Spatial autocorrelation

Vojtěch Barták

LS 2022

Section 1

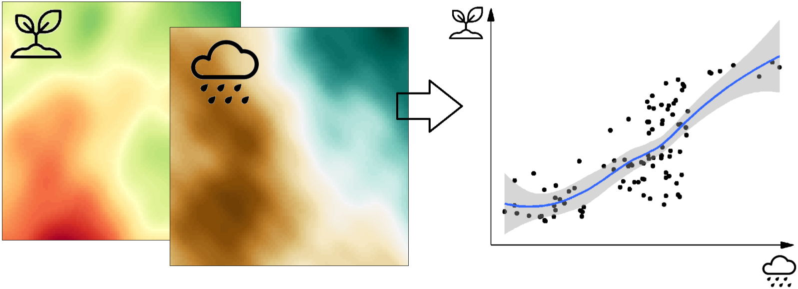
Spatial autocorrelation and why it matters

Types of data in spatial statistics

- Geostatistical data
- Point pattern data
- Areal (lattice) data

Problem with spatially autocorrelated data

Spatial or **structural** relationship?



Possible explanations:

1. A direct causal effect
2. An indirect causal effect
3. An accidental spatial association

Chapman 2010

Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2010) **19**, 831–841

RESEARCH PAPER

Weak climatic associations among British plant distributions

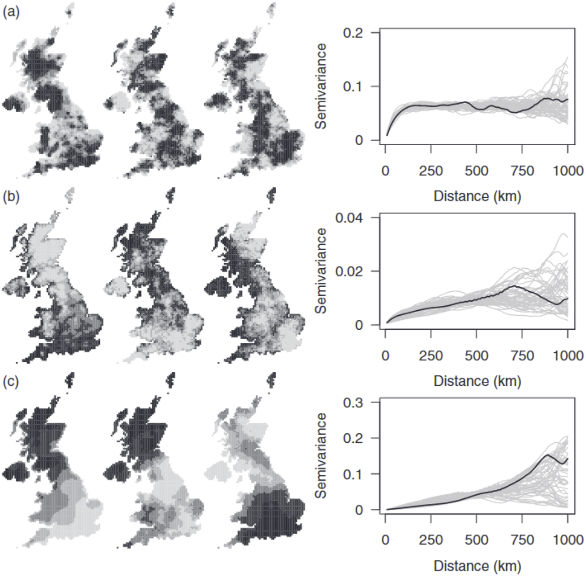
Data:

- 100 UK plant species at 10x10 km grid
- 23 climatic variables
- simulated variables with same properties

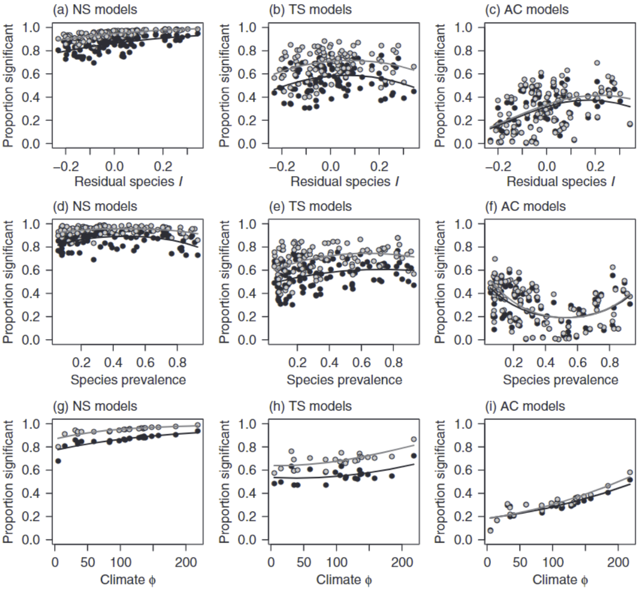
Models:

- GLM, random forests
- non-spatial models, trend surfaces, autoregression
- Single predictors, PCA

Chapman 2010



Chapman 2010



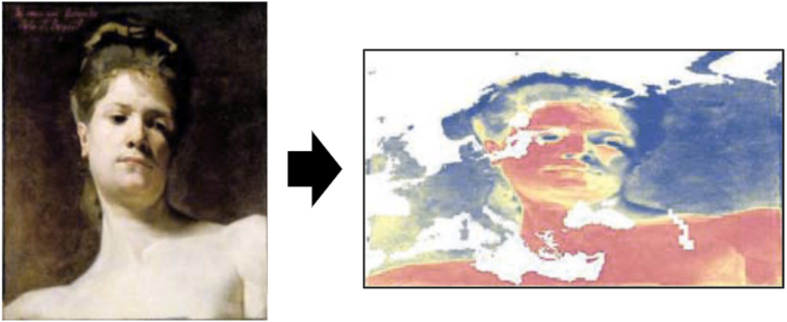
Fourcade, Besnard, Secondi 2018

MACROECOLOGICAL METHOD

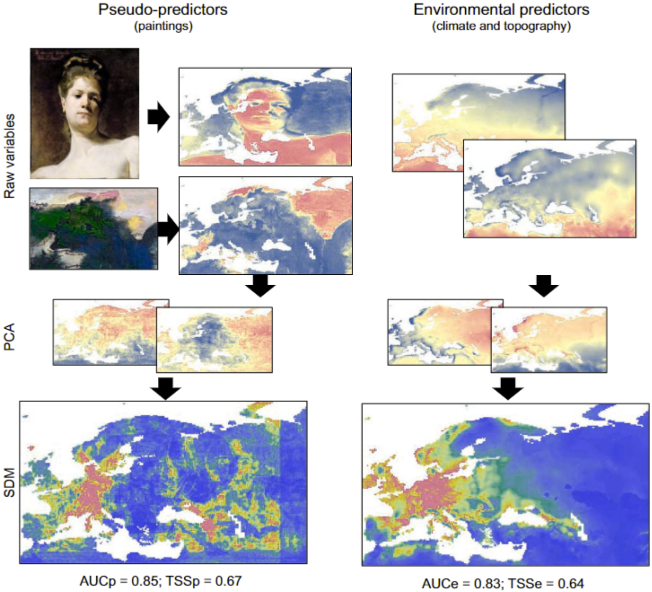
WILEY Global Ecology and Biogeography A Journal of Macroecology

Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics

- 497 species from European Red List (GBIF)
- real environmental predictors vs paintings



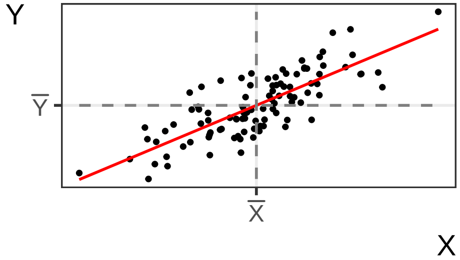
Fourcade, Besnard, Secondi 2018



Section 2

Measuring SAC

Correlation of two variables



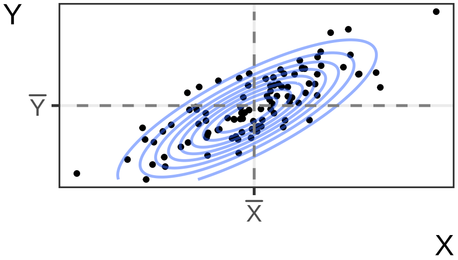
Sample covariance: $S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$

Correlation coefficient: $r = \frac{S_{XY}}{S_X S_Y}$, $-1 \leq r \leq 1$

S_X, S_Y ... sample standard deviations of X, Y

Correlation of two variables

Mathematical model: $(X, Y) \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$



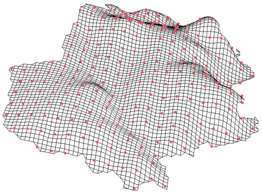
Covariance matrix: $\boldsymbol{\Sigma} = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}$

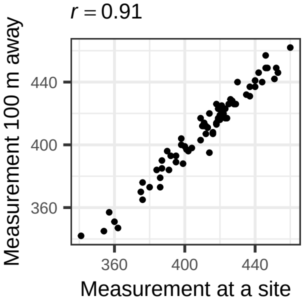
Covariance: $Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$

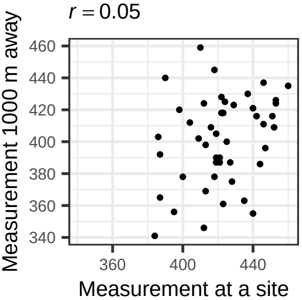
Correlation coefficient: $\rho = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$

Spatial autocorrelation

Point observations in space:

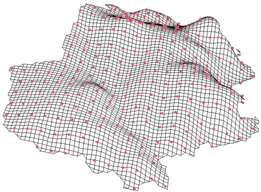




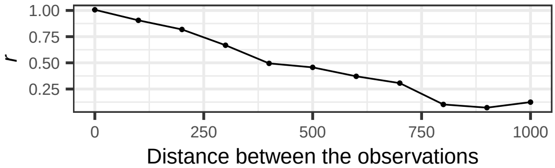


Spatial autocorrelation

Point observations in space:



Correlation as a function of distance (correlogram):



- Assumption: the underlying spatial process is **stacionary** and **isotropic**!

Moran's I

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{i=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- w_{ij} measures the magnitude of spatial interactions between sites i and j
- originally designed for lattice data!
- Standard metric used for correlograms
- Under the null hypothesis of spatial independence: $E(I) = -1/(n-1)$
- Alternative measure: Geary's C

Local Moran's I

The (global) Moran's I can be decomposed into components assosiated with the individual sites:

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^{n} w_{ij}(y_j - \bar{y})}{\frac{\sum_{j=1}^{n}(y_i - \bar{y})^2}{n}}$$

$$\sum_{i=1}^{n} I_i = I$$

Section 3

Gaussian processes, correlation function

Gaussian (spatial random) process

Gaussian process: $\left\{S(x) : x \in \mathbb{R}^2\right\}$

For any $x_1, x_2, ..., x_n$ -> $\{S(x_1), S(x_2), ..., S(x_3)\} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\mu(x) = E\left[S(x)\right] \ldots$ mean function
- $C(x, x') = Cov\left(S(x), S(x')\right) \ldots$ covariance function

Stationary, isotropic process: $\mu(x) = \mu$, $C(x, x') = C(u)$

We can then define:

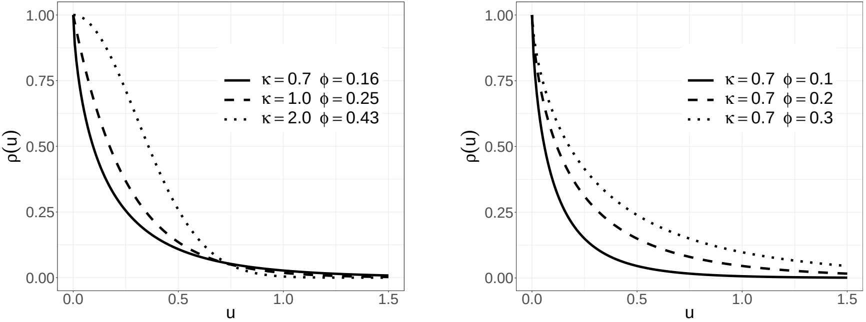
- $\sigma^2 = C(0) \ldots$ (constant) variance of the process
- $\rho(u) = C(u)/\sigma^2 \ldots$ **correlation function** of the process

Power-Exponential family

$$\rho(u) = exp\left\{-\left(\frac{u}{\phi}\right)^{\kappa}\right\}$$

ϕ ... scale, $\phi > 0$

κ ... shape, $0 < \kappa \leq 2$

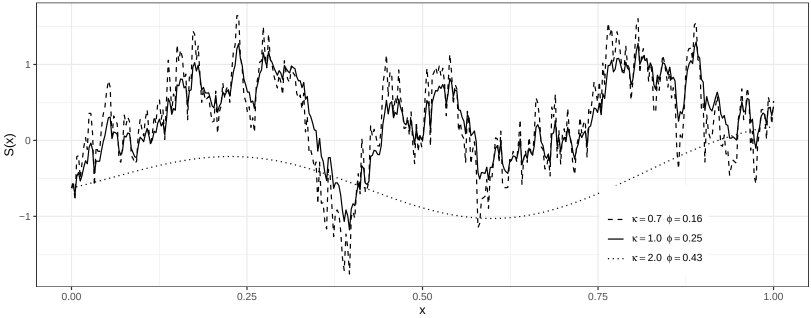


Power-Exponential family

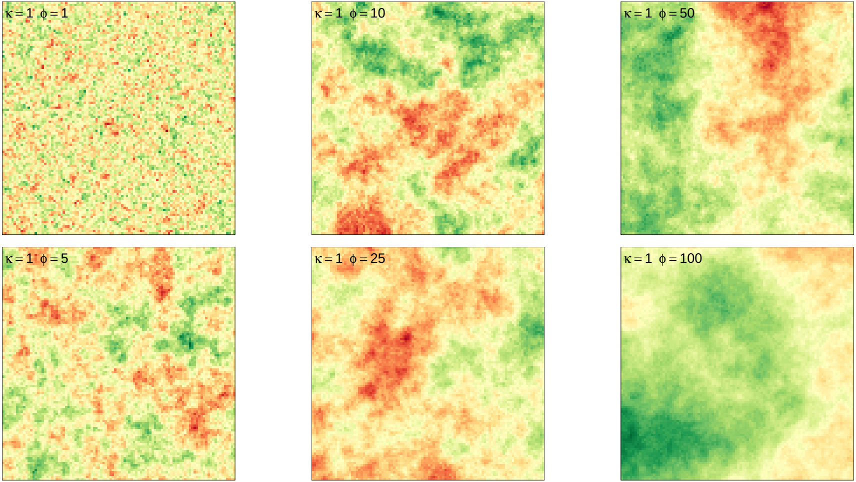
$$\rho(u) = exp\left\{-\left(\frac{u}{\phi}\right)^{\kappa}\right\}$$

ϕ . . . scale, $\phi > 0$

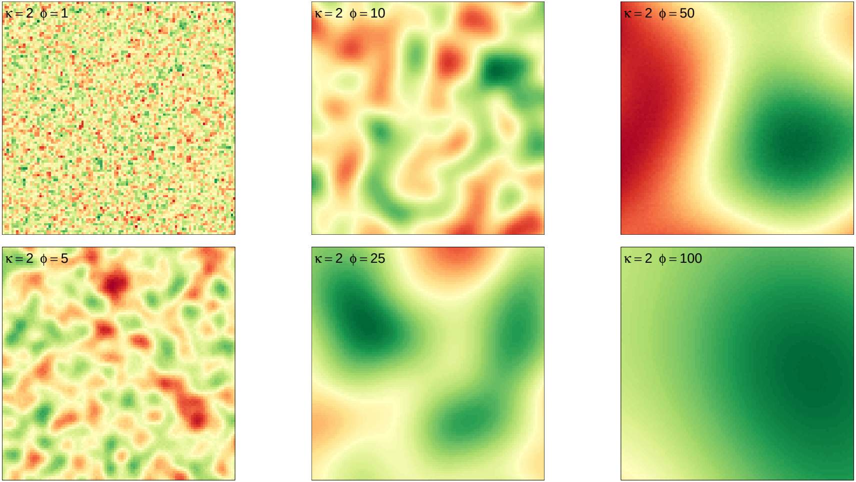
κ . . . shape, $0 < \kappa \leq 2$



Power-Exponential family



Power-Exponential family



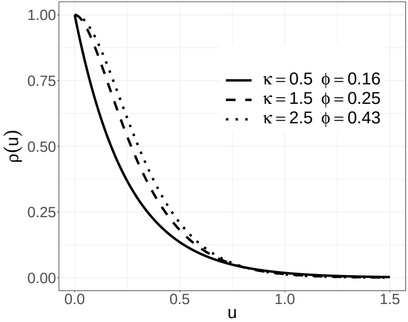
Matérn family

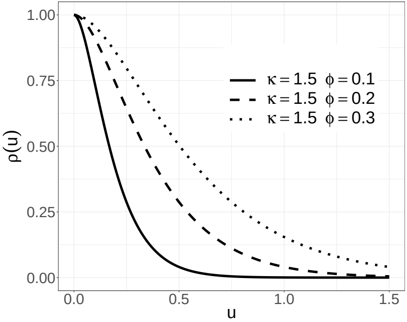
$$\rho(u) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}\left(\frac{u}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{u}{\phi}\right)$$

ϕ ... scale, $\phi > 0$

κ ... order, $\kappa > 0$

K_κ ... modifikovaná Besselova funkce druhého typu řádu κ





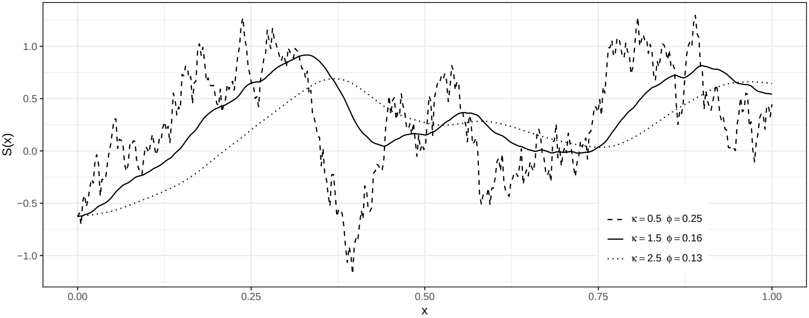
Matérn family

$$\rho(u) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}\left(\frac{u}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{u}{\phi}\right)$$

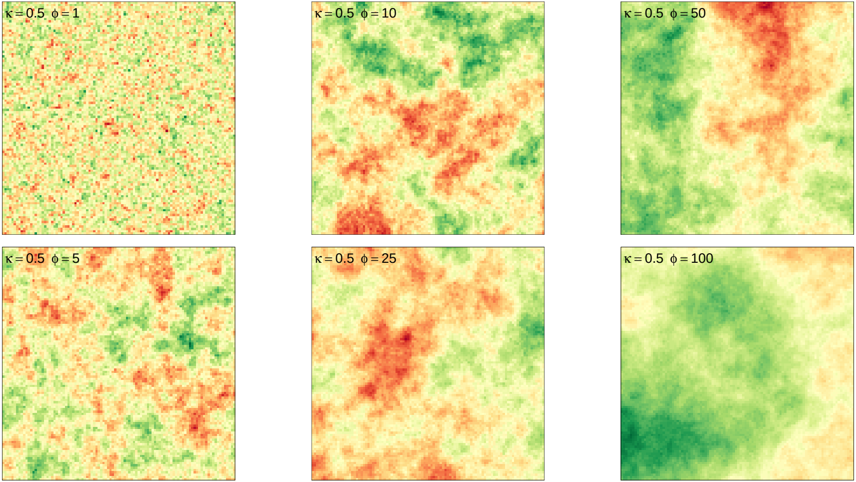
ϕ ... scale, $\phi > 0$

κ ... order, $\kappa > 0$

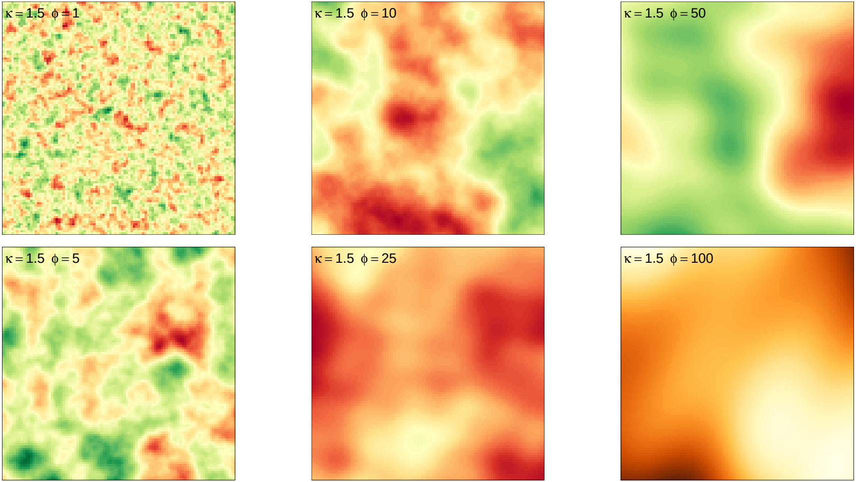
K_κ ... modifikovaná Besselova funkce druhého typu řádu κ



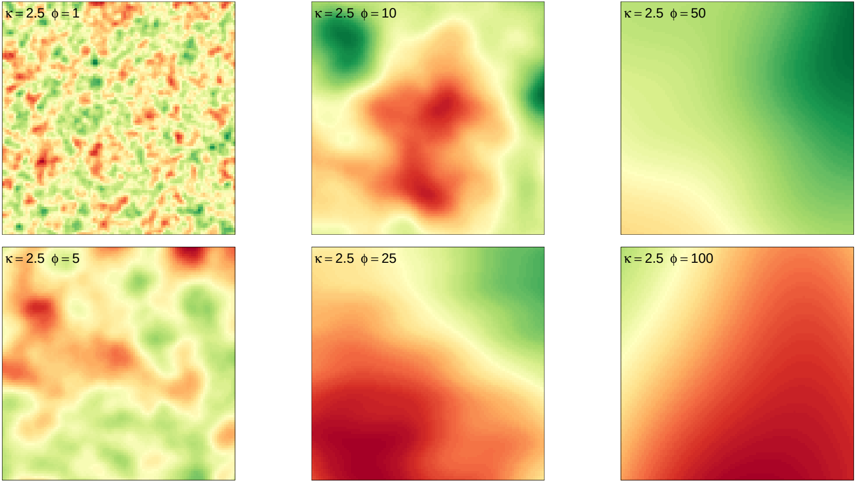
Matérn family



Matérn family



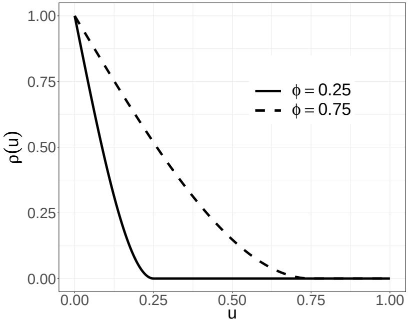
Matérn family

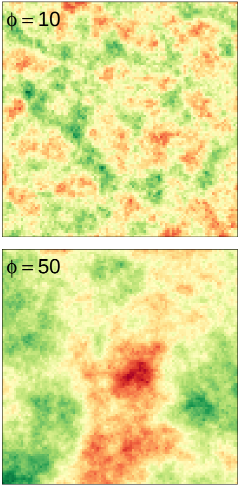


Spherical family

$$\rho(u) = \begin{cases} 1 - \frac{3}{2}\left(\frac{u}{\phi}\right) + \frac{1}{2}\left(\frac{u}{\phi}\right)^3 & 0 \le u \le \phi \\ 0 & u > \phi \end{cases}$$

ϕ . . . range, $\phi > 0$





Section 4

Variogram

Variogram

(Semi)Variogram ... alternative description of the autocorrelation structure:

$$V(x, x') = \frac{1}{2} Var\left(S(x) - S(x')\right)$$

For stationary isotropic process, it is easy to show that:

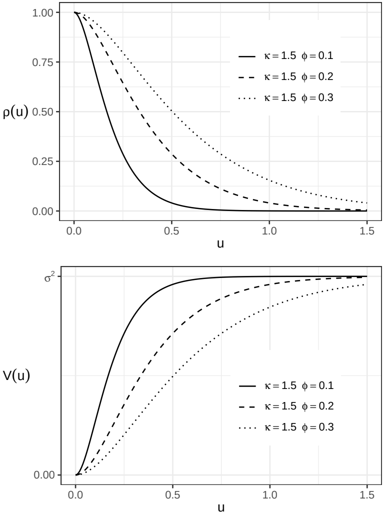
$$V(u) = \sigma^2 - C(u)$$

$$V(u) = \sigma^2 \left\{1 - \rho(u)\right\}$$

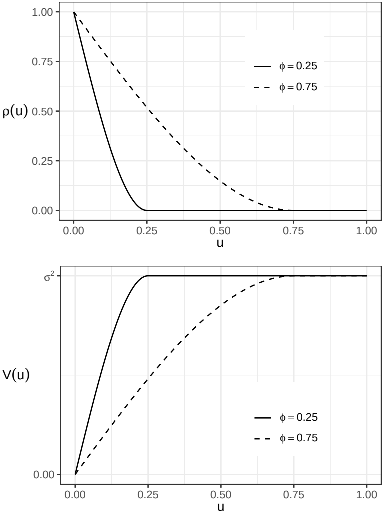
-> Variogram has an *inverse shape* compared to the correlation function.

Variogram

Matérn:



Spherical:



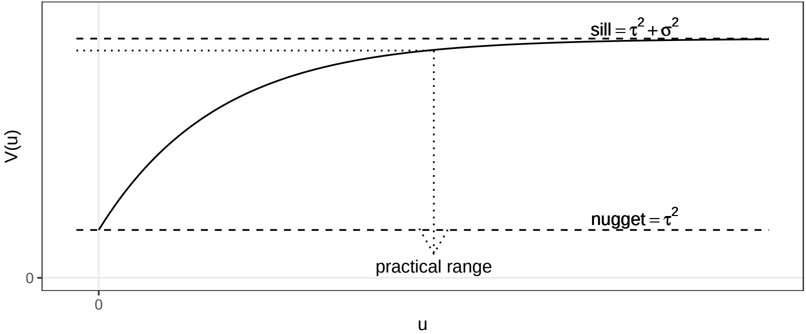
Nugget effect

Realistic assumption: data (x_i, y_i) are generated by a stationary "observational" process

$Y_i = S(x_i) + \epsilon_i$, where $z_i \sim iid, E(\epsilon_i) = 0, Var(\epsilon_i) = \tau^2$

Variogram is then:

$$V_Y = \tau^2 + \sigma^2 \{1 - \rho(u)\}$$



Nugget effect

First approach:

- we decide to *model the nugget effect*
- τ^2 is estimated from data
- two-fold interpretation:
  1. measurement error
  2. variability of $S(x)$ in the scale smaller than the scale of the measurement
- cannot be separated, unless we have repeated measurements at the same site
- interpolation does not fit the data points precisely

Second approach:

- we decide to *set the* $\tau^2 = 0$ *fixly* (no measurement error)
- interpolation exactly fits the data points

Variogram estimation

1. Compute $v_{ij} = \frac{1}{2}\left(Y_i - Y_j\right)^2$
2. Plot v_{ij} against u -> **variogram cloud**
3. Smooth out the variogram cloud -> **empirical variogram**