t_j)$
$$C((s_i, t_i), (s_j, t_j)) = Cov(X(s_i, t_i), X(s_j, t_j))$$

- **Positive semidefinite property** in the space time case is:
$$\sum_i^m \sum_j^m a_i C((s_i, t_i), (s_j, t_j)) a_j \geq 0$$
for each choice of $a_l \in R^m$ and $\{s_1, \dots, s_m\}$ and integer m.

- **Second-order** and **weakly stationarity** can be easily extended to a space time random field.

- For instance X is **weakly stationarity** if $E(X(s,t)) = m$ and
$$Cov(X(s,t), X(s+h, t+u)) = C(h,u)$$
.

- **Gaussian random field can be easily extended to the space time context**

- Hereafter, we focus on space-time Gaussian random fields that are characterized by the mean and the space time covariance function (matrix)

Space time Covariance and variogram

For the stationary space time covariance function $C(h, u)$.

- $C(h, 0)$ is called the **spatial marginal covariance**
- $C(0, u)$ is called the **temporal marginal covariance**
- For weakly stationary space time random field the semivariogram is given by;

$$\gamma(h, u) = C(0, 0) - C(h, u)$$

where

$$2\gamma(h, u) = Var(X(s + h, t + u) - X(s, t)).$$

is the variogram function

Space time random field decomposition

As in the spatial case we decompose the random field

$$Y(s,t) = m(s,t) + X(s,t) + \varepsilon(s,t)$$

where

- $m(s,t)$ is a **deterministic space time trend function**. In the simplest case, the space time trend function $m(s,t)$, decomposes as the sum of a purely spatial and a purely temporal trend component. The purely spatial component then can be modeled in the ways as already discussed. Temporal trends are often periodic, reflecting diurnal or seasonal effects, and can be modeled with trigonometric functions or nonparametric alternatives. The trend component might depend on environmental temporal and/or spatial covariates.

  The most commonly used parametric mean model is a linear function, given by

  $$m(s,t) = M(s,t)^T \beta$$

  where $M(s,t)$ is a vector of covariates (explanatory variables) observed at s and time t, and β is an unrestricted parameter vector.

- X is a second order **space time weakly stationary random field with zero mean** and covariance function

$$Cov(X(s,t), X(s+h,t+u)) = C(||h||, |u|)$$

- ε is a zero mean process independent of X and such that

$$Cov(\varepsilon(s,t), \varepsilon(s+h,t+u)) = C_{\varepsilon}(0,0)$$

where $C_{\varepsilon}(0,0) = \tau_{st}^2 I((h,u) = (0,0)) + \tau_s^2 I(h = 0) + \tau_t^2 I(u = 0)$ and 0 otherwise, where I denotes an indicator function. Thus we can have three type of nugget effect: spatiotemporal, purely spatial and purely temporal. Hereafter for simplicity we assume $\tau_s^2 = \tau_t^2 = 0$.

Spatial isotropy and temporally symmetry

- A **space-time** covariance is **spatially isotropic** and **temporally symmetric** if

$$Cov(X(s,t), X(s+h, t+u)) = C(||h||, |u|),$$

$||h|| \geq 0, |u| \geq 0.$

that is if it depends on the spatial distance and the temporal distance.

- Most of the space time parametric covariance models proposed in the literature are **spatially isotropic** and **temporally symmetric**.

- A simple example

$$Cov(h, u) = \sigma^2 e^{-\frac{||h||}{\alpha_s}} e^{-\frac{|u|}{\alpha_t}}.$$

where α_s and α_t are spatial and temporal dependence parameter and σ^2 is the variance parameter.

- **R practice**

Space time Covariance parameters

For a **spacetime weakly stationary** random field Y with covariance function $C(h, u, \theta)$ where $\theta \in \Theta \subset R^p$: The covariance function is given by:

$$C(h,u) = \begin{cases} \sigma^2 + \tau^2 & h = 0, u = 0 \\ \sigma^2 \rho(h,u,\alpha) & otherwise \end{cases}$$

Here σ^2 is the variance parameter, τ^2_{st} is the nugget parameter and α include all the correlation parameters (range and other parameters).
Using the alternative parametrization:

$$C(h,u) = \begin{cases} \sigma^2 & h = 0, u = 0 \\ \sigma^2(1-\tau^2)\rho(h,u,\alpha) & otherwise \end{cases}$$

with $0 \leq \tau^2 \leq 1$.

Covariance parameters

Variance parameters:

- The **nugget** parameter τ^2 ($\sigma^2\tau^2$ in the second parametrization) represents the variability at distances smaller than the typical sample spacing and it can be interpreted as an error measure. It is a discontinuity of the covariance function at the origin.
- The total variance (or sill) given by $\sigma^2+\tau^2$ (σ^2 in the second parametrization)

Correlation parameters :

- The **spatial scale** parameter and the **temporal scale** parameter
- **Other** parameters associated to the correlation model (**spatial and temporal smoothing parameters or separability parameters).**

Space time correlation models: separable models

- We need **flexible models able to capture spatial and temporal dependence** and the interaction between them.

- How to obtain valid space-time correlation models?

- **Separable models**:

$$\rho(h, u) = \rho_S(h)\rho_T(u).$$

Given X_1 and X_2 two independent spatial and temporal random fields with covariance $\rho_S(h)$ and $\rho_T(u)$ then it is easy to see that $\rho(h, u)$ is the correlation function of the process X_{12} defined as

$$X_{ST}(s) = \{X_S(s)X_T(t), s \in \mathbb{R}^d, t \in \mathbb{R}^1\}$$

- **Drawback of the separable models**: lack of interaction of time and space.

- Advantage: separable models simplify the calculation associated to the correlation matrix. Given $\mathbf{X} = [X(s_1, t_1), \ldots, X(s_n, t_1), \ldots, X(s_n, t_m)]^T$ the vector of nm observations then the correlation matrix $\Sigma = Var(\mathbf{X})$ can be written as:

$$\Sigma = \Sigma_T \bigotimes \Sigma_S$$

where Σ_T is a $m \times m$ temporal correlation matrix and Σ_S is a $n \times n$ spatial correlation matrix. The inverse and determinant of Σ are then easier to calculate since:

$$\Sigma^{-1} = \Sigma_T^{-1} \bigotimes \Sigma_S^{-1}, \quad |\Sigma| = |\Sigma_T \bigotimes \Sigma_S| = |\Sigma_T|^n |\Sigma_S|^m$$

- An example

$$\rho(h, u) = \mathcal{M}_{\nu_s, \beta_s}(||h||) \mathcal{M}_{\nu_t, \beta_t}(|u|)$$

where ν_s, ν_t are smoothness parameters and β_s, β_t are scale parameters.

Space time correlation models: non separable models

- Need of **more flexible models able to capture the possible interaction among space and time**.

- **Full symmetry**: If

$$cor(X(s_1, t_1), X(s_2, t_2)) = cor(X(s_1, t_2), X(s_2, t_1))$$

  or equivalently $\rho(h, u) = \rho(h, -u)$ then the correlation is fully symmetric. In general this is not true. **However most of the proposed correlation models share this feature**.

- Different proposals in the last years starting **from different approaches**

- A possible approach is **using spectral representation** of the space time covariance function

- Using this approach Gneiting (2002) proposes the following class:

$$\rho(h, u) = \frac{1}{(\phi(|u|^2))^{d/2}} \psi\left(\frac{||h||^2}{\phi(|u|^2)}\right)$$

  It is a **valid space time correlation model** if $\psi(\cdot)$ is completely monotone function and $\phi(\cdot)$ is a Bernstein function.

- An example is choosing $\psi(\cdot) = \mathcal{M}_{\nu_s,\beta_s}(\sqrt{(\cdot)})$ as completely monotonic function and $\phi(\cdot) = (1 + \frac{(\cdot)^\alpha}{a})^\beta$ as Bernstein function. After a useful reparametrization we obtain:

$$\rho(h,u) = \frac{1}{\left(1+\frac{|u|^{\nu_t}}{\beta_t}\right)^\gamma}\mathcal{M}_{\nu_s,\beta_s}\left(\frac{||h||}{\left(1+\frac{|u|^{\nu_t}}{\beta_t}\right)^{\beta/2}}\right)$$

where β_s, β_t are positive **temporal and spatial scale parameters**, ν_t and ν_s are **temporal and spatial smoothness parameters**, $\gamma \geq d/2$ and $0 \leq \beta \leq 1$ is an interaction parameter that allows for **separability** $\beta = 0$ or **non separability** $0 < \beta \leq 1$.

- **R practice**

- Another approach is to consider **scale mixtures of two correlation functions**. (De Iaco Myers Posa (2002) and Ma(2003) for instance). If μ is finite positive measure on A and for each $a \in A$ $\rho_S(h,a), \rho_T(u,a)$ are stationary covariances then.

$$\rho(h,u) = \int \rho_S(h,a)\rho_T(u,a)d\mu(a), \quad h \in \mathbb{R}^d, u \in R$$

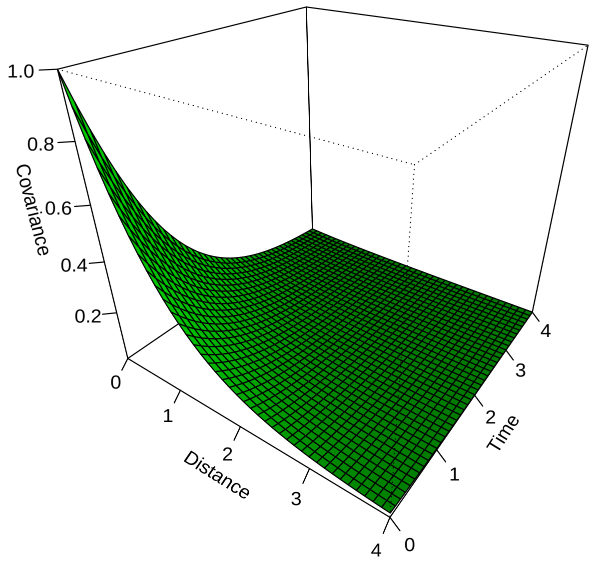
.

is a **valid space time correlation function**. For instance choosing μ as the gamma distribution and ρ_S and ρ_t of the powered exponential type we obtain the family:

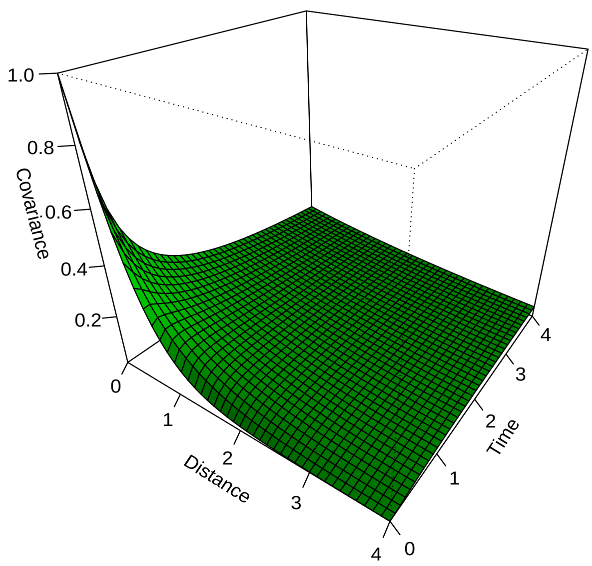
$$\rho(h,u) = \sigma^2 \left(1 + \left(\frac{||h||}{b} \right)^{\alpha} + \left(\frac{|u|}{c} \right)^{\beta} \right)^{-\gamma}$$

with $0 < \alpha, \beta \leq 2$, $\gamma, a, b > 0$

Space time correlation models



Separable (double exponential model)



Non-separable (Gneiting model)

Suggested:

- **Geostatistical Space-Time Models, Stationarity,Separability, and Full Symmetry T. Gneiting, M. Genton, and P.Guttorp. In Statistical Methods for Spatio-Temporal Systems Chapman and Hall**
- **Continuous Parameter Spatio-Temporal random fields T. Gneiting and P.Guttorp. In Handbook of spatial statistics Chapman and Hall.**
- **Gneiting, T. (2002), Nonseparable, stationary covariance functions for space-time data, Journal of the American Statistical Association, 97, 590–600.**
- **Porcu, E, Furrer, R, Nychka, D. 30 Years of spacetime covariance functions. WIREs Comput Stat. 2021;13: e1512**

Anisotropy

In anisotropic random fields, the correlation between two locations is direction dependent. The form of that anisotropy generally falls into one of two cases: geometric or zonal anisotropy.

- Geometric anisotropy: the ranges vary depending on the direction while the sill is constant. Correcting geometric anisotropy involves rotating the coordinate system i.e. to multiplying the coordinates of the spatial random field in by a matrix A of size $d \times d$ that deforms the space appropriately. That is the random field X is geometrically anisotropic, but $X(As)$ is isotropic with covariance $C(||Ah||)$.

- Zonal anisotropy: in this case both sill and ranges vary depending on the direction. In this case linear combination of (possibly anisotropic) variograms that depends on some components of the vector h can be useful to solve the problem. For instance the variogram:

$$\gamma(h) = \gamma_1(||h||) + \gamma_2(|h_2|)$$

  has different sills.

- **Suggested: Chapter 1.3 in Gaetan and Guyon (2010)**
- **Suggested: Chapter 2.5.2 in Chiles and Delfiner (2010)**

Correlation model defined on the sphere

- The spatial, spatio-temporal and multivariate correlation models seen so far **are defined in the euclidean space and using the euclidean distance**

- However when **data are collected over the whole planet**, or a big portion of it, projecting the coordinates results in **big distortion** of the distances

- In the last ten years new correlation models have been proposed **using the great-circle distance of the sphere** with radius $R > 0$

$$\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : ||x|| = R\}.$$

with $d = 3$ and $R = 6371$ **it is a good approximation of the planet earth**. Hereafter $d = 3$, $R = 1$.

- Correlation functions that depends on the geodesic distance

$$d_{GC}(x_i, x_j) = [arccos(x_i^T x_j)], \quad x_i, x_j \in \mathbb{S}^2.$$

with $0 \leq d_{GC} \leq \pi$ are called **geodesically isotropic**

- Flexible correlation models have parametric restriction when replacing the eucliden distance with the geodesic distance

- For example the **Matern model** using the great circle distance

$$\mathcal{M}_{\nu,\beta}(d_{GC}) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{d_{GC}}{\beta}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{d_{GC}}{\beta}\right),$$

  is valid if $0 < \nu \leq 0.5$ in $\mathbb{S}^2$.

- The **Generalized Wendland model** using the great circle distance

$$\mathcal{GW}_{\nu,\mu,\beta}(d_{GC}) = \begin{cases} K\left(1-\left(\frac{d_{GC}}{\beta}\right)^2\right)^{\nu+\mu} {}_2F_1\left(\frac{\mu}{2}, \frac{\mu+1}{2}; \nu+\mu+1; 1-\left(\frac{d_{GC}}{\beta}\right)^2\right) & 0 \leq d_{GC} \leq \\ 0 & d_{GC} > \beta, \end{cases}$$

  is valid if $\mu > 2 + \nu$ in $\mathbb{S}^2$.

- The **Generalized Cauchy model** using the great circle distance

$$\mathcal{C}_{\delta,\lambda,\gamma}(d_{GC}) = \left(1 + (d_{GC}/\gamma)^{\delta}\right)^{-\lambda/\delta}$$

  is valid if $\delta \in (0, 1]$ in $\mathbb{S}^2$.

- New contributions **for spatial, spatio temporal and multivariate spatial cases** can be found in Alegría et al. (2020). Porcu et al. (2016) and Bevilacqua et al. (2020) respectively.

- Using chordal distance **no restrictions** on the parameters since it is a euclidean distance in $\mathbb{R}^3$.

Suggested:

- **Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. Bernoulli, 19(4):1327–1349.**
- **Porcu, E., Bevilacqua, M., and Genton, M. G. (2016). Spatio-temporal covariance and crosscovariance functions of the great circle distance on a sphere. Journal of the American Statistical Association, 111(514):888–898.**
- **Alegría A., Cuevas-Pacheco F., Diggle P., Porcu E., The F-family of covariance functions: A Matérn analogue for modeling random fields on spheres, Spatial Statistics, Volume 43, 2021,**
- **Bevilacqua M., Porcu E., Diggle P. (2020) Families of Covariance Functions for Bivariate Random Fields on Spheres. Spatial Statistics.**

Semi-Variogram estimation

- The natural empirical estimator of $\gamma(h)$ is the moment estimator :

$$\widehat{\gamma}_n(h) = \frac{1}{2\#N(h)} \sum_{(s_i,s_j)\in N(h)} (X(s_i) - X(s_j))^2, \qquad h \in \mathbb{R}^d. \tag{A.3}$$

where $\#$ is the set cardinality.

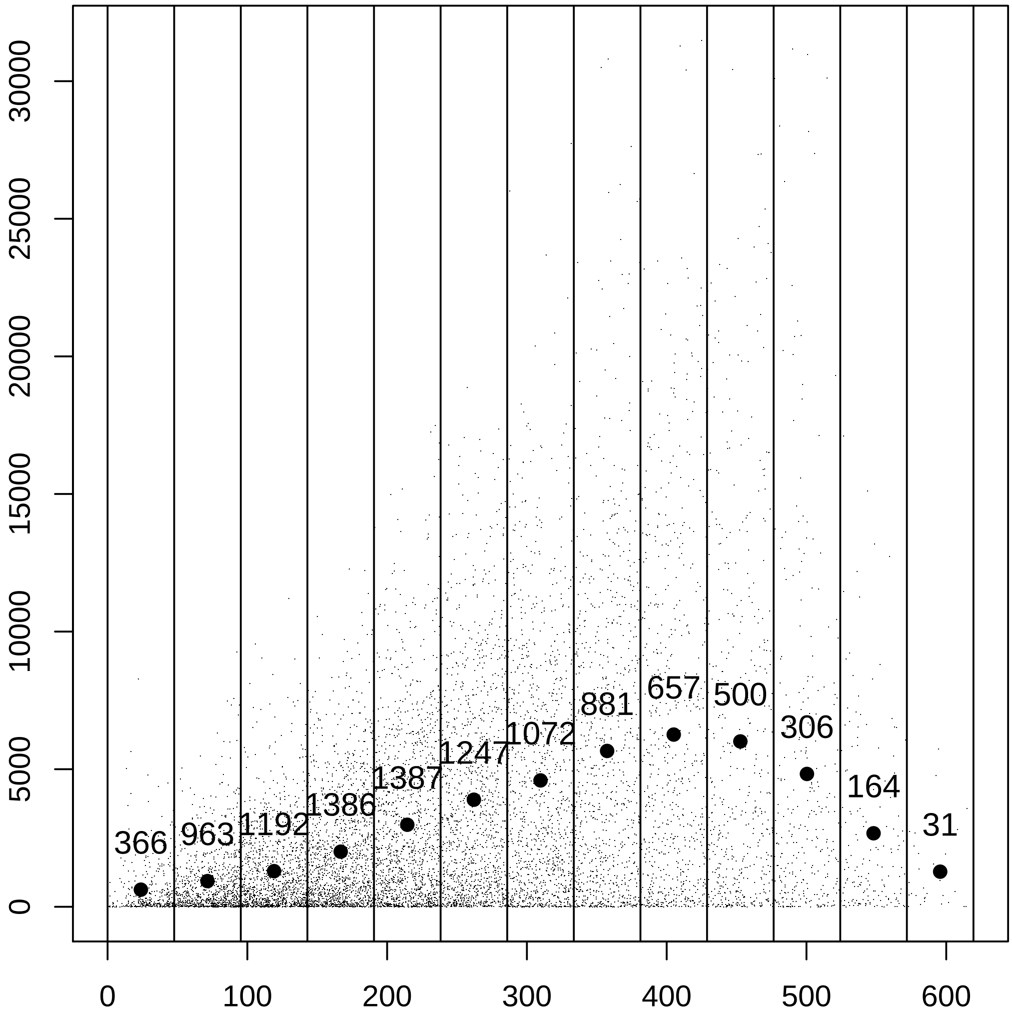
- In the isotropic case and in a regular grid (lattice), $N(h)$ is the set defined as:

$$N(h) = \{(s_i, s_j) : \|s_i - s_j\| = ||h|| \,;\, i, j = 1, \ldots, n\}.$$

In the more general case $N(h)$ is defined as:

$$N(h) = \{(s_i, s_j) : \; ||h|| - \Delta \leq ||h|| \leq ||h|| + \Delta \,;\, i, j = 1, \ldots, n$$

where $\Delta > 0$ defined a tolerance region around the distance $||h||$.



- In practice, we estimate the semivariogram $\gamma(\cdot)$ at a finite number k of lags:

$$\mathcal{H} = \{h_1, h_2, \ldots, h_k\},$$

- Semi-Variogram estimation **can be not efficient for large lags**. In practice the semivariogram is computed **for lags lower than the maximum distance**. A practical rule is half the maximum distance

- A priori, the semi-variogram is not isotropic and it is prudent to consider several orientations of the vector h and evaluate the semi-variogram in different directions. In this way we can empirically detect possible anisotropy in the semi-variogram.

- **Bounded or unbounded empirical semi-variogram** give informations about **the weakly stationarity of the process** and about **the presence of a nugget**.

- If X is second-order stationary, the covariance can be empirically estimated by

$$\widehat{C}_n(h) = \frac{1}{\#N(h)} \sum_{s_i, s_j \in N(h)} (X(s_i) - \overline{X})(X(s_j) - \overline{X}), \qquad h \in \mathbb{R}^d, \tag{A.4}$$

where $\widehat{\mu} = \overline{X} = n^{-1} \sum_{i=1}^{n} X(s_i)$.

- $2\widehat{\gamma}_n(h)$ in comparison to $\widehat{C}_n(h)$ does not require a prior estimate of the mean μ.

- Generalization to the space time case:

$$\widehat{\gamma}_n(h, u) = \frac{1}{2\#N(h, u)} \sum_{(s_i, s_j, t_l, t_k) \in N(h, u)} (X(s_i, t_l) - X(s_j, t_k))^2, \tag{A.5}$$

where $\#$ is the set cardinality and

$$N(h, u) = \{(s_i, s_j, t_l, t_k) :$$
$$(s_i, s_j) : ||h|| - \Delta \leq ||h|| \leq ||h|| + \Delta \; ; \; i, j = 1, \ldots, n$$

$$(t_l, t_k) : |t_l - t_k| = |u|; l, k = 1, \ldots, T\}.$$

- **R practice**

Least squares estimation for parametric variogram models

Let $\gamma(\cdot;\theta)$ with $\theta \in \mathbb{R}^p$ a parametric variogram model.

- **ordinary least squares** (OLS)

$$\widehat{\theta}_{OLS} = \underset{\theta\in\Theta}{\text{argmin}} \sum_{i=1}^{k} (\widehat{\gamma}_n(h_i) - \gamma(h_i;\theta))^2, \tag{A.6}$$

  where k is the number bins

- **generalized least squares** (GLS) The OLS method generally performs poorly as the $\widehat{\gamma}_n(h_i)$ are neither independent nor have the same variance.

$$\widehat{\theta}_{GLS} = \underset{\theta\in\Theta}{\text{argmin}} (\widehat{\gamma}_n - \gamma(\theta))' \{Cov_\theta(\widehat{\gamma}_n)\}^{-1} (\widehat{\gamma}_n - \gamma(\theta)), \tag{A.7}$$

  where

  $\widehat{\gamma}_n = (\widehat{\gamma}_n(h_1), \ldots, \widehat{\gamma}_n(h_k))'$.
  and $\gamma(\theta) = (\gamma(h_1,\theta), \ldots, \gamma(h_k,\theta))'$.

- **weighted least squares** (WLS). As calculating $Cov_\theta(\widehat{\gamma}_n)$ is often difficult, we use

$$\widehat{\theta}_{WLS} = \underset{\theta\in\Theta}{\text{argmin}} \sum_{i=1}^{k} \frac{\#N(h_i)}{\gamma^2(h_i;\theta)} (\widehat{\gamma}_n(h_i) - \gamma(h_i;\theta))^2. \tag{A.8}$$

- These three methods operate under the same principle, *least squares estimation* (LSE): for $V_n(\theta)$ a $k \times k$ p.d. symmetric matrix with known parametric form, we want to minimize the distance $U_n(\theta)$ between $\gamma(\theta)$ and $\widehat{\gamma}_n$:

$$\widehat{\theta}_{LSE} = \underset{\theta \in \Theta}{\text{argmin}}\, U_n(\theta), \tag{A.9}$$

  where

$$U_n(\theta) = (\widehat{\gamma}_n - \gamma(\theta))' V_n(\theta)(\widehat{\gamma}_n - \gamma(\theta)).$$

  Closed form expression for $Cov_\theta(\widehat{\gamma}_n)$ exist in the Gaussian case (see for instance) Convergence and asymptotic normality of the LSE (Lahiri, Lee and Cressie (2002)).

- **Drawback**: Least square methods are not the best methods from statistical efficiency point of view.

- **Advantage**: they only requires the variogram knowledge.

Variogram estimation in a presence of a parametric trend

- Let $X = \{X(s), s \in \mathbb{R}^d\}$ be a weakly second order random field
- Then we decompose the random field:

$$X(s) = Z(s)^T \beta + \varepsilon(s)$$

- where we assume $\mathrm{E}(\varepsilon(s)) = 0$ and covariance function $Cov(\varepsilon(s+h), \varepsilon(s)) = C(h, \theta)$. Here $C(h, \theta)$ is a parametric covariance model with $\theta \in \Theta \subset R^p$.
- Let $x = (x(s_1), \dots, x(s_n))^T$ a realization from X.
- In correspondence we can observe a set of q covariates $i.e$ $z_i = (z_i(s_1), \dots, z_i(s_n)^T$. Then $z = [z_1, z_2, \dots, z_q]$ is a $(n \times q)$ full rank regression matrix.
- In matrix notation:

$$x = z\beta + \varepsilon, \qquad \beta \in \mathbb{R}^q, \tag{A.10}$$

with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2 \Sigma$, where $\Sigma = \Sigma(\theta)$

The model parameters are $(\beta, \sigma^2, \theta) \in \mathbb{R}^{p+q+1}$. By premultiplying by $\Sigma^{-\frac{1}{2}}$ we obtain the model:

$$x^* = z^* \beta + \varepsilon^* \tag{A.11}$$

where $x^* = \Sigma^{-\frac{1}{2}}x$, $z^* = \Sigma^{-\frac{1}{2}}z$ and $E(\varepsilon^*) = 0$ and $Var(\varepsilon^*) = \sigma^2 I$.
This model is thus a standard linear model and we know (Gauss-Markov theorem) that the best linear estimator, unbiased and with minimal variance for β is :

$$\beta = (z^{*T}z^*)^{-1}(z^{*T}x^*)$$

that is

$$\beta_{GLS}(\Sigma) = (z^T\Sigma^{-1}z)^{-1}(z^T\Sigma^{-1}X)$$

the GLS estimator. Since Σ is unknown the parameters are estimated with the following alghorithm:

1. Estimate β by $\widehat{\beta}$ using a method that does not require knowledge of θ, for example OLS.
2. Calculate the residuals $\widehat{\varepsilon} = x - z'\widehat{\beta}$
3. Based on $\widehat{\varepsilon}$ estimate θ using a previously developed method LS square methods for instance.
4. Estimate Σ by $\hat{\Sigma} = \Sigma(\tilde{\theta})$ then β with $\beta_{GLS}(\hat{\Sigma})$
5. Steps 2 to 4 can be iterated until convergence of estimations (see Gaetan's book)

Maximum likelihood estimation

- Let
$$Z = \{Z(s), s \in \mathbb{R}^d\}$$
be a **weakly second order Gaussian spatial random field**.

- We suppose $\mathsf{E}[Z(s)] = X(s)^T\beta$ and $C(Z(s), Z(s+h)) = \sigma^2\rho(h, \theta)$.

- $\rho(h, \theta)$ is a **parametric correlation model**

- where

  - $\sigma^2 > 0$ is the variance
  - $\beta \in \mathbb{R}^q$ is the **mean vector parameter** and
  - $\theta \in \Theta \subset \mathbb{R}^p$ is the vector of **correlation parameters**.

  We denote with $\psi = (\beta, \theta, \sigma^2)^T \in \Psi \subset \mathbb{R}^{p+q+1}$ the vector parameters.

- Suppose we have a **realization** at a finite number of location sites $(s_1, \ldots, s_n)^T$ that is a realization $z = (z(s_1), \ldots, z(s_n))^T$. of the random vector $Z = (Z(s_1), \ldots, Z(s_n))^T$.

- In correspondence we can observe a set of q covariates $i.e$ we observe the n column vector $x_i = (x_i(s_1), \ldots, x_i(s_n))^T$.

  Then $X = [x_1, x_2, \ldots, x_q]$ is a $(n \times q)$ full rank regression matrix.

- Then $Z \sim N(X\beta, \Sigma)$ where $\Sigma = \Sigma(\theta, \sigma^2) = \sigma^2 R(\theta)$ and

$$R(\theta) = \rho(||s_i - s_j||, \theta)$$

is the **correlation matrix**.

- The **loglikelihood function** is:

$$l_Z(\psi) = -0.5(nlog(2\pi) + log|\Sigma| + (z - X\beta)^T)\Sigma^{-1}(z - X\beta)$$

- Joint maximization of the previous function leads to **ML estimator**.

$$\widehat{\psi}_{ML} = (\hat{\beta}, \hat{\theta}, \hat{\sigma}^2)^T = \underset{\psi \in \Psi}{\text{argmax}}\, l_Z(\psi)$$

- The problem of **maximizing the loglikelihood leads to a nonlinear optimization problem** for which a closed-form solution exists only in very special cases. In general, therefore, ML estimates must be obtained numerically. using iterative procedure for maximization such as **methods based on gradient** (Newton Raphson and Fisher scoring for instance) or **gradient free methods** such as the simplex algorithm (Nelder and Mead, 1965).

- **Drawback**: Computation of maximum likelihood estimation (or its gradient) requires $O(n^3)$ operations. **This is an issue for large datasets**.

- Some problems with **multimodality** of likelihood with specific covariance model (Mardia, Watkins (1989)).

- Computing the estimating equations of the loglikelihood function we can obtain:

$$\hat{\beta}(\theta) = (X^T R X)^{-1}(X^T R Z)$$

$$\hat{\sigma}^2(\beta, \theta) = \frac{(Z - X\beta)^T R^{-1} (Z - X\beta)}{n}$$

Then, as alternative maximization method we can use the profile likelihood function:

$$l_{Prof}(\hat{\beta}, \hat{\sigma}^2, \theta) = -0.5(n log(2\pi) + n log(\hat{\sigma}^2) + log|R| + n)$$

So first we maximize $l_{Prof}(\hat{\beta}, \hat{\sigma}^2, \theta)$ and we obtain $\hat{\theta}$

Then we obtain $\hat{\beta}(\hat{\theta})$ and then $\hat{\sigma}^2(\hat{\beta}, \hat{\theta})$.

- **Asymptotic properties of ML estimators**.
Classical asymptotic results under **increasing domain** (Mardia and Marshall 1984) says that ML estimator is **consistent and asymptotically normal** with variance-covariance matrix the inverse of the **Fisher Information matrix** $I(\psi) = -E[l^{''}(\psi)]$ whose generic entries is

$$I(\psi)_{i,j} = \left(\frac{dX\beta}{d\psi_i}\right)^T \Sigma^{-1} \frac{dX\beta}{d\psi_j} + \frac{1}{2} tr\left(\frac{d\Sigma}{d\psi_i}\Sigma^{-1}\frac{d\Sigma}{d\psi_j}\Sigma^{-1}\right)$$

The entries of the Fisher Information entries have closed form a part for the correlation parameters:

$$I(\psi) = \begin{bmatrix} \frac{X^T R^{-1} X}{\sigma^2} & 0 & 0 \\ - & \frac{n}{2\sigma^4} & \frac{1}{2\sigma^2} tr\left(\frac{dR}{d\psi_j} R^{-1}\right) \\ - & - & \frac{1}{2} tr\left(\frac{dR}{d\psi_i} R^{-1} \frac{dR}{d\psi_j} R^{-1}\right) \end{bmatrix}$$

This implies:

– $\widehat{\psi}_{ML} \xrightarrow{p} \psi$

– $\widehat{\psi}_{ML} \overset{d}{\approx} N(\psi, I(\psi)^{-1})$

Square root of the diagonal elements of $I(\psi)^{-1}$ **provide standard error estimation for each parameter**.

- Model selection. Akaike information Criteria

$$AIC = -2(l(\widehat{\psi}_{ML}) - p)$$

Given a set of candidate (covariance) models for the data, the preferred model is the one with the minimum AIC value.

- **Results under infill asymptotics**:
Under fixed domain asymptotics, **no general results are available for the asymptotic properties of ML estimators**.
It turns out that, the the parameters of the covariance function cannot be consistently estimated under these

framework.
For instance for the Matern model $\sigma^2\mathcal{M}_{\nu,\alpha}$ it can be shown that **only the parameter**:

$$\frac{\sigma^2}{\alpha^{2\nu}}$$

can be estimated consistently and in this case.

- $\frac{\hat{\sigma}^2}{\hat{\alpha}^{2\nu}} \xrightarrow{p} \frac{\sigma^2}{\alpha^{2\nu}}$, and
- $\frac{\hat{\sigma}^2}{\hat{\alpha}^{2\nu}} \overset{d}{\approx} \mathcal{N}(\frac{\sigma^2}{\alpha^{2\nu}}, \frac{2(\sigma^2/\alpha^{2\nu})^2}{n})$.

Similar results for the **Generalized Wendland model**.

- **R practice** (ML space, ML spacetime, ML sphere)

Suggested:

- **Mardia, K. V. and Marshall, J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 71, 135–146.**
- **Zhang, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. Journal of the American Statistical Association 99, 250–261.**
- **Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2019), "Estimation and Prediction using generalized Wendland functions under fixed domain asymptotics," The Annals of Statistics, 47, 828–856.**

Alternative methods of estimation

- LS estimation method are **statistically inefficient**.

- Likelihood based method is the most efficient. **However it is computationally intensive since Gaussian likelihood evaluation involves $O(n^3)$ operation and $O(n^2)$ memory storage.** This is a problem for large datasets.

- This **computational bottleneck is due to the computation of the inverse and of the determinant of the covariance matrix** which is usually done trough cholesky decomposition.

- Different methods of estimation have been proposed in the last years with a **good trade-off between computational complexity and statistical efficiency.**

- For instance: methods based on modification of the covariance structure (low rank methods, Markov fields approximation, covariance tapering), methods based on modification of the likelihood (composite likelihood, Vecchia methods)

Suggested:

- **Heaton, M.J., Datta, A., Finley, A.O. et al. A Case Study Competition Among Methods for Analyzing Large Spatial Data. JABES 24, 398–425 (2019).**

- **Sun, Y., Li, B., and Genton, M. G. (2012), "Geostatistics for large datasets," in Advances and challenges in space-time modelling of natural events, Springer, pp. 55–77.**

Maximum Composite likelihood estimation

Let us consider (for simplicity) a spatial random field

$$Z = \{Z(s), s \in A\}.$$

For any set of distinct points $(s_1, \ldots, s_n)^T$, $s_i \in A$, let

$$Z = (Z(s_1), \ldots, Z(s_n))^T$$

the associate **multivariate random vector** with multivariate parametric density $f_Z(z;\theta)$.

- Suppose that $f_Z(z;\theta)$ is **difficult to evaluate or to specify**, but that **it is possible to compute or specify distribution for some subsets of Z.**
- It may be expedient to consider instead a pseudolikelihood **compounding the likelihood of such subsets**.
- Let B_k be a marginal or conditional set of Z. The log-composite likelihood is an objective function defined as a sum of K sub-likelihoods

$$CL(\theta) = \sum_{k=1}^{K} l_{B_k}(\theta) w_k, \tag{A.12}$$

where $l_{B_k}(\theta)$ is a log-likelihood calculated **by considering only the random variables in** B_k and w_k are suitable weights that do not depend on θ.

- **Drawback**: general loss of statistical efficiency is expected from CL estimation with respect to ML methods.
- **Benefit I**: computational tractability.
- **Benefit II**: it requires only model assumptions on lower dimensional marginal densities, and not detailed specification of the full joint distribution.

Maximum Composite likelihood estimation based on pairs

- An interesting setting (from statistical and computational efficiency point of view) **is by considering composite likelihood based on pairs**.

- Setting $B_k = Z_{ij} = (Z(s_i), Z(s_j))^T$ or $B_k = Z_{i|j} = Z(s_i)|Z(s_j)$ we obtain the **pairwise marginal** log-likelihood $l_{ij} = log(f_{Z_{ij}})$ and the **pairwise conditional** log-likelihood $l_{i|j} = log(f_{Z_{i|j}})$ respectively.

- The corresponding weighted **composite log-likelihoods function** are given by:

$$wpl_M(\theta) = \sum_{i=1}^{n}\sum_{j\neq i}^{n} l_{ij}(\theta)w_{ij}, \qquad wpl_C(\theta) = \sum_{i=1}^{n}\sum_{j\neq i}^{n} l_{i|j}(\theta)w_{ij}.$$

- Assuming symmetric weights, $i.e\ w_{ij} = w_{ji}$ then the previous expression can be simplified, up to a proportional constant, as

$$wpl_M(\theta) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} l_{ij}(\theta)w_{ij}, \quad wpl_C(\theta) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} [2l_{ij}(\theta) - (l_i(\theta) + l_j(\theta))]$$

  where $l_k(\theta) = log(f_{Z_k})$ is the marginal log-likelihood.

- Then

$$\widehat{\theta}_X = \underset{\theta\in\Theta}{\text{argmax}}\,(wpl_X(\theta)), \quad X = M, C$$

  is the maximum CL estimator based on pairs (marginal and conditional).

- The **role of the weights is to improve the statistical efficiency of the method**

- A symmetric weight function based on distance frequently used is:

$$w_{ij}(k) = \begin{cases} 1 & ||s_i - s_j|| < k \\ 0 & \text{otherwise} \end{cases}.$$

  where $k > 0$ is an arbitrary distance greater than the minimum distance of the location points.

- An alternative (a)symmetric weight function based on neighborhoods is:

$$w_{ij}(m) = \begin{cases} 1 & s_i \in N_m(s_j) \\ 0 & \text{otherwise} \end{cases}$$

  where $N_m(s_l)$ is the set of the neighbors of order $m = 1, 2, \ldots$ of the point s_l.

- Under increasing domain it can be shown that for $X = M, C$

  – $\widehat{\theta}_X \xrightarrow{p} \theta$

  – $\widehat{\theta}_X \stackrel{d}{\approx} N(\theta, G(\theta)^{-1})$

  where

$$G_X(\theta) = H_X(\theta) J_X(\theta)^{-1} H_X(\theta)^{\intercal}$$

  with $H_X(\theta) = -\mathsf{E}[wpl''_X(\theta)], \qquad J_X(\theta) = \mathsf{E}[wpl'_X(\theta) wpl'_X(\theta)^T].$

- Estimation of G can be complicated. A useful tool: parametric bootstrap.

  Moreover, model selection can be performed by considering an information criterion, defined as

$$PLIC = -2\left(wpl_X(\widehat{\theta}_X) - tr(J_X(\widehat{\theta})H_X^{-1}(\widehat{\theta})\right)$$

- **R practice**

  Suggested:

  - **Varin, C., Reid, N., and Firth, D. (2011), "An overview of composite likelihood methods," Statistica Sinica, 21, 5–42.**
  - **Bevilacqua, M., and Gaetan, C. (2015), "Comparing composite likelihood methods based on pairs for spatial Gaussian random fields," Statistics and Computing, 25, 877–892.**