Geostatistics

Mathieu Ribatet—Full Professor of Statistics

Outline

References

[1] J.-P. Chilès and P. Delfiner. *Geostatistics: Modelling Spatial Uncertainty*. Wiley, New York, 1999.

[2] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. John Wiley & Sons inc., New York, 1993.

[3] P. Diggle, P.J. Ribeiro, and P. Justiniano. *Model-based Geostatistics*. Springer Series in Statistics. Springer, 2007.

[4] Carlo Gaetan and Xavier Guyon. *Spatial Statistics and Modeling*. Springer Series in Statistics. Springer Science+Business Media, LLC, New York, NY, 2010.

[5] M. L. Stein. *Interpolation for spatial data: Some theory for kriging*. Springer, New York, 1999.

[6] H. Wackernagel. *Multivariate geostatistics: An introduction with applications*. Springer, New York, third edition edition, 2003.

1. Introduction

Motivation

- Many variables are spatial in extent, e.g., rainfall, petroleum, elevation
- The use of univariate or even multivariate statistical models may be too restrictive.
- An example would be to try to estimate the expected surface of a pollutant exceeding some critical level u_{Craft} in a study region $\mathcal{X} \subset \mathbb{R}^d$, i.e.,

$$\text{Area}(u_{\text{crit}}) = \mathbb{E}\left[\int_{\mathcal{X}} 1_{\{Y(s)>u_{\text{crit}}\}} \text{d}s\right],$$

where $Y(s)$ is the amount of pollutant at location s.

The use of univariate models may still be useful provided the focus is on pointwise quantities, e.g., quantiles at $s_* \in \mathcal{X}$.

Different type of spatial data

- geostatistical data: data are defined continuously on $\mathcal{X}$, e.g., rainfall;
- punctual data: the data are points falling randomly over some space $\mathcal{X}$, e.g., tree locations.
- lattice data: data are aggregated over sub-regions, e.g., number of citizen in counties.

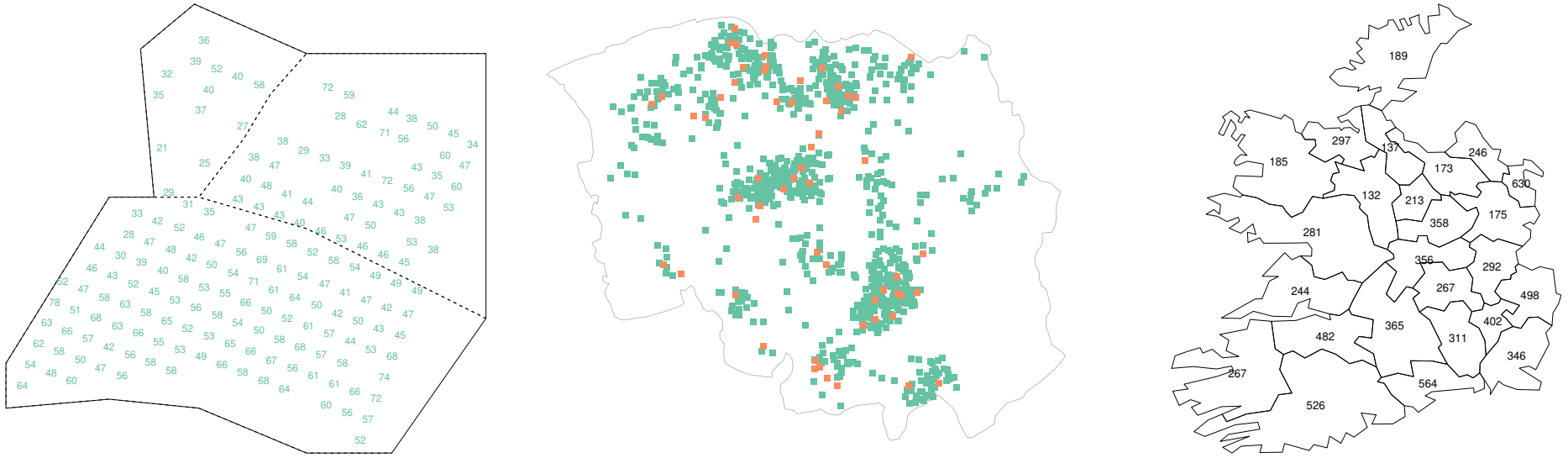
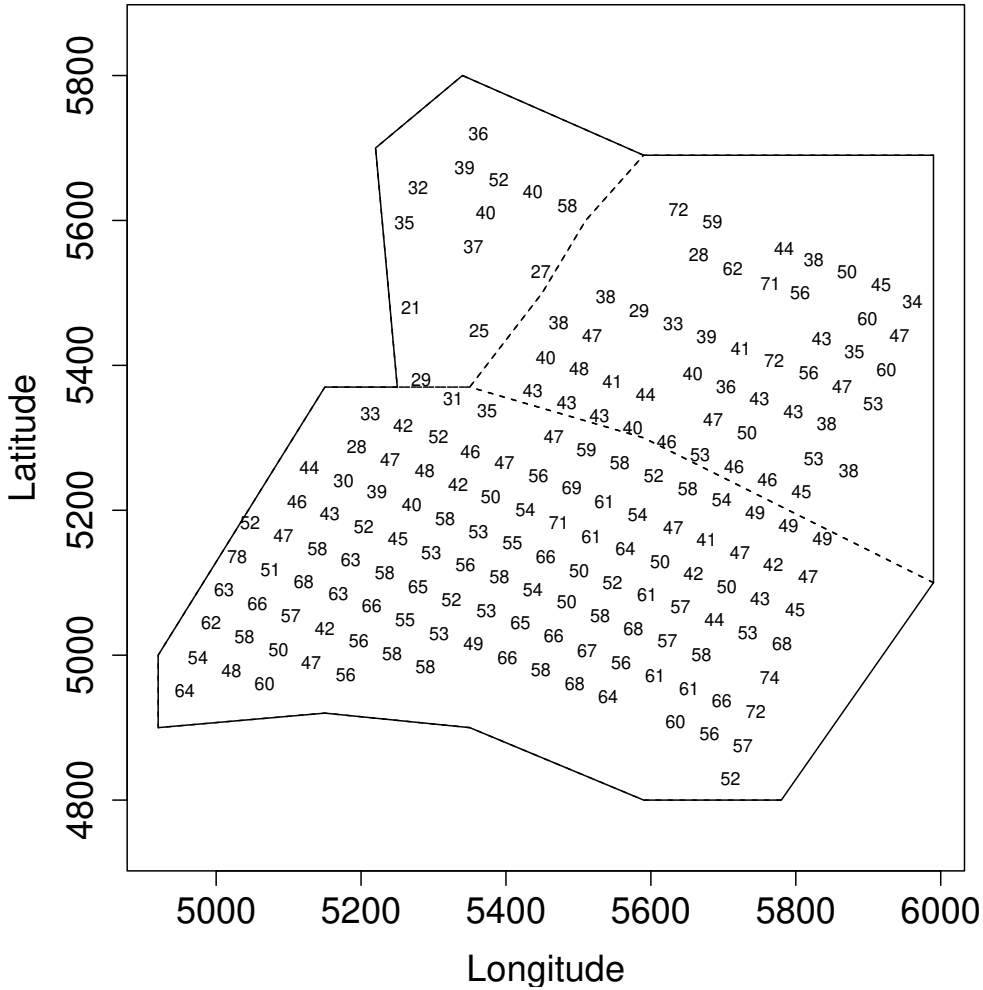


Figure 1: *The three different type of spatial data. From left to right: geostatistical (Calcium concentration), punctual (location of lung and larynx cancer in Chorely–Ribble) and lattice data (town population in the 25 counties of Ireland).*

The Ca_{20} data set



Focus is on geostatiscal data only!

2. Theoretical Framework

Stochastic processes (reminder)

Definition 1. A stochastic process defined on $\mathcal{X}$ is a collection of random variables indexed by $\mathcal{X}$ on the same probability space $(\Omega, \mathcal{F}, \Pr)$.

Proposition 1. *A stochastic process $\{Y(s)\colon s \in \mathcal{X}\}$ is completely characterised from its finite dimensional distribution functions, i.e., for any $k \geq 1$ and $s_1, \ldots, s_k \in \mathcal{X}$*

$$\Pr\{Y(s_1) \leq A_1, \ldots, Y(s_k) \leq A_k\}, \qquad A_1, \ldots, A_k \text{ Borel sets},$$

(provided they satisfy the hypothesis of the Kolmogorov extension theorem, i.e., invariance to permutation and consistent marginalisation)

Strictly stationary processes

Definition 2. A stochastic process $\{Y(s)\colon s \in \mathcal{X}\}$ is said (strictly) stationary if its finite dimensional distribution functions are invariant by translation, i.e., for any $k \geq 1$, $s_1, \ldots, s_k \in \mathcal{X}$ and $h \in \mathcal{X}$ we have

$$\Pr\{Y(s_1+h) \leq A_1, \ldots, Y(s_k+h) \leq A_k\} = \Pr\{Y(s_1) \leq A_1, \ldots, Y(s_k) \leq A_k\},$$

where A_j are Borel sets.

☞ In practice, strict stationarity is too strong and cannot be checked. Need a weaker hypothesis.

Definition 3. A second order stochastic process is a stochastic process whose second order moment exists, i.e., $\text{Var}[Y(s)] < \infty$ for all $s \in \mathcal{X}$.

- Working with second order processes allows to define
  - the mean function / trend / drift

$$\begin{aligned} \mu \colon \mathcal{X} &\longrightarrow \mathbb{R} \\ s &\longmapsto \mathbb{E}[Y(s)], \end{aligned}$$

  - the covariance function

$$\begin{aligned} K \colon \mathcal{X} \times \mathcal{X} &\longrightarrow \mathbb{R} \\ (s, s') &\longmapsto \text{Cov}\{Y(s), Y(s')\}. \end{aligned}$$

Definition 4. A second order process is said weakly stationary, or just stationary, if for any $s, s' \in \mathcal{X}$ and $h \in \mathcal{X}$ we have

$$\mu(s+h) = \mu(s), \qquad K(s+h, s'+h) = K(s, s'). \qquad \texttt{(translation invariance)}$$

Definition 5. A stochastic process $\{Y(s) \colon s \in \mathcal{X}\}$ is said isotropic if for any rotation matrix R, i.e., $|R| = 1$ and $R^{-1} = R^T$, we have

$$\{Y(Rs) \colon s \in \mathcal{X}\} \overset{\mathsf{d}}{=} \{Y(s) \colon s \in \mathcal{X}\}. \qquad \texttt{(rotation invariance)}$$

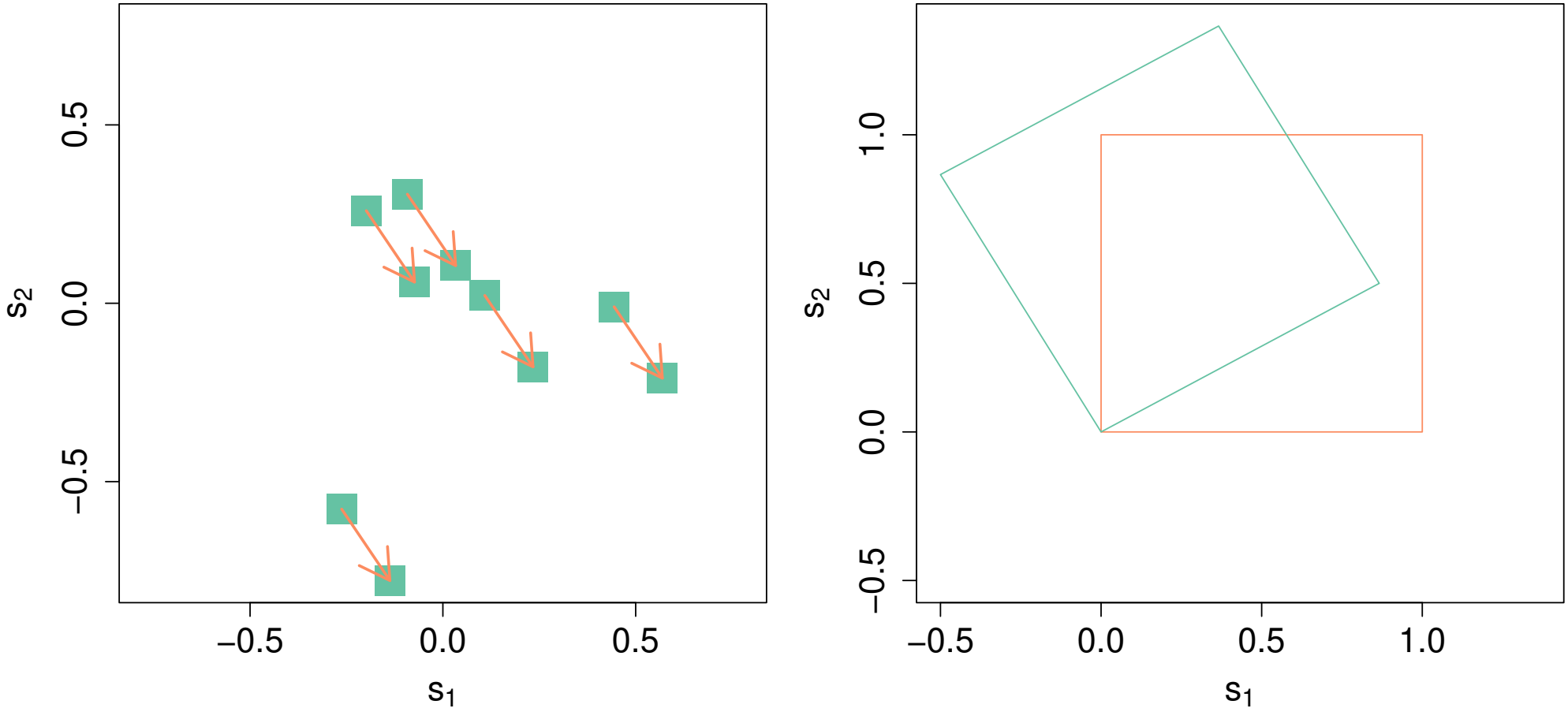


Figure 2: *Illustration of stationarity and isotropy.*

Consequences

- If a process is stationary we have

$$K(s, s') = K(o, s' - s) = K(h),$$

where $h = s - s'$ and is even since

$$\text{Cov}\{Y(s), Y(s')\} = \text{Cov}\{Y(s'), Y(s)\}.$$

- If we further assume isotropy, the covariance function now satisfies

$$\begin{aligned} K(Rh) &= K(h) \\ &= K(\|h\|) \\ &= K(-\|h\|). \end{aligned}$$

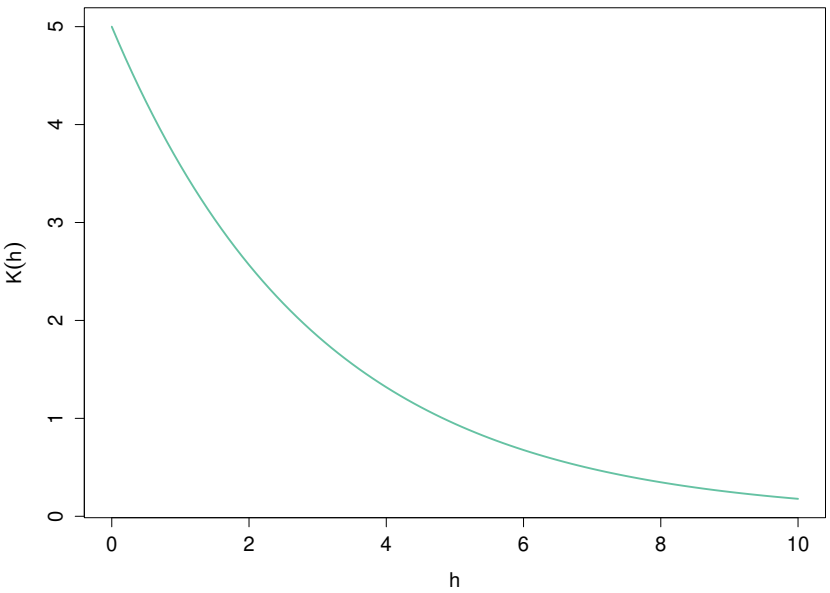


Figure 3: *Plot of a stationary isotropic covariance function. What is $K(0)$?*

Some parametric family of stationary isotropic correlation functions

Exponential

$$\rho(h) = \exp\left(-\frac{h}{\lambda}\right), \qquad \lambda > 0$$

Gaussian

$$\rho(h) = \exp\left\{-\left(\frac{h}{\lambda}\right)^2\right\}, \qquad \lambda > 0$$

Stable // Powered exponential

$$\rho(h) = \exp\left\{-\left(\frac{h}{\lambda}\right)^\kappa\right\}, \qquad 0 < \kappa \leq 2, \quad \lambda > 0$$

Whittle–Matérn

$$\rho(h) = \frac{2^{1-\kappa}}{\Gamma(\kappa)}\left(\frac{u}{\lambda}\right)^\kappa K_\kappa\left(\frac{u}{\lambda}\right), \qquad \kappa > 0, \quad \lambda > 0.$$

- The Whittle-Matèrn family is very flexible:
  - As $\kappa \to \infty$, it converges to the Gaussian family
  - When $\kappa = 1/2$, it is the exponential family
- The shape parameter (as usual) is difficult to estimate
- Some people suggest to considered a bunch of fixed values, e.g., $\kappa = 0.25, 0.5, 0.75, \ldots$.

Properties of the covariance function

Proposition 2. *The covariance function K of a stationnary process satisfies*

- $K(h) \leq K(0)$;
- *For any $k \geq 1$, $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ and $s_1, \ldots, s_k \in \mathcal{X}$ we have*

$$\sum_{i,j=1}^{k} \lambda_i \lambda_j K(s_i - s_j) \geq 0.$$

The last properties states that the covariance function is a (semi) definite positive function.

Proposition 3. *Let K_1 and K_2 be two definite positive functions. Then*

- $aK_1(\cdot) + bK_2(\cdot)$ *is positive definite whenever* $a, b \geq 0$*;*
- $K_1(\cdot)K_2(\cdot)$ *is positive definite.*

Further, if $\{K_n : n \geq 1\}$ *is a sequence of positive definite functions such that* $K_n(s) \to K(s)$ *as* $n \to \infty$ *for all* $s \in \mathcal{X}$*, then* K *is positive definite.*

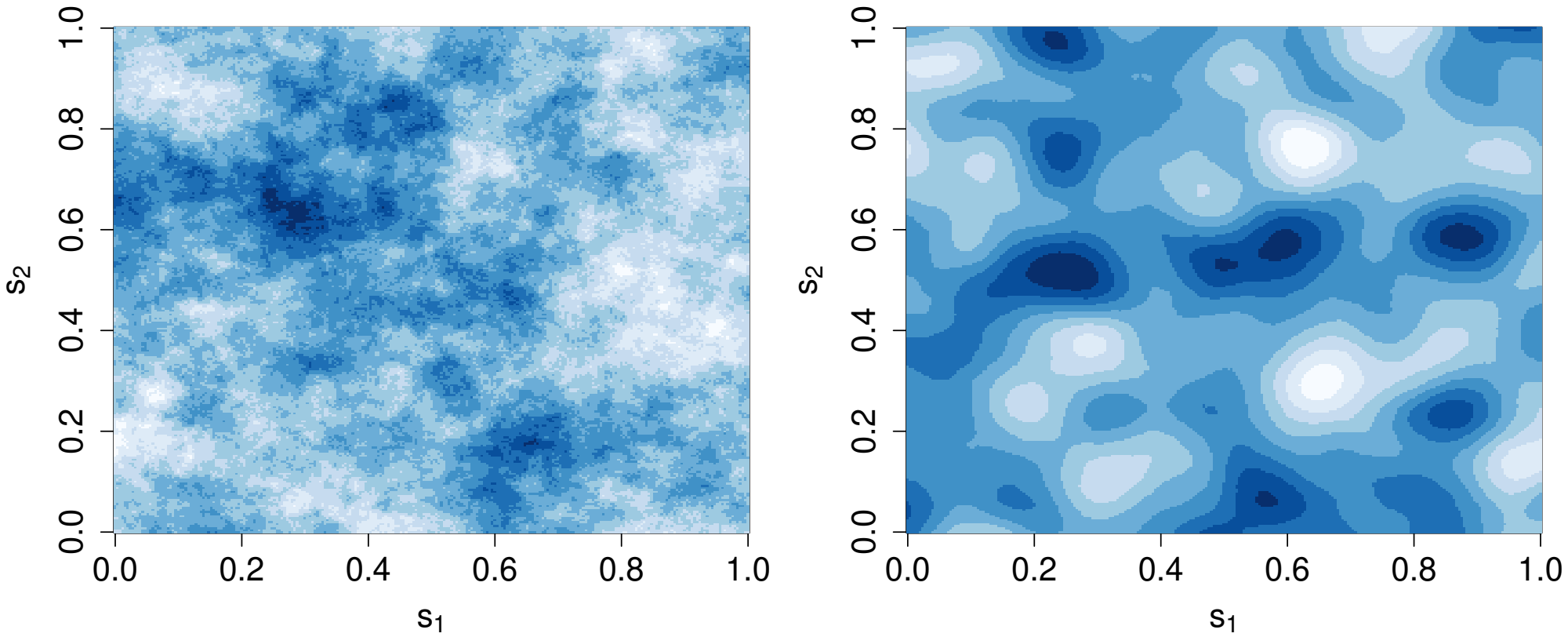
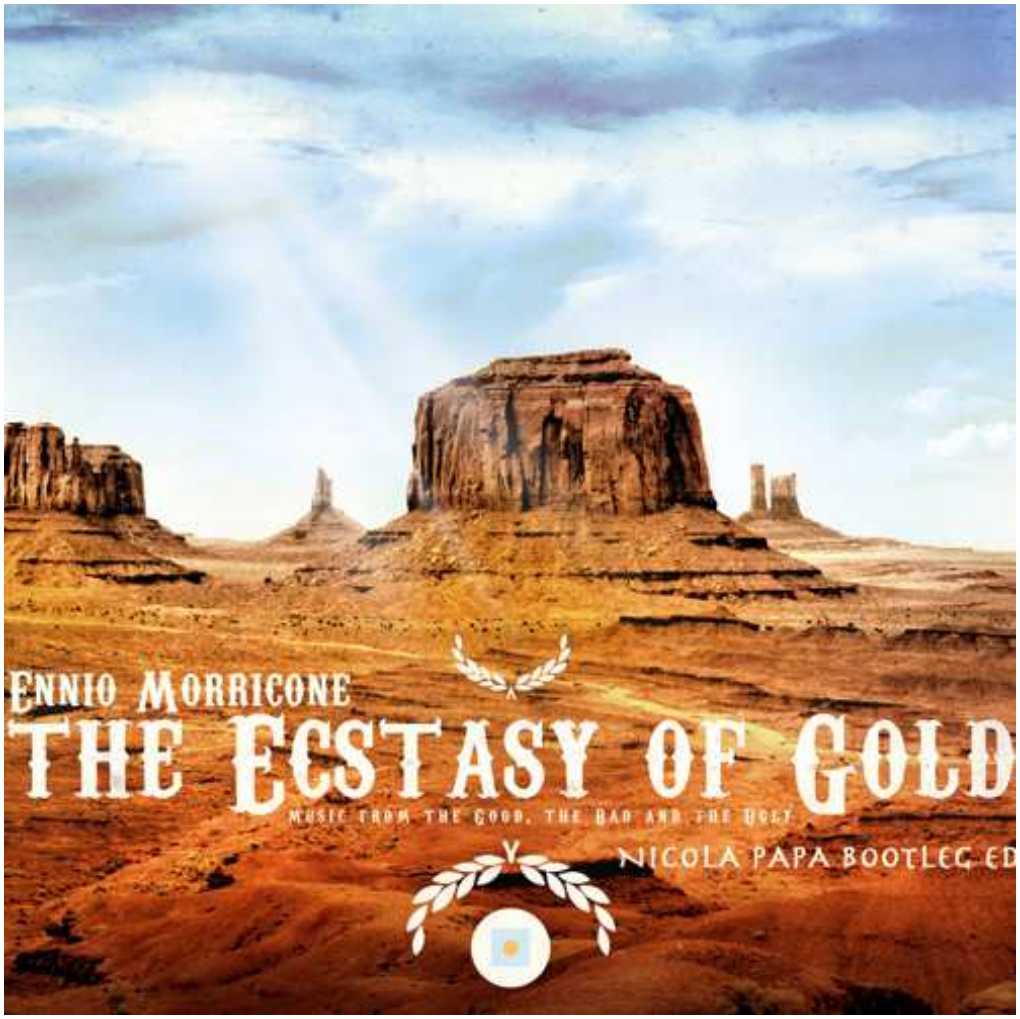


Figure 4: *Two realisations of a random field. (Gaussian random field with a powered exponential correlation function. Left:* $\kappa = 1$*. Right:* $\kappa = 2$*.)*

Proposition 4. *Let K be the covariance function of a stationary process. If K is continuous at the origin, then it is continuous everywhere.*

Corollary 1. *Under the same setting, discontinuity is only possible at the origin, i.e.,*

$$K(h) = c\delta_0(h) + K_c(h).$$


ENNIO MORRICONE
THE ECSTASY OF GOLD
MUSIC FROM THE GOOD, THE BAD AND THE UGLY
NICOLA PAPA BOOTLEG ED

Nugget effect

- This potential discontinuity at the origin is called the nugget effect η, i.e.,

$$K(h) = \begin{cases} \eta + \tau, & h = 0, \\ \tau\rho(h), & h > 0, \end{cases}$$

  where ρ is a correlation function.
- The nugget effect may have two possible interpretations:
  - error in measurements, i.e., $Y(s) = S(s) + \varepsilon(s)$
  - spatial variation on a scale smaller than the minimum distance between measurements (if no replicate)

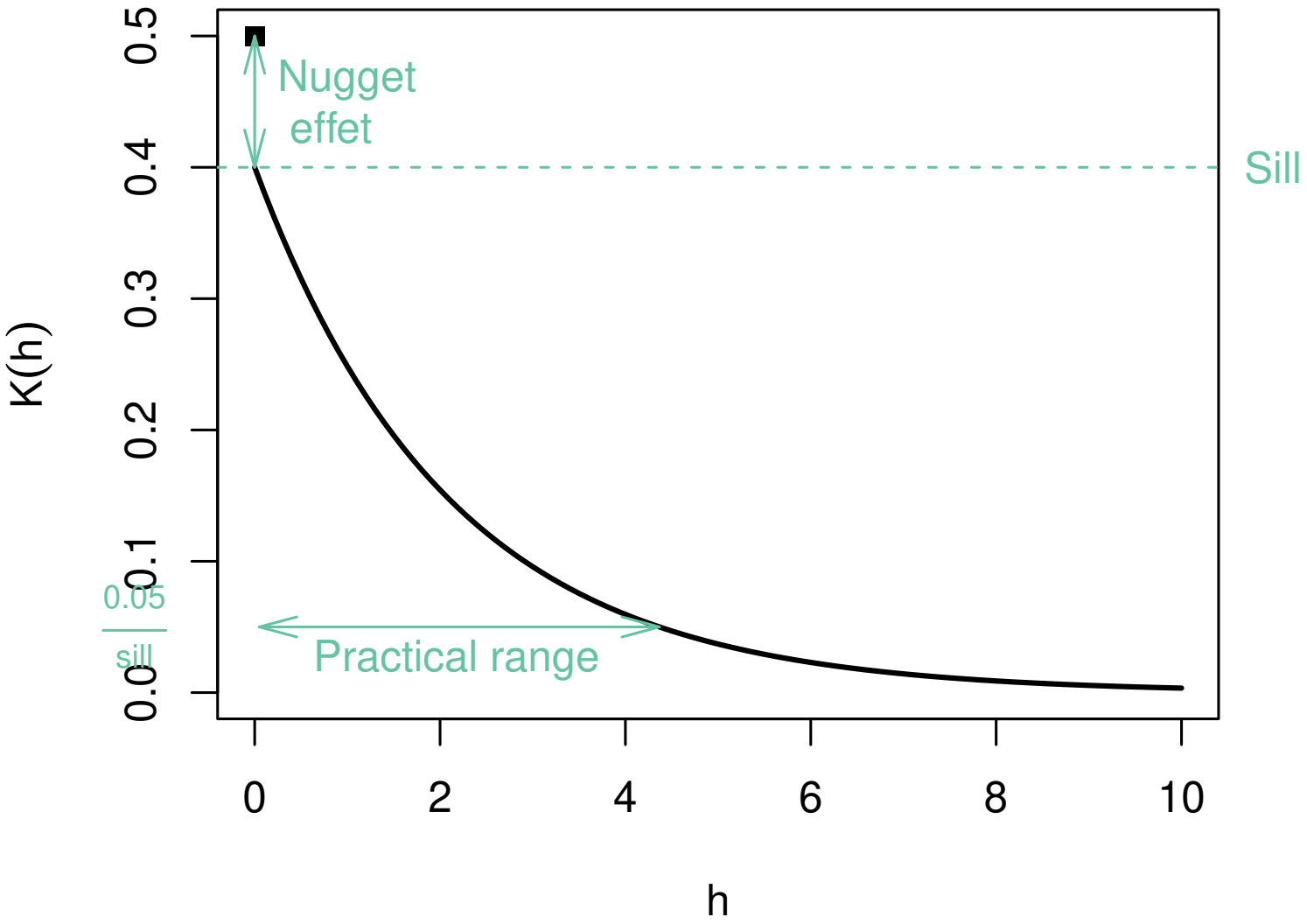


Figure 5: *Illustration of the nugget effect, the sill parameter and the practical range.*

Continuity in L^2

Definition 6. A stationary process $Y(\cdot)$ is said to be quadratic mean continuous if

$$\mathbb{E}\left[\{Y(o) - Y(h)\}^2\right] \longrightarrow 0, \qquad \|h\| \to 0.$$

Theorem 1. *A stationary process is quadratic mean continuous if and only if the covariance function is continuous at the origin.*

Example: L^2 continuous but not a.s. continuous

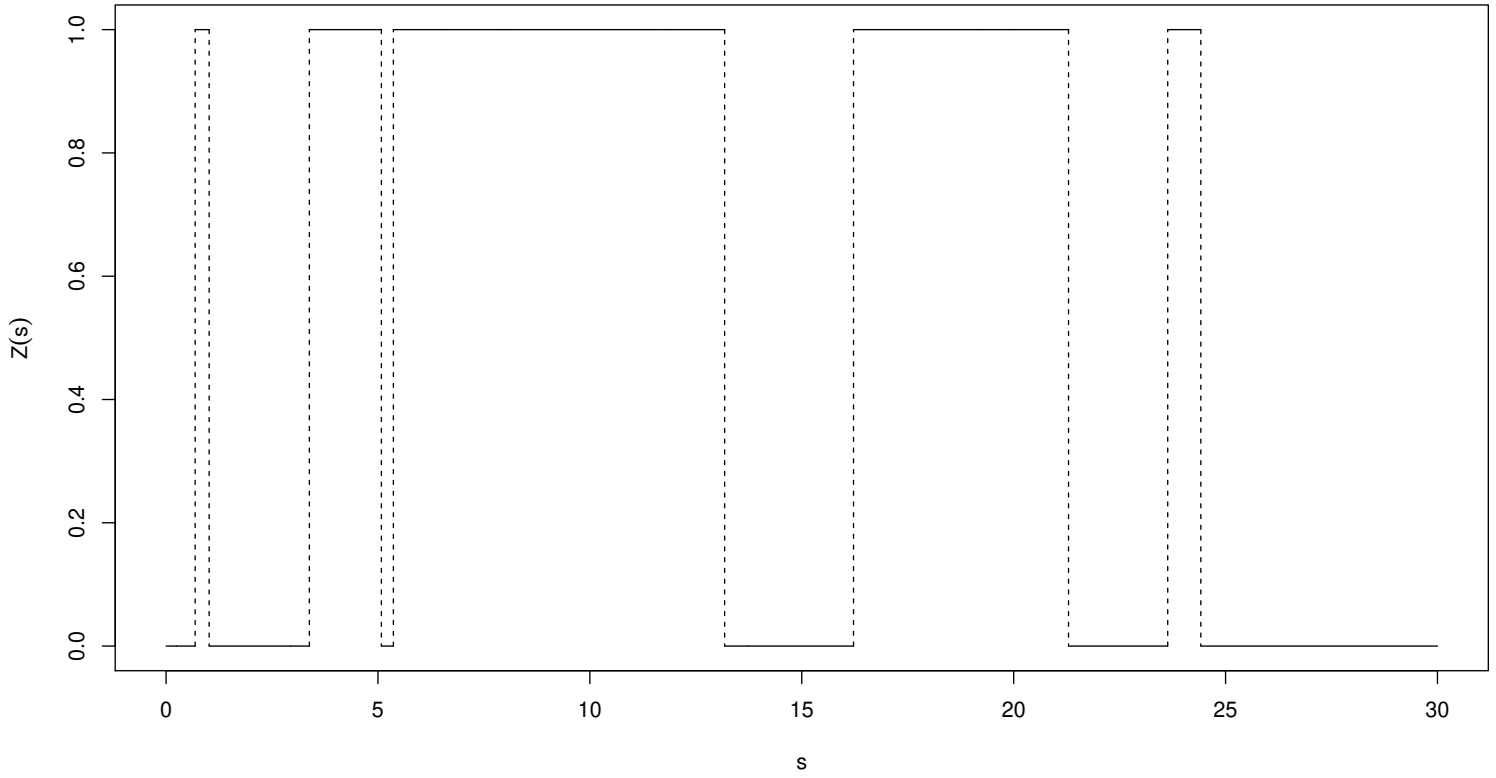


Figure 6: *Realization of a L_2–continuous process whose sample path are not continuous a.s.*

$$Z(s) = X_{I(s)}, \qquad I(s) = \inf\{i \geq 1 \colon \xi_i \geq s\},$$

whith $\{\xi_i \colon i \geq 1\} \sim PPP(1)$ and $X_i \overset{\text{iid}}{\sim} \text{Ber}(p)$.

Mean square differentiability

Definition 7. A stationnary process $Y(\cdot)$ is L^2 differentiable if there exists a second order process $D(\cdot)$ such that

$$\mathbb{E}\left[\left\{\frac{Y(s+h)-Y(s)}{\|h\|}-D(s)\right\}^2\right]\longrightarrow 0, \qquad \|h\|\to 0.$$

The process $D(\cdot)$ is then called the mean square derivative process of $Y(\cdot)$ and will be naturally be written $Y'(\cdot)$.

Proposition 5 (Bartlett, 1955)**.** *A stationnary process is mean square differentiable if and only if γ is twice differentiable at its origin. The mean square derivative process has covariance function $h\mapsto -K''(h)$.*

Processes with stationary increments

Definition 8. A stochastic process $\{Y(s)\colon s \in \mathcal{X}\}$ is said to have stationary increments if for all $h \in \mathcal{X}$, the distribution of the process $\{Y(s+h) - Y(s)\colon s \in \mathcal{X}\}$ is the same as $\{Y(s)\colon s \in \mathcal{X}\}$.

- The motivation for using stationary increments processes is that we are no longer restricted to stationary processes.
- We can even work with non second order processes and simply assume

$$\text{Var}[Y(h) - Y(o)] < \infty.$$

In general we thus have $\mathbb{E}[Y(s+h) - Y(s)] = f(h)$, but it is common practice (and we will do!) to further assume $\mathbb{E}[Y(s+h) - Y(s)] = 0$.

Example 1. Consider the following random walk defined on $\mathcal{X} = \mathbb{Z}$

$$Y(s+1) = Y(s) + \varepsilon_{s+1}, \qquad \varepsilon_j \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

It has indeed stationary increments since

$$Y(s+h) - Y(s) = \sum_{j=0}^{h-1} \underbrace{\{Y(s+h-j) - Y(s+h-j-1)\}}_{=\varepsilon(s+h-j)} = \sum_{j=0}^{h-1} \varepsilon_{s+h-j} \sim N(0, h\sigma^2).$$

but is not stationary. Even worse we have $\text{Var}\{Y(s)\} \to \infty$ as $s \to \infty$.

> Extension of the above random walk to $\mathcal{X} = \mathbb{R}^d$ leads to the so-called Brownian random fields. If we further assume dependence across increments we get fractional Brownian processes.

Semi–variogram

- The covariance function is a summary statistic of the spatial dependence function for at most second order processes.
- To get an analogue for stationary increment processes we rather consider the semi–variogram

$$\gamma(h) = \frac{1}{2}\mathsf{Var}[Y(h) - Y(o)] = \frac{1}{2}\mathbb{E}\left[\{Y(h) - Y(o)\}^2\right].$$

If the process is indeed second order we have

$$\gamma(h) = \frac{1}{2}\{2K(o,o) - 2K(o,h)\} = K(o,o)\{1 - \rho(h)\},$$

where $h \mapsto \rho(h)$ is the correlation function and

$$\gamma(h) \longrightarrow K(o,o), \qquad \|h\| \to \infty, \quad \texttt{(as long as } \rho(h) \to 0\texttt{)}$$

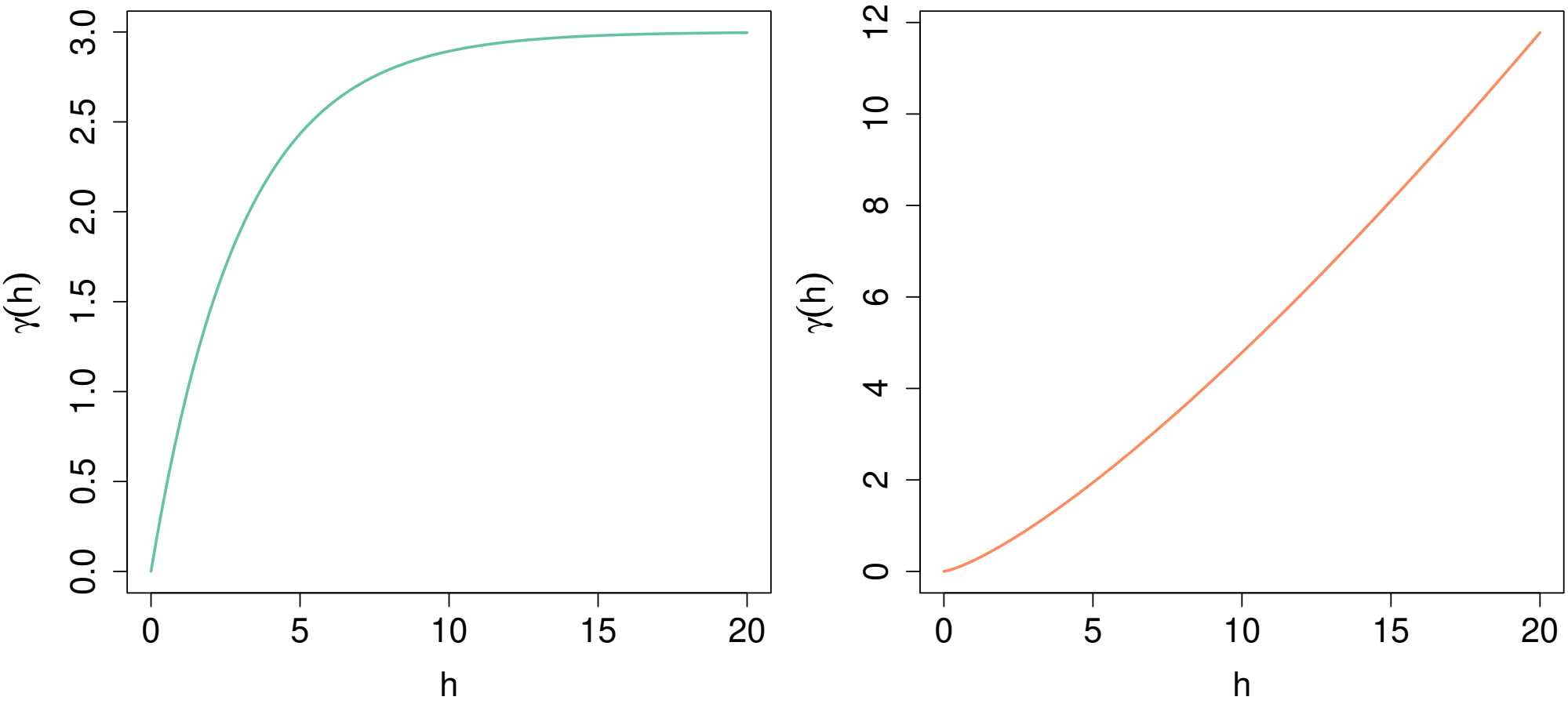


Figure 7: *Bounded (left) and unbounded (right) semi–variograms. If it exists, what is $\gamma(\infty)$?*

Some isotropic stationary correlation functions and variograms

Family	$\rho(h)$	$\gamma(h)$	Support
Exponential	$\exp(-h/\lambda)$	$1-\exp(-h/\lambda)$	$\lambda>0$
Gaussian	$\exp\left\{-(h/\lambda)^2\right\}$	$1-\exp\left\{-(h/\lambda)^2\right\}$	$\lambda>0$
Stable / Powered exponential	$\exp\left\{-(h/\lambda)^\kappa\right\}$	$1-\exp\left\{-(h/\lambda)^\kappa\right\}$	$\lambda>0, 0\leq\kappa\leq 2$
Whittle–Matérn	$\frac{2^{1-\kappa}}{\Gamma(\kappa)}\left(\frac{u}{\lambda}\right)^\kappa K_\kappa\left(\frac{u}{\lambda}\right)$	$1-\frac{2^{1-\kappa}}{\Gamma(\kappa)}\left(\frac{u}{\lambda}\right)^\kappa K_\kappa\left(\frac{u}{\lambda}\right)$	$\lambda>0, \kappa>0$
Fractional	—	$(h/\lambda)^\kappa$	$0\leq\kappa\leq 2$

- The parameters λ and κ are known as the range and smooth parameters.
- Associated covariance functions are derived using a sill parameter τ, i.e.,

$$\gamma(h)=\tau\rho(h), \qquad \tau>0. \qquad (\tau=\mathtt{K}(\mathtt{o}))$$

- The smooth and range parameters drives respectively the smoothness of the random process and the range of spatial dependence.

The practical range h_p is the distance such that $\rho(h_p)=0.05$.

Proposition 6. *Let $Y(\cdot)$ be a process with stationnary increments, $s_1, \ldots, s_k \in \mathcal{X}$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that $\sum_{j=1}^{k} \lambda_j = 0$. Then*

$$\mathbb{E}\left\{\sum_{j=1}^{k} \lambda_j Y(s_j)\right\} = 0$$

$$\mathit{Var}\left\{\sum_{j=1}^{k} \lambda_j Y(s_j)\right\} = -\sum_{i,j=1}^{k} \lambda_i \lambda_j \gamma(s_i - s_j) \geq 0$$

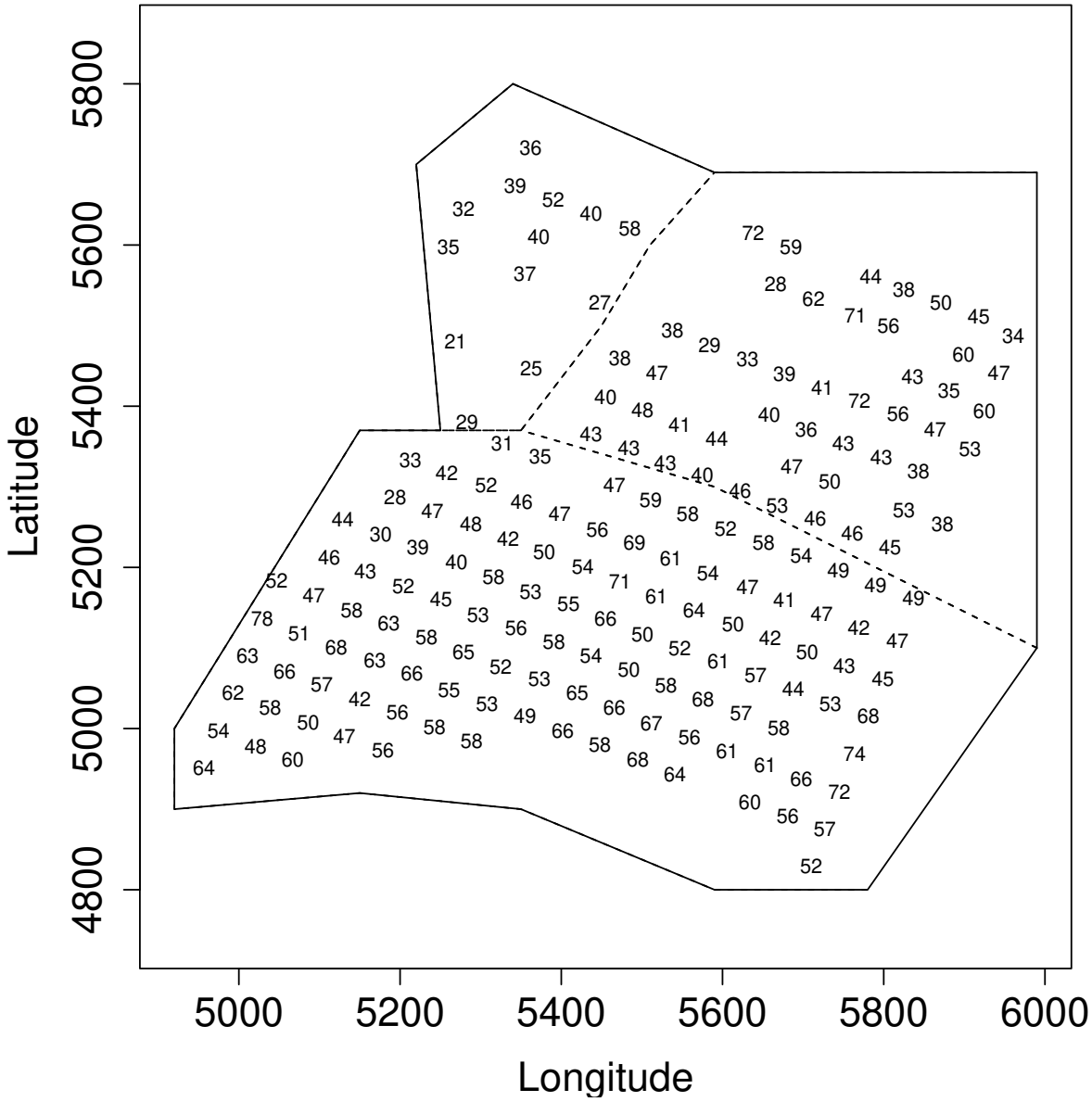
The last property states that the variogram is conditionally definite negative.

3. Kriging

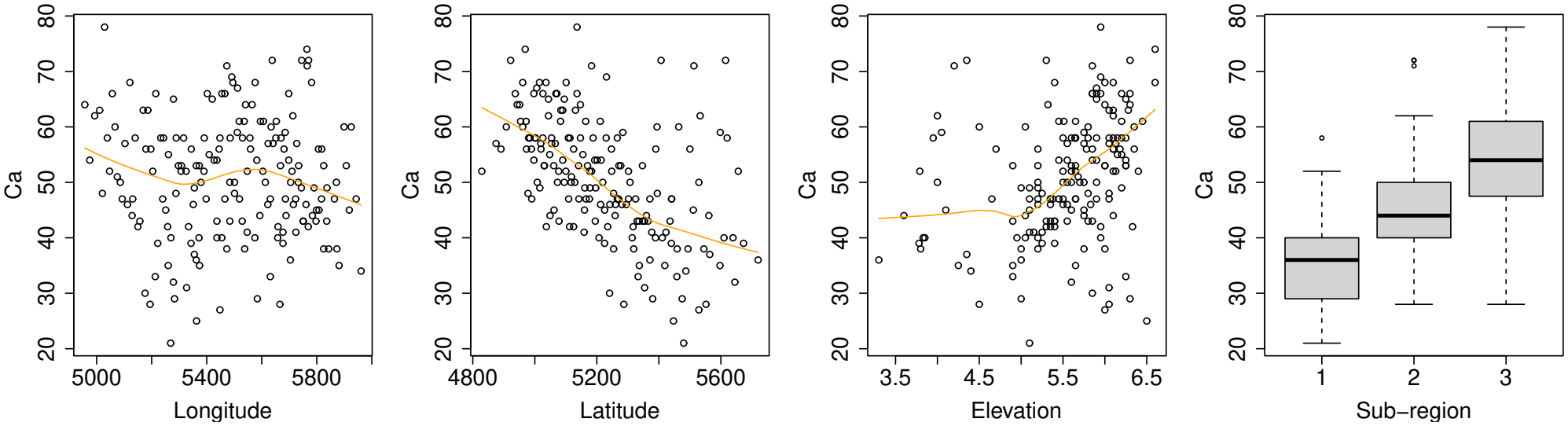
Descriptive analysis

- Before trying to model the data we need to check whether the data can safely be assumed stationary / isotropic / ...
- Essentially we start with a descriptive analysis which, four our context, consists in
  - checking for any trend in the mean function $s \mapsto \mu(s)$
  - inspecting the semi–variogram.
- The first stage is very simple. Just plot data w.r.t. some covariates, e.g., longitude, latitude, ...

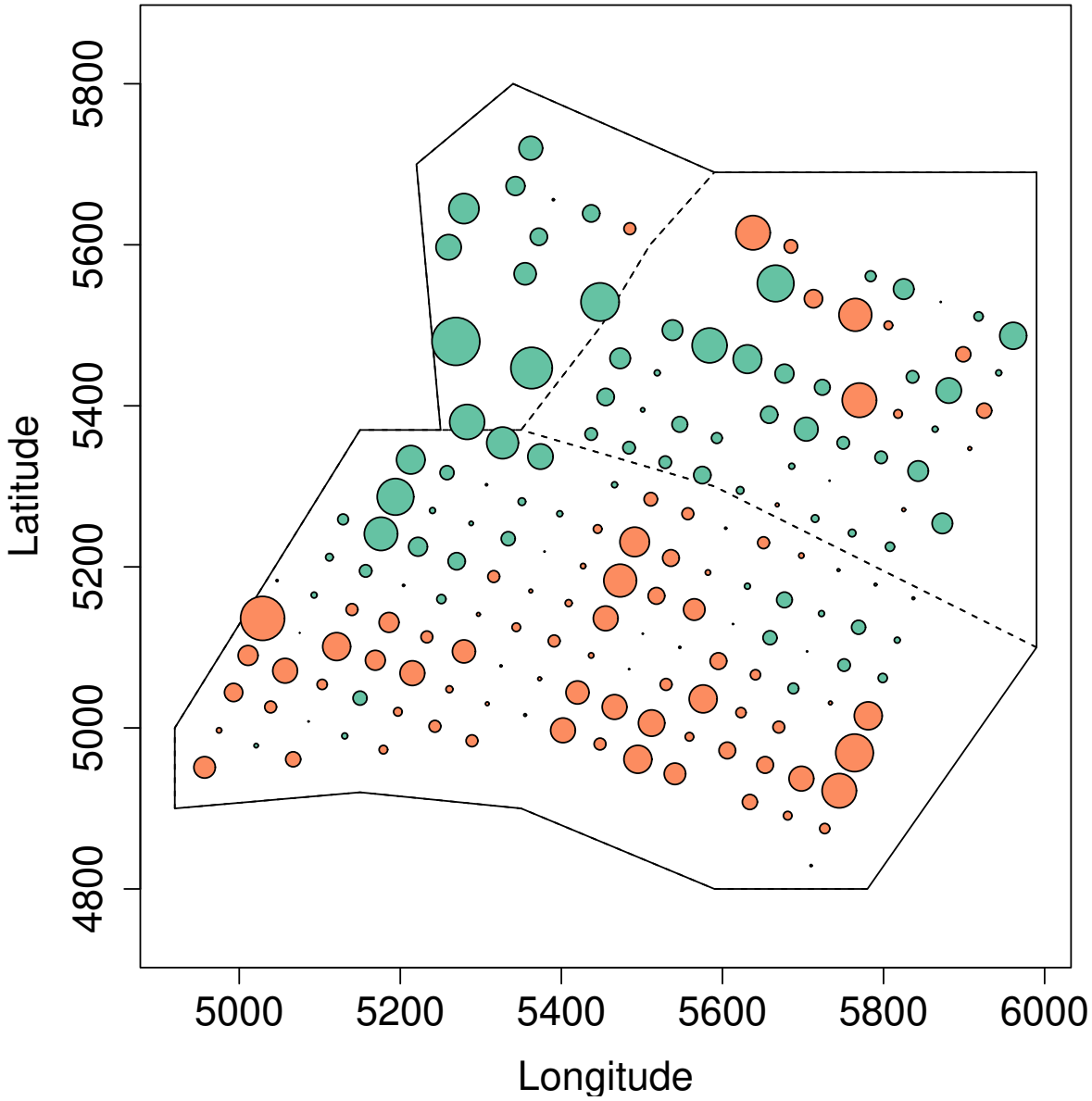
Ca20 data set.



Ca20 data set.



Ca20 data set.



Empirical variograms

- Given some data $\mathcal{D}_n = \{Y_i(s_j) : i = 1, \ldots, n,\ j = 1, \ldots, k\}$, we easily estimate the semi–variogram

$$\hat{\gamma}(h_{j,\ell}) = \frac{1}{2n} \sum_{i=1}^{n} \{Y_i(s_j) - Y_i(s_\ell)\}^2, \qquad h_{j,\ell} = \|s_j - s_\ell\|.$$

- We may have $n = 1$ so that the above estimator has huge variance and we rather use a binned version, i.e.,

$$\tilde{\gamma}(h_b) = \frac{1}{2|B_b|} \sum_{i=1}^{n} \sum_{j,\ell=1}^{k} \{Y_i(s_j) - Y_i(s_\ell)\}^2 1_{\{\|s_j - s_\ell\| \in B_b\}},$$

where $\{B_b : b = 1, \ldots, B\}$ is a partition of $(0, \max h_{j,\ell})$ and h_b is the centroid of B_b.

The binned estimator is however biased but has a (much) lower variance.

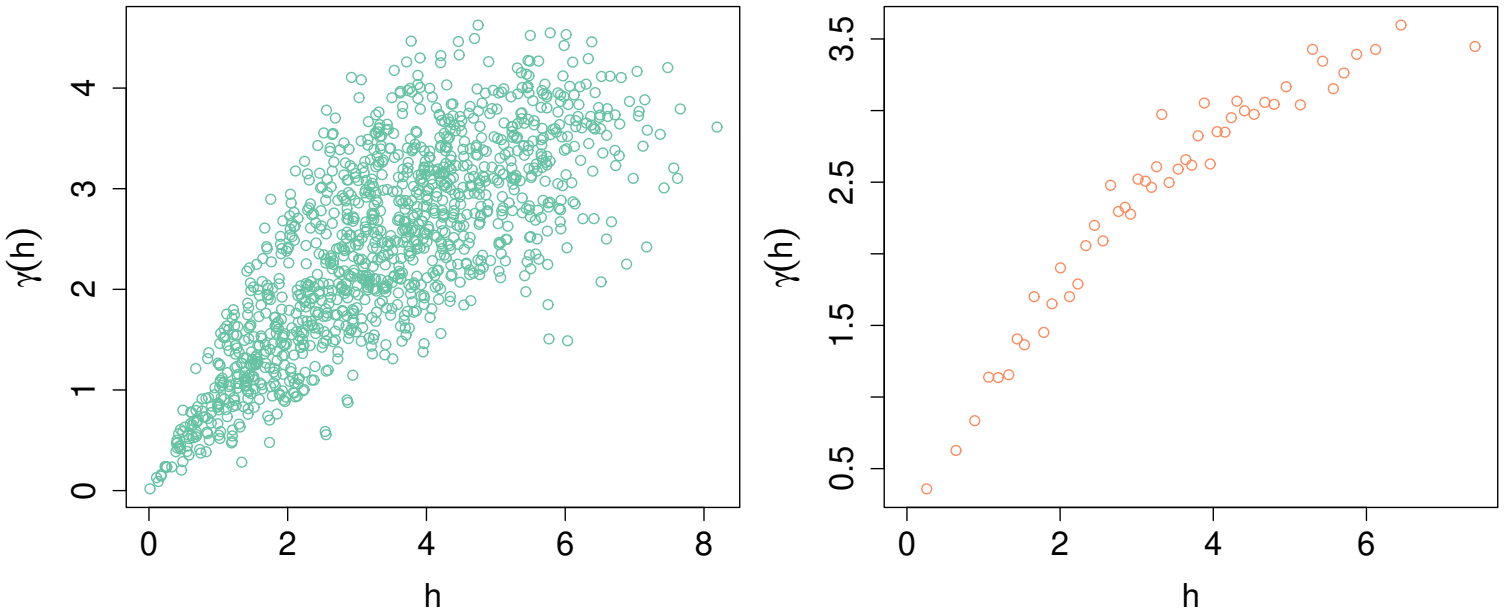


Figure 8: *Empirical variograms. Left: raw. Right: binned.*

The above estimator makes sense only if your data can be considered as stationary or at least with stationary increments.

You may want to remove any possible trends (using a linear model for instance) and estimate the variogram on the residuals.

Least square fitting of a variogram

- Suppose we have fitted a mean function, e.g., from linear models.
- We can fit any parametric variogram $\gamma(\cdot;\psi)$ minimizing using the (weighted) least square estimator on the empirical variogram $\hat{\gamma}$, i.e.,

$$\hat{\psi} = \underset{\psi\in\Psi}{\arg\min} \sum_{j,\ell} \omega_{j,\ell} \left\{\hat{\gamma}(h_{j,\ell}) - \gamma(h_{j,\ell};\psi)\right\}^2 .$$

- The two fitted quantities are all we need to enable predictions!

Nasty optimization problem: use several initial values!
Often fix the smooth parameter to some values, e.g., $\kappa = 0.25, 0.5, \ldots, 2$.
Always question yourself if a nugget effect makes sense.

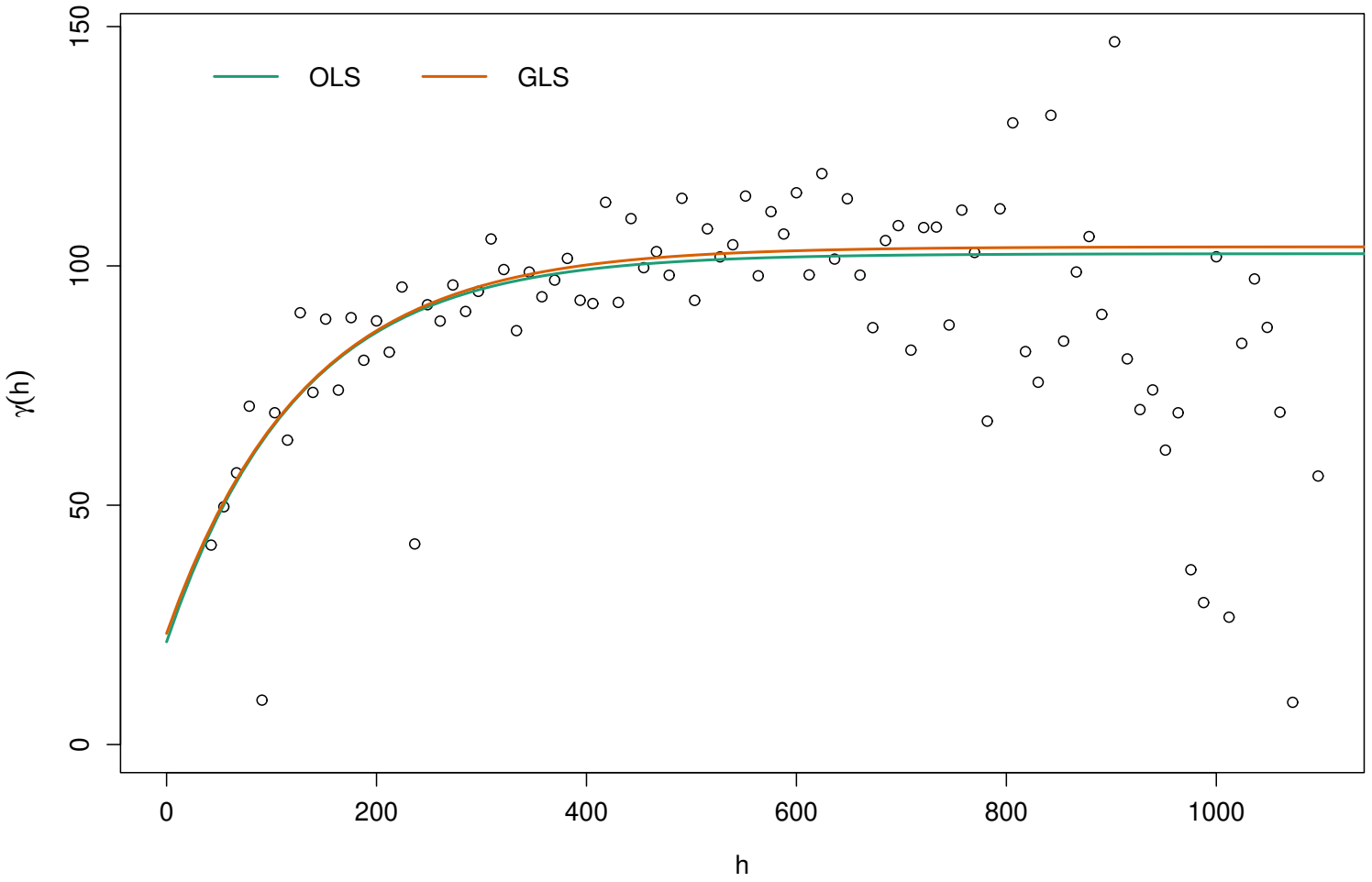
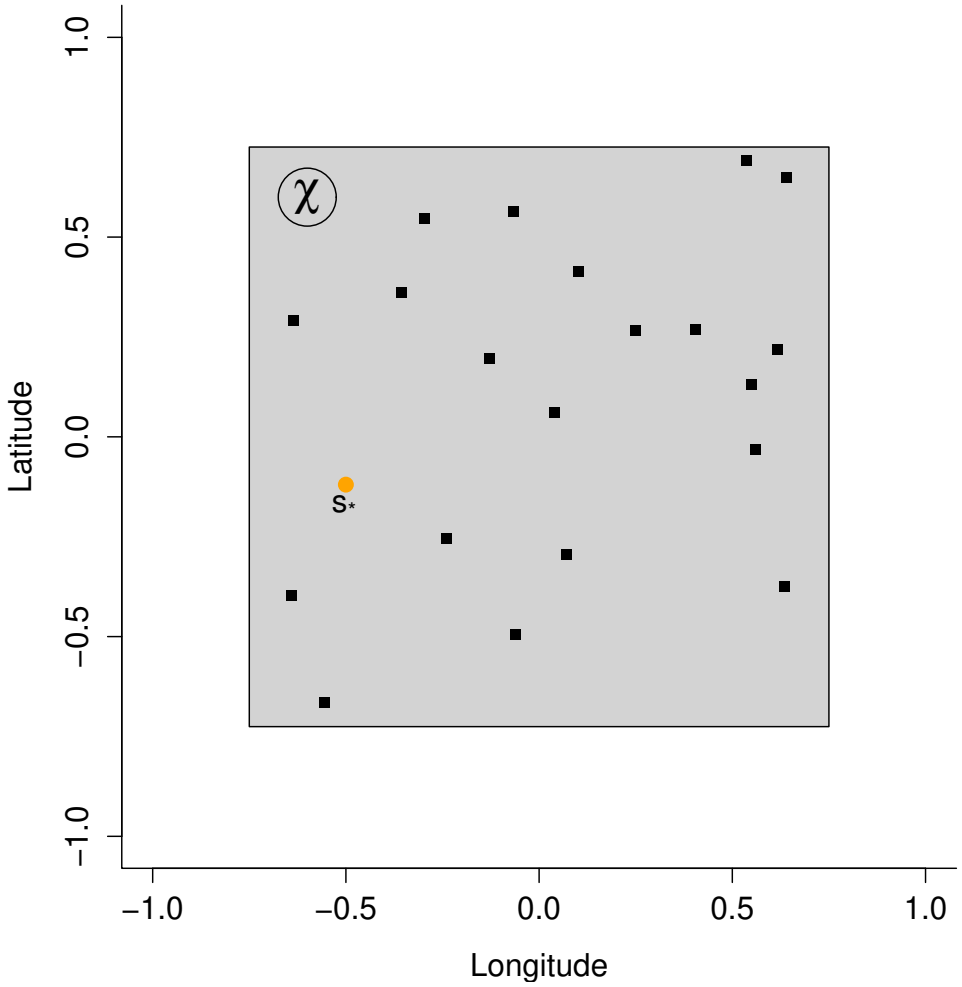


Figure 9: *Least square fitting of a parametric variogram on the Calcium data set.*



- Prediction of $Y(s_*)$ based on observations $Y(s_1), \ldots, Y(s_k)$.
- Restriction to unbiased linear estimators, i.e.,

$$\hat{Y}(s_*) = \sum_{j=1}^{k} \lambda_j Y(s_j),$$

with $\mathbb{E}[\hat{Y}(s_*)] = \mu(s_*)$.

- Estimator is the one minimizing the mean squared error, i.e.,

$$\hat{Y}(s_*) = \arg\min_{T} \mathbb{E}\left[\{T - Y(s_*)\}^2\right].$$

(Universal) Kriging

- There are several type of kriging:

  Simple $\mu(s) \equiv 0$
  Ordinary $\mu(s) = m$, m unknown parameter
  Universal $\mu(s) = \mathbf{x}(s)^\top \boldsymbol{\beta}$,$\boldsymbol{\beta}$ unknown parameter, $\mathbf{x}(s)$ vector of covariates,
  Co-kriging Y is multivariate

  and their intrinsic counterpart.
- Explicit expressions for $\hat{Y}(s_*)$ are available but not given here (nasty).
- We can also get expression for the kriging variance, i.e.,

$$\text{Var}\left\{\hat{Y}(s_*) - Y(s_*)\right\}.$$

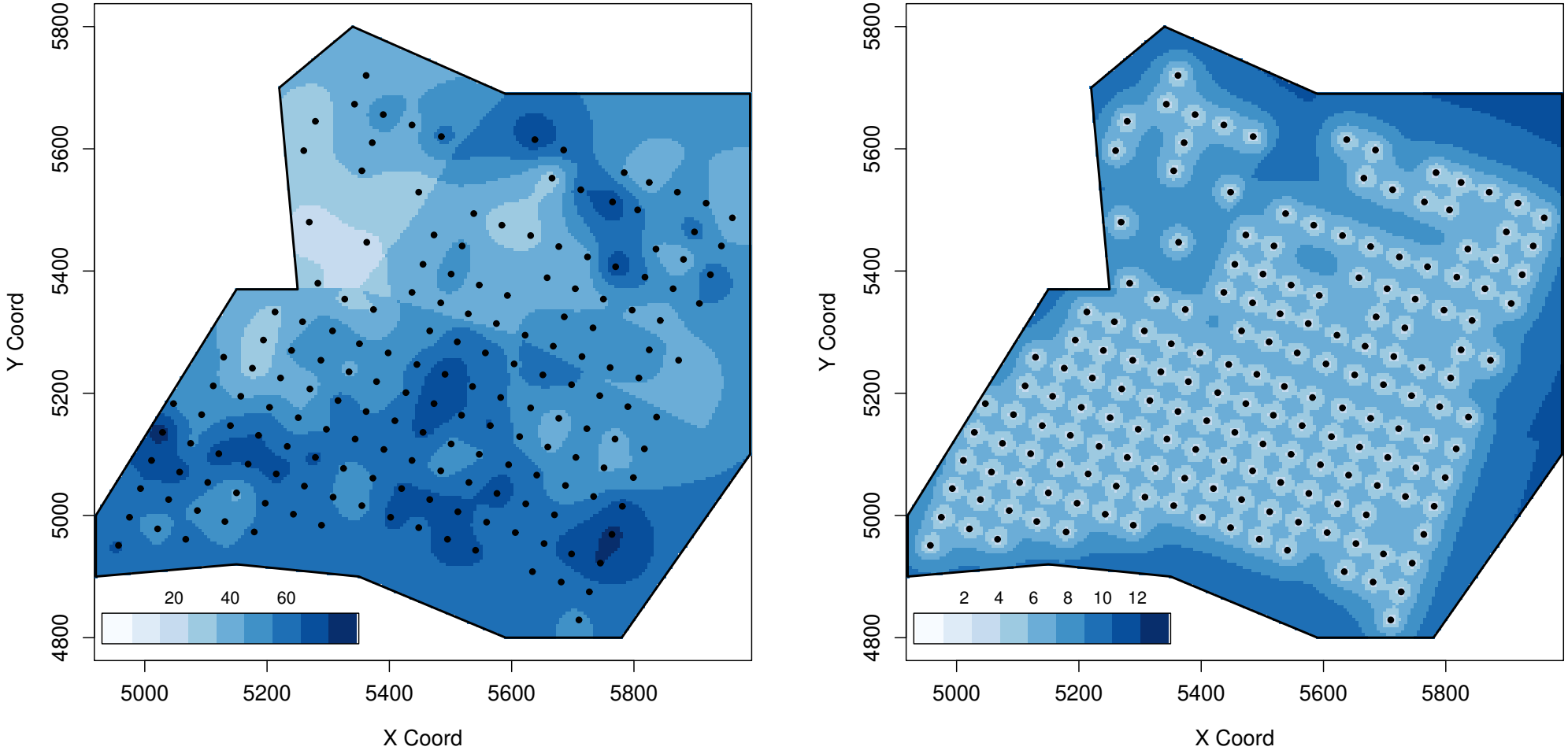


Figure 10: *Kriging estimator (left) and kriging standard error (right) for the* Ca_{20} *data set.*

Exercise 1. Let $\{Y(s)\colon s \in \mathcal{X}\}$ be a stationnary process with (fitted) variogram γ. We have observed a realization of this process at locations $s_1, \ldots, s_k \in \mathcal{X}$ and get $Y(s_j) = y_j$, $j = 1, \ldots, k$. We consider the kriging estimator at $s_* \in \mathcal{X}$

$$\hat{Y}(s_*) = \sum_{j=1}^{k} \lambda_j Y(s_j).$$

1. Find the expression of the simple kriging estimator?
2. What about the kriging error variance, i.e., Var$\{\hat{Y}(s_*) - Y(s)\}$?
3. Find the conditions on the λ_j's to ensure that the ordinary kriging estimator is indeed unbiased.

Change of support

- Our focus is now to estimate the following areal quantity

$$Y(S) = \frac{1}{|S|} \int_S Y(s) \mathrm{d}s, \qquad S \subset \mathcal{X}, \tag{1}$$

  where $|S|$ is the volume of the subset S.
- One can simply use a plugin-estimator based on the kriging estimator, i.e.,

$$\hat{Y}_{\text{plugin}}(S) = \frac{1}{|S|} \int_S \hat{Y}(s) \mathrm{d}s, \qquad \hat{Y}(s) \text{ kriging estimator at } s$$

- However the integral would still have to be discretized

Quantity (1) makes senses for additive variables.

Block kriging

Exercise 2. Let $\{Y(s)\colon s \in \mathcal{X}\}$ be a stationnary random field, $s_0 \in \mathcal{X}$ and $S \subset \mathcal{X}$.

1. Show that

$$K(s_o, S) := \mathsf{Cov}\{Y(s_o), Y(S)\} = \frac{1}{|S|}\int_S K(s_0 - s)\mathsf{d}s.$$

2. Give the ordinary kriging estimator for $Y(S)$.
3. Give an expression for the kriging variance.
4. How would you implement it in practice?

4. Model based geostatistics

Definition 9. The multivariate Gaussian distribution defined on $\mathbb{R}^d$, $d \geq 1$, has probability density function

$$f(\mathbf{y}) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}-\mu)^\top\Sigma^{-1}(\mathbf{y}-\mu)\right\}, \qquad \mathbf{y} \in \mathbb{R}^d, \tag{2}$$

where $\mu \in \mathbb{R}^d$ is the mean vector and $\Sigma \in M_d(\mathbb{R})$ is the covariance matrix.

The Mahalanobis distance is given by

$$a^2(\mathbf{y}) = (\mathbf{y}-\mu)^\top\Sigma^{-1}(\mathbf{y}-\mu)$$

Definition 10. A Gaussian process $\{Y(s) : s \in \mathcal{X}\}$ is a stochastic process whose finite dimensional distribution functions are multivariate Gaussian.

Proposition 7. *A Gaussian process is completely characterized through its mean function and covariance function.*

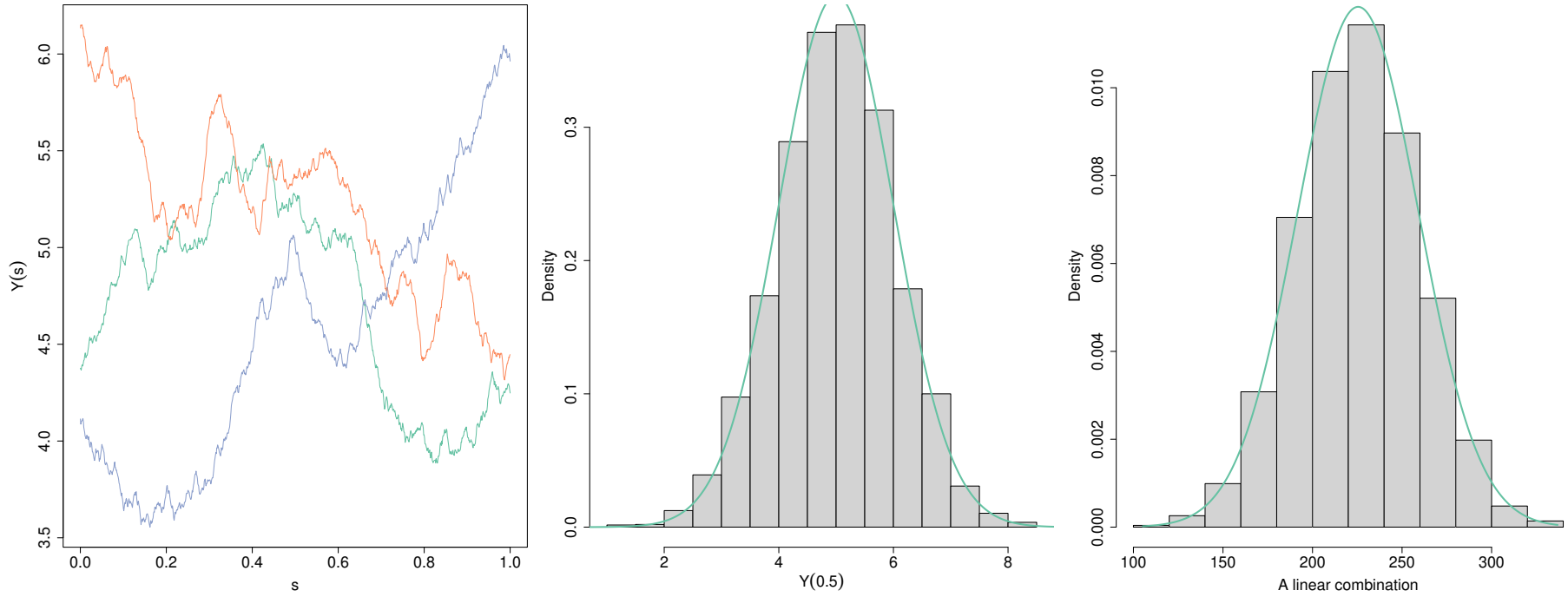


Figure 11: *Numerical illustration of a Gaussian process.*

(semi) Definite positive functions

Definition 11. A function $f : \mathbf{s} \in \mathbb{R}^d \mapsto f(\mathbf{s})$ is said to be (semi) definite positive if it is symmetric and

$$\boldsymbol{\lambda}^\top A \boldsymbol{\lambda} > 0, \qquad A = (a_{i,j} = f(s_i - s_j) : i, j = 1, \ldots, d), \quad x_1, \ldots, x_p \in \mathbb{R}^d,$$

for any non–zero vector $\boldsymbol{\lambda} \in \mathbb{R}^p$. It is semi definite positive if the above inequality is not strict.

The covariance function γ is (semi) definite positive to ensure that the Mahalanobis distance

$$a^2(\mathbf{s}, \mathbf{y}) = (\mathbf{y} - \mu)^\top \Sigma^{-1}(\mathbf{s})(\mathbf{y} - \mu), \qquad \Sigma(\mathbf{s}) = \{\sigma_{i,j} = \gamma(s_i, s_j)\}$$

is always positive and the multivariate Gaussian density is properly defined.

Fitting a Gaussian process

- Having observed n independent observations at k spatial locations, i.e., $\mathcal{D}_n = \{y_i(s_j) \colon i = 1, \ldots, n, j = 1, \ldots, k\}$ $s_1, \ldots, s_k$, we define the log–likelihood as

$$\ell(\mu, \gamma; \mathcal{D}_n) = -\frac{nd}{2} \log(2\pi) - \frac{n}{2} \log |\Sigma(\mathbf{s})| - \frac{1}{2} \sum_{i=1}^{n} a^2(\mathbf{s}, \mathbf{y}_i).$$

💔 no likelihood theory here, but Gaussian processes can (easily) be estimated by maximizing the log–likelihood.

Parametric assumptions

- The above likelihood has some flaws:
  - it has $d + d(d+1)/2$ parameters to estimate which is typically too large;
  - cannot enable prediction at a new location s_* since both mean and covariance function cannot be computed at s_*.
- Hence we further place some parametric structures on
  - the mean function $\mu(s)$, e.g.,

$$\mu(s; \boldsymbol{\beta}) = \mathbf{x}(s)^\top \boldsymbol{\beta},$$

    where $\mathbf{x}(s)$ is a vector of additional covariates at s and $\boldsymbol{\beta}$ a parameter vector to be estimated.
  - the covariance function $\gamma(s, s') = \gamma(s, s'; \psi)$ using some prescribed parametric expressions as the ones presented earlier.

Non isotropic/stationary covariance functions

- Defining non isotropic / stationary covariance functions is a current research field and is far from being trivial.
- A quick and dirty way to get non isotropic covariance functions is to use any isotropic correlation function on a transformed space $\mathcal{X}'$ given by

$$\begin{aligned} \phi\colon \mathcal{X} &\longrightarrow \mathcal{X}' \\ s &\longmapsto \phi(s;\kappa), \end{aligned}$$

  for some prescribed parametric one–one mapping $\phi(\cdot;\kappa)$.
- A specific case, known as geometric anisotropy, is to set

$$\phi(s;\kappa) = C(\kappa)s, \qquad C(\kappa) = \begin{bmatrix} \cos\kappa_1 & -\sin\kappa_1 \\ \sin\kappa_1 & \cos\kappa_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \kappa_2^{-1} \end{bmatrix},$$

  κ_1, κ_2 are respectively the anisotropy angle and ratio.

Interpolation

- As usual the best estimator we can reached (in a L^2 sense) is the conditional expectation, i.e.,

$$\hat{Y}(s_*) = \mathbb{E}\left\{Y(s_*) \mid Y(s_1), \ldots, Y(s_k)\right\}.$$

- For the Gaussian case, the conditional expectation is linear in the $Y(s_j)$.
- Hence the above estimator is actually the kriging estimator!

🔊 You will sometimes hear: "kriging is the optimal estimator" in a L^2 sense. It is wrong unless if we assume Gaussian. However it is indeed optimal if we restrict to linear unbiased estimators.

By the way...

- What if my data are not Gaussian, e.g., rainfall amount.
- A quick and dirty way is to work on a transformation of your data, e.g., $\log Y(s)$, so that Gaussian is a sensible choice.
- One widely used choice for positive variable is the Box–Cox transformation

$$y \longmapsto \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log y, & \lambda = 0. \end{cases}$$

- However it implicitly assumes that the data are stationary so you need to remove any trend first to estimate the shape parameter λ in the Box–Cox transformation.

5. Simulation

(Unconditional) Simulations

- It is rather straightforward to simulated Gaussian process at a moderate number of locations, e.g., $k \leq 3000$, from the Cholesky decomposition of the covariance matrix.
- More precisely for any $\mathbf{s} = (s_1, \ldots, s_k) \in \mathcal{X}$, we have

$$Y(\mathbf{s}) \overset{\mathsf{d}}{=} \mu(\mathbf{s}) + C(\mathbf{s})^\top \boldsymbol{\varepsilon}, \qquad \Sigma(\mathbf{s}) = C(\mathbf{s})^\top C(\mathbf{s}),$$

where $\boldsymbol{\varepsilon}$ is a vector of k independent standard normal random variables.

More sophisticated techniques, e.g., turning bands, circulant embedding methods, exist to get faster simulations on large (gridded) number of locations.

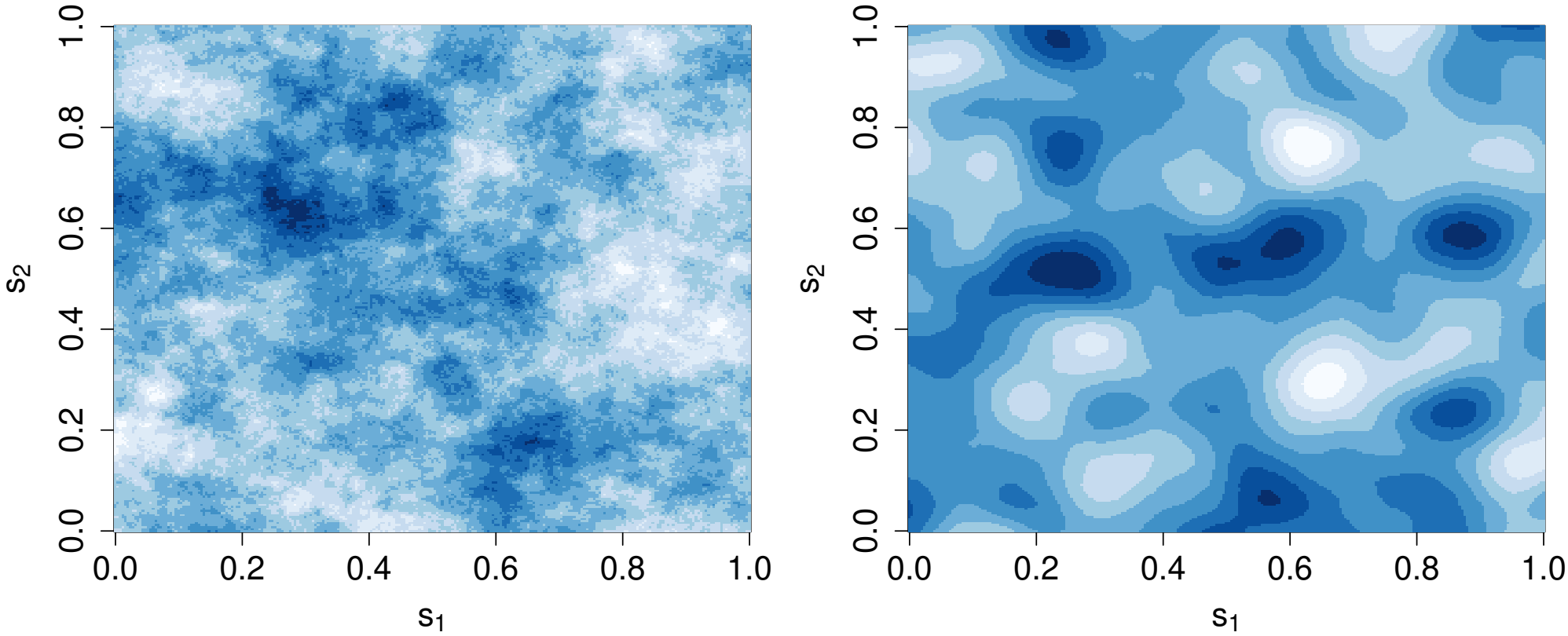


Figure 12: *Two realizations of a random fields with a powered exponential correlation function. Left:* $\kappa = 1$. *Right:* $\kappa = 2$.

Conditional simulations

- Estimating areal quantities from kriging may be too smooth.
- Conditional simulations can be used to get Monte Carlo estimate (and thus the entire distribution) of it.
- Conditional simulations are random simulations that honors some constraints, e.g., simulating from

$$Y(s_*) \mid Y(\mathbf{s}) = \mathbf{y},$$

  where $\mathbf{y}$ is the vector of held fixed values at prescribed location $\mathbf{s}$.
- Under the Gaussian setting, one can use the decomposition

$$Y(s_*) \mid \{Y(\mathbf{s}) = \mathbf{y}\} \overset{\mathsf{d}}{=} \underbrace{Y_{\mathsf{krig}}(s_*)}_{\text{kriging of } Y} + \tilde{Y}(s_*) - \underbrace{\tilde{Y}_{\mathsf{krig}}(s_*)}_{\text{kriging of } \tilde{Y}},$$

  where $\tilde{Y}$ is an independent copy of Y.

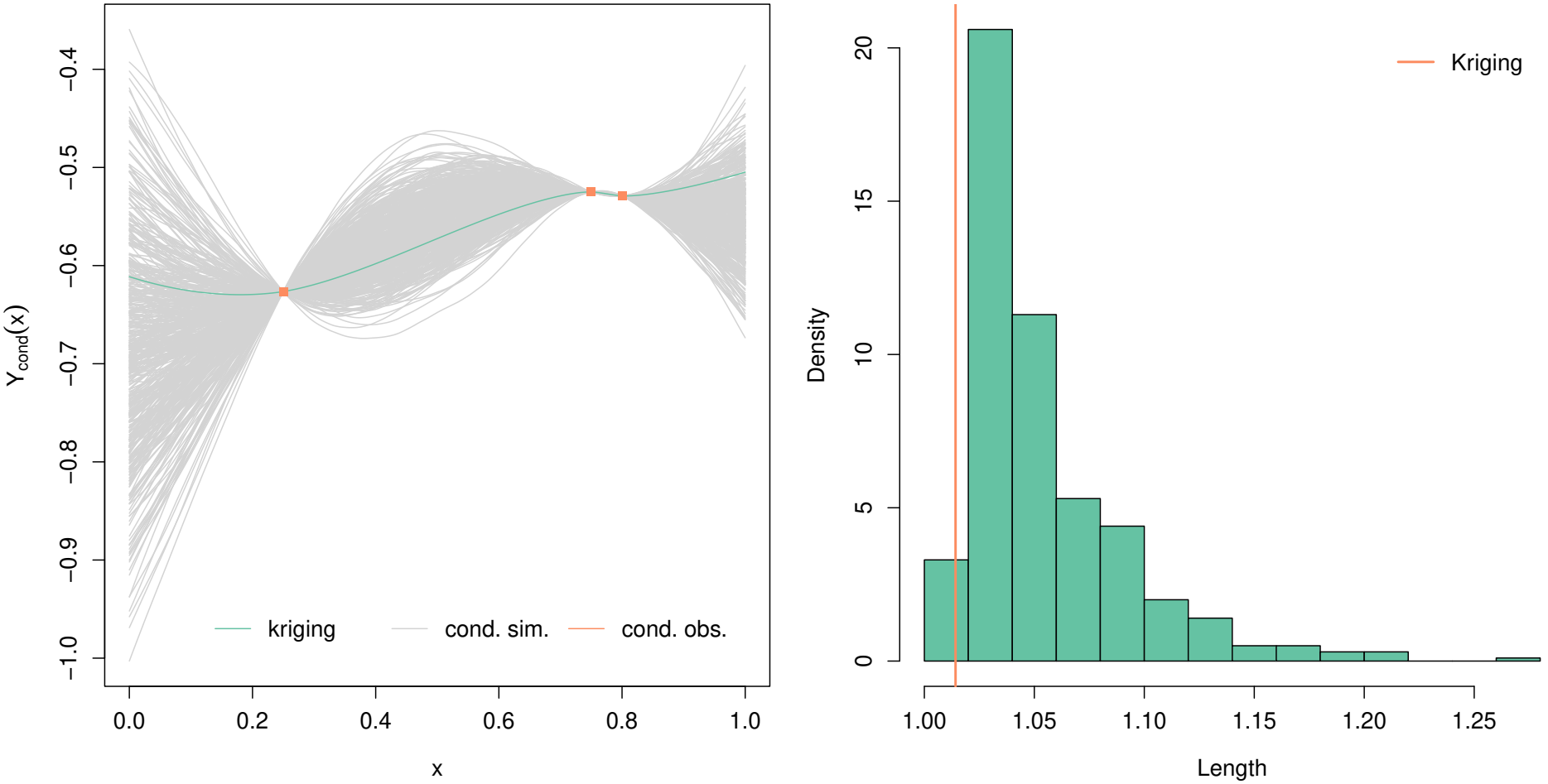


Figure 13: *Comparison between conditional simulations and kriging. Right: length of the curve estimated from kriging and conditional simulations.*

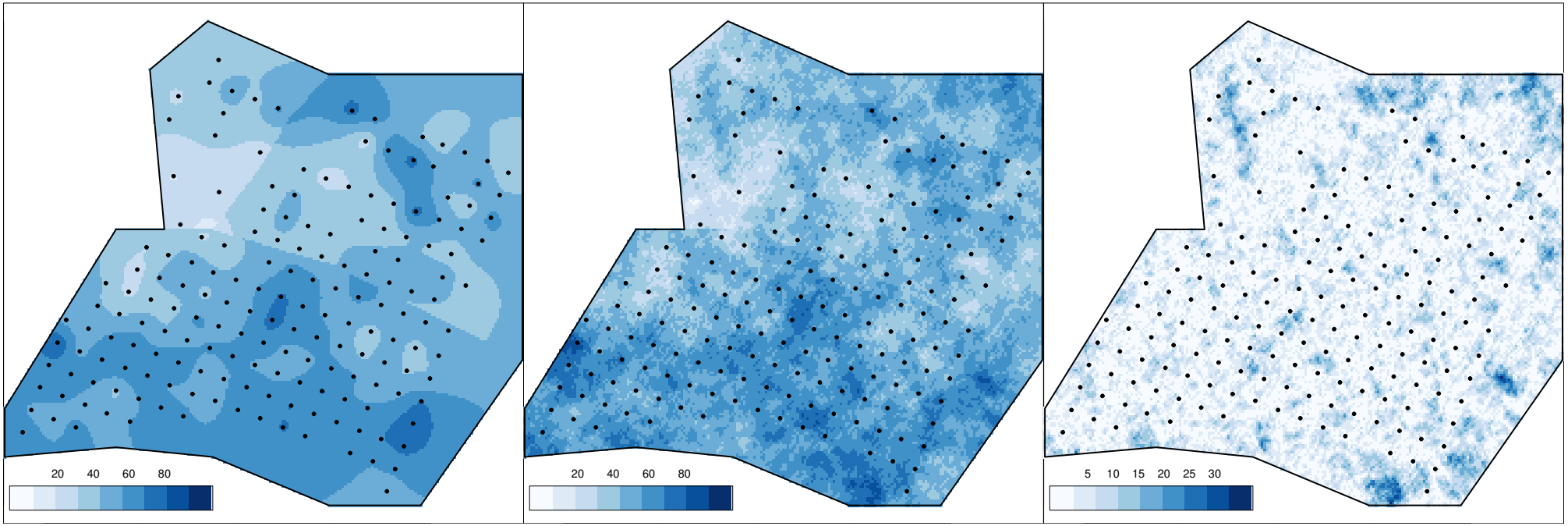


Figure 14: *Comparison between kriging (left) and a conditional simulation (middle). Right: absolute difference of the two.*

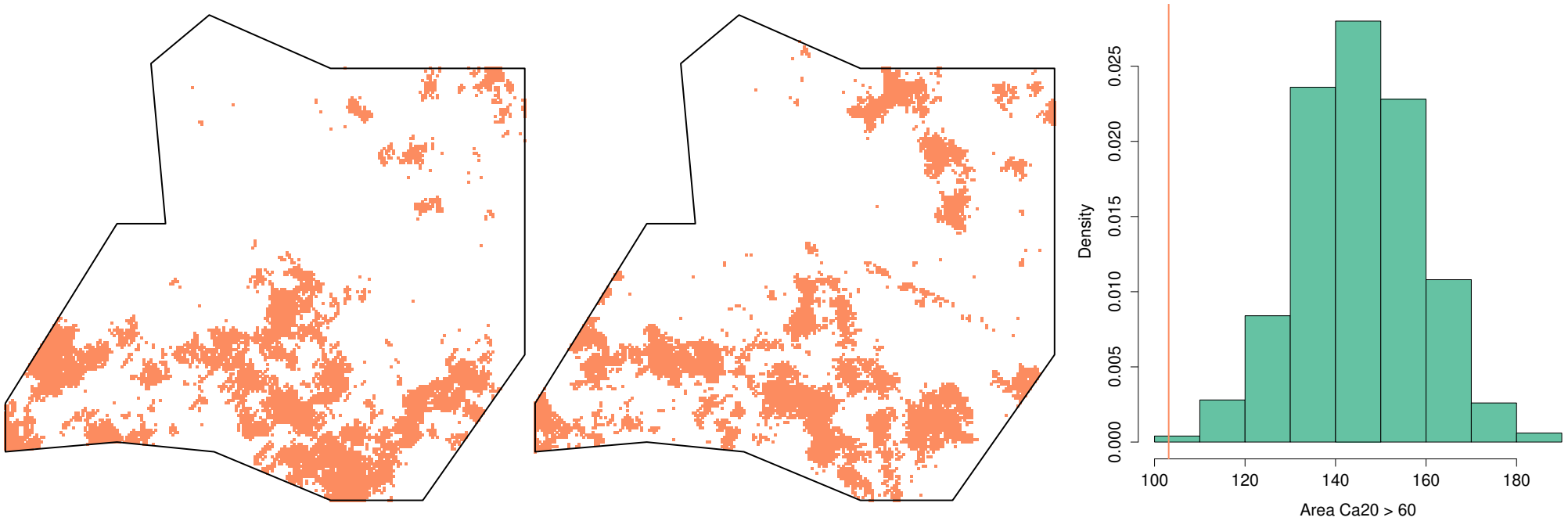


Figure 15: *Left and middle: Two sampled level sets with $u_{crit} = 60$. Right: Distribution of the expected level set area from conditional simulations (histogram) and kriging (vertical line).*

☞ As expected, the kriging–based estimator underestimates.

This is the end. . .