NRCSE

2.1 Spatial covariances

Valid covariance functions

Bochner's theorem: The class of covariance functions is the class of positive definite functions C:

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

Why?

$$\sum_i \sum_j a_i a_j C(s_i, s_j) = Var(\sum a_i Z(s_i))$$

Spectral representation

By the spectral representation any isotropic continuous correlation on R^d is of the form

$$\rho(v) = E\left(e^{iu^T X}\right), v = \|u\|, X \in R^d$$

By isotropy, the expectation depends only on the distribution G of $\|X\|$. Let Y be uniform on the unit sphere. Then

$$\rho(v) = Ee^{iv\|X\|Y} = E\Phi_Y(v\|X\|)$$

Isotropic correlation

$$\Phi_Y(u) = \left(\frac{2}{u}\right)^{\frac{d}{2}-1} \Gamma\left(\frac{d}{2}\right) J_{\frac{d}{2}-1}(u)$$

$J_v(u)$ is a Bessel function of the first kind and order v.

Hence

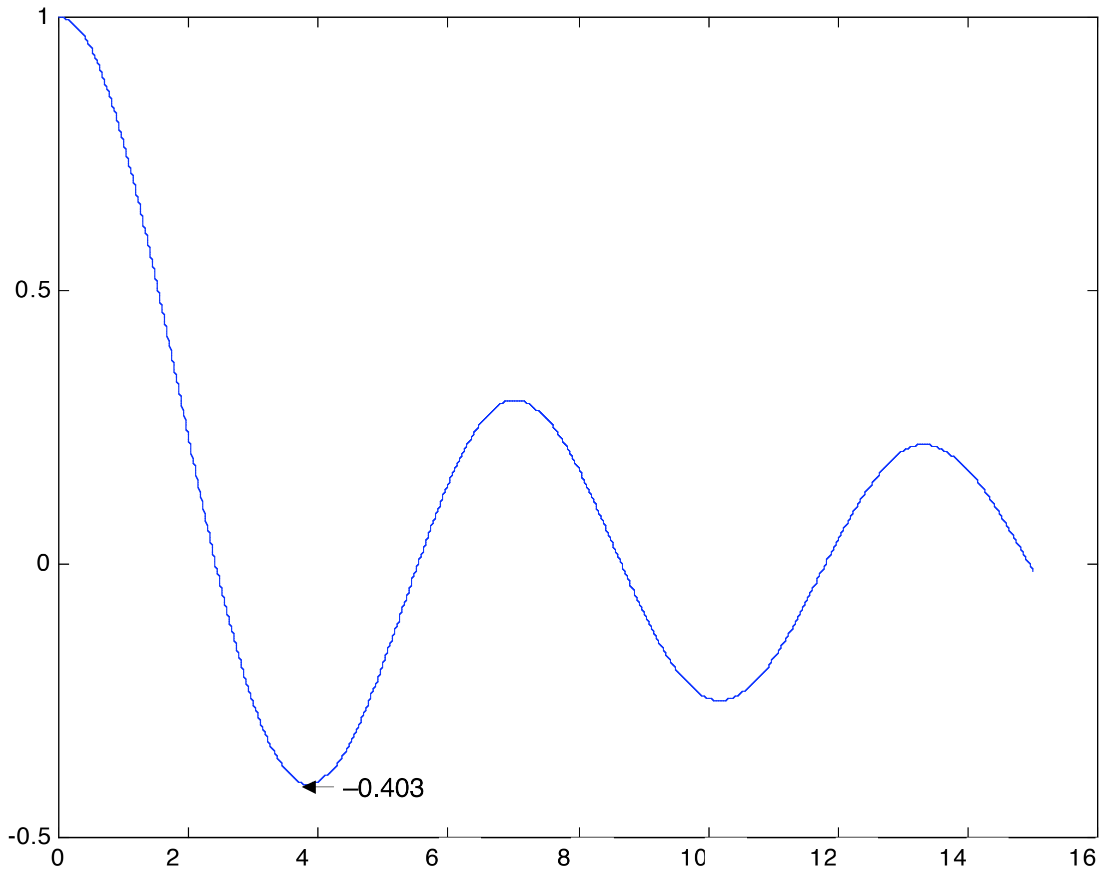
$$\rho(v) = \int_0^\infty \Phi_Y(sv)\,dG(s)$$

and in the case d=2

$$\rho(v) = \int_0^\infty J_0(sv)\,dG(s)$$

(Hankel transform)

The Bessel function J_0



The exponential correlation

A commonly used correlation function is $\rho(v) = e^{-v/\phi}$. Corresponds to a Gaussian process with continuous but not differentiable sample paths.

More generally, $\rho(v) = c(v{=}0) + (1\text{-}c)e^{-v/\phi}$ has a *nugget* c, corresponding to measurement error and spatial correlation at small distances.

All isotropic correlations are a mixture of a nugget and a continuous isotropic correlation.

The squared exponential

Using $$G'(x) = \frac{2x}{\phi^2} e^{-4x^2/\phi^2}$$ **yields**

$$\rho(v) = e^{-\left(v/\phi\right)^2}$$

corresponding to an underlying Gaussian field with analytic paths.

This is sometimes called the Gaussian covariance, for no really good reason.

A generalization is the *power(ed) exponential* correlation function,

$$\rho(v) = \exp\left(-\left[v/\phi\right]^\kappa\right), 0 < \kappa \leq 2$$

The spherical

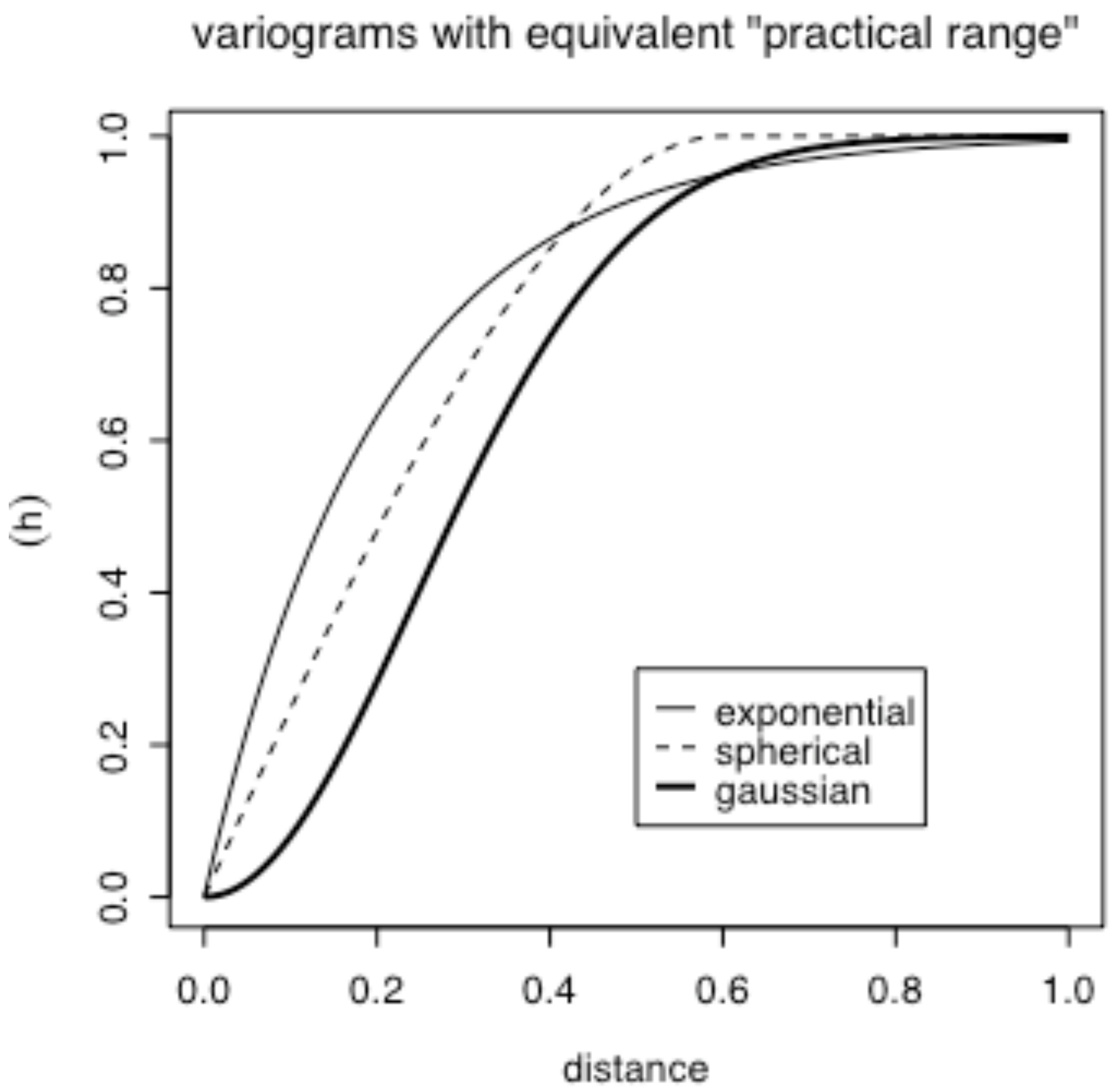
$$\rho(\mathbf{v}) = \begin{cases} 1 - 1.5\mathbf{v} + 0.5\left(\frac{\mathbf{v}}{\phi}\right)^3; & \mathbf{h} < \phi \\ 0, & \text{otherwise} \end{cases}$$

Corresponding variogram

$$\boxed{\tau^2} + \frac{\sigma^2}{2}\left(3\frac{t}{\phi} + \left(\frac{t}{\phi}\right)^3\right); \quad 0 \le t \le \phi$$

$$\boxed{\tau^2 + \sigma^2}; \quad t > \boxed{\phi}$$

nugget → τ^2; sill → $\tau^2 + \sigma^2$; range → ϕ


variograms with equivalent "practical range"
(h)
distance
0.0
0.2
0.4
0.6
0.8
1.0
exponential
spherical
gaussian

The Matérn class

$$G'(x) = \frac{2\kappa}{\phi^{2\kappa}} \frac{x}{(x^2 + \phi^{-2})^{1+\kappa}}$$

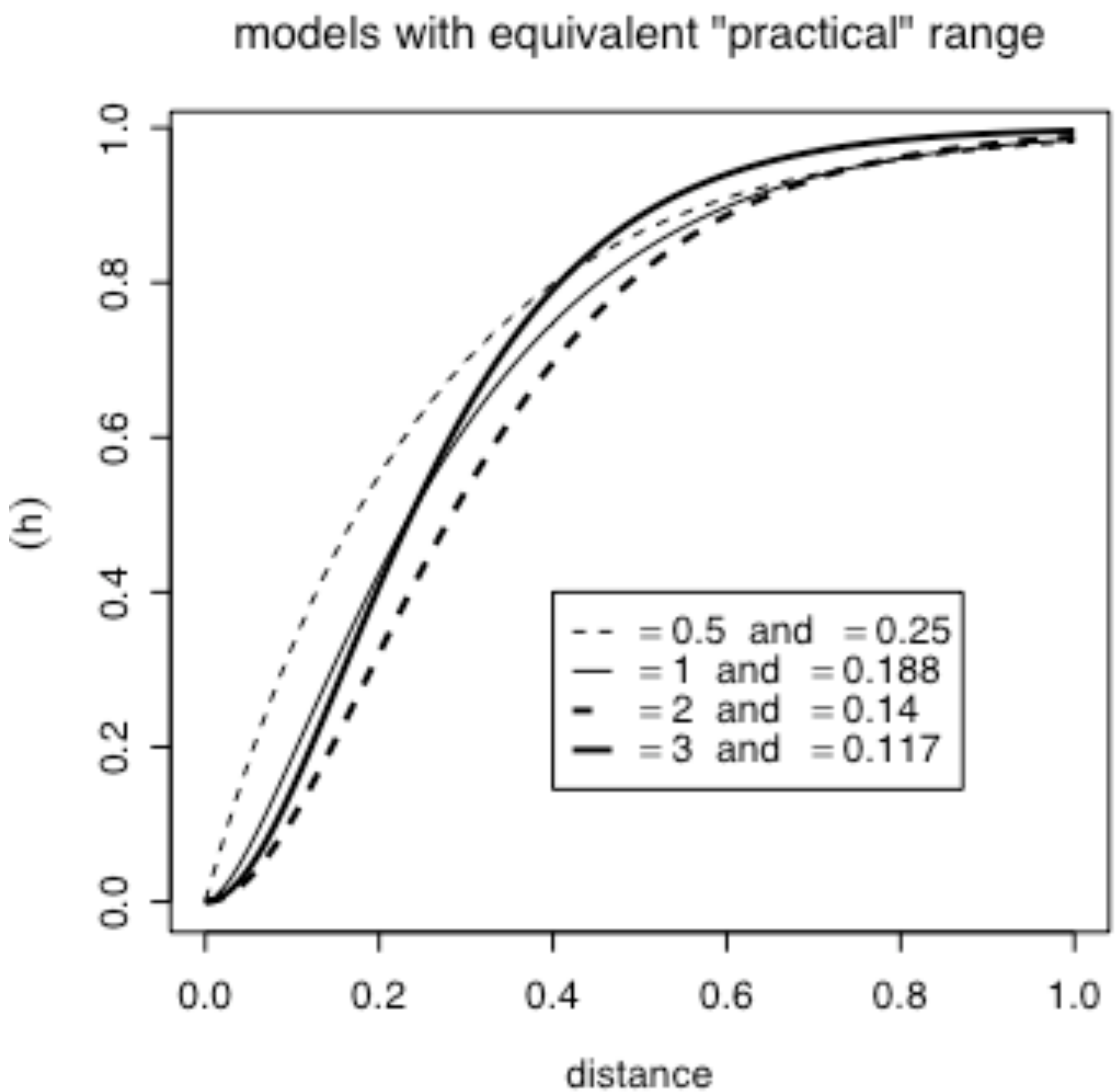
$$\rho(v) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{v}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{v}{\phi}\right)$$

where K_{κ} is a modified Bessel function of the third kind and order κ. It corresponds to a spatial field with $\kappa-1$ continuous derivatives

κ=1/2 is exponential;

κ=1 is Whittle's spatial correlation;

$\kappa \to \infty$ yields squared exponential.


models with equivalent "practical" range
(h)
distance
0.0
0.2
0.4
0.6
0.8
1.0
= 0.5 and = 0.25
= 1 and = 0.188
= 2 and = 0.14
= 3 and = 0.117

Some other covariance/variogram families

Name	Covariance	Variogram
Wave	$\sigma^2 \frac{\sin(\phi t)}{\phi t}$	$\tau^2 + \sigma^2 (1 - \frac{\sin(\phi t)}{\phi t})$
Rational quadratic	$\sigma^2 (1 - \frac{t^2}{1 + \phi t^2})$	$\tau^2 + \frac{\sigma^2 t^2}{1 + \phi t^2}$
Linear	None	$\tau^2 + \sigma^2 t$
Power law	None	$\tau^2 + \sigma^2 t^{\phi}$

Estimation of variograms

Recall $\gamma(v) = \sigma^2(1-\rho(v))$

Method of moments: square of all pairwise differences, smoothed over lag bins

$$\bar{\gamma}(h) = \frac{1}{|N(h)|}\sum_{i,j\in N(h)}(Z(s_i)-Z(s_j))^2$$

$$N(h) = \left\{(i,j) : h - \frac{\Delta h}{2} \le |s_i - s_j| \le h + \frac{\Delta h}{2}\right\}$$

Problems: Not necessarily a valid variogram

Not very robust

A robust empirical variogram estimator

$(Z(x)-Z(y))^2$ is chi-squared for Gaussian data

Fourth root is variance stabilizing

Cressie and Hawkins:

$$\tilde{\gamma}(h) = \frac{\left\{ \frac{1}{|N(h)|} \sum |Z(s_i) - Z(s_j)|^{1/2} \right\}^4}{0.457 + \frac{0.494}{|N(h)|}}$$

Least squares

Minimize

$$\theta \mapsto \sum_i \sum_j \left([(Z(s_i) - Z(s_j)]^2 - \gamma(\|s_i - s_j\|;\theta) \right)^2$$

Alternatives:

- **fourth root transformation**
- **weighting by $1/\gamma^2$**
- **generalized least squares**

Maximum likelihood

$Z \sim N_n(\mu,\Sigma)$ $\Sigma = \alpha[\rho(s_i - s_j;\theta)] = \alpha\, V(\theta)$

Maximize

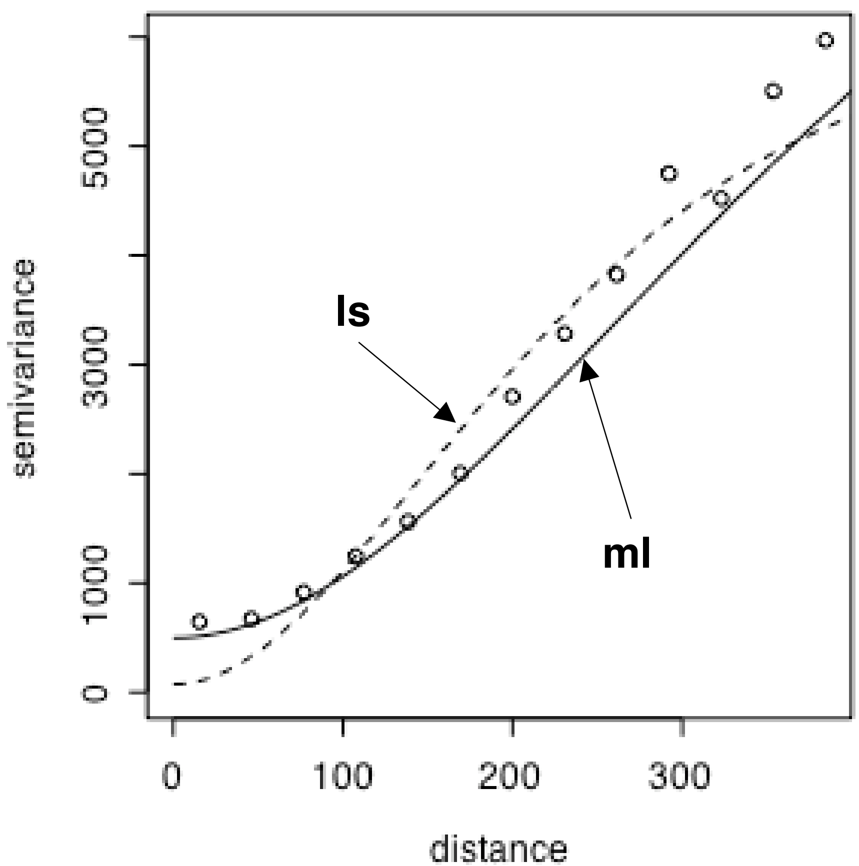
$$\ell(\mu,\alpha,\theta) = -\frac{n}{2}\log(2\pi\alpha) - \frac{1}{2}\log\det V(\theta) + \frac{1}{2\alpha}(Z-\mu)^T V(\theta)^{-1}(Z-\mu)$$

$$\hat{\mu} = 1^T Z / n \quad \hat{\alpha} = G(\hat{\theta}) / n \quad G(\theta) = (Z-\hat{\mu})^T V(\theta)^{-1}(Z-\hat{\mu})$$

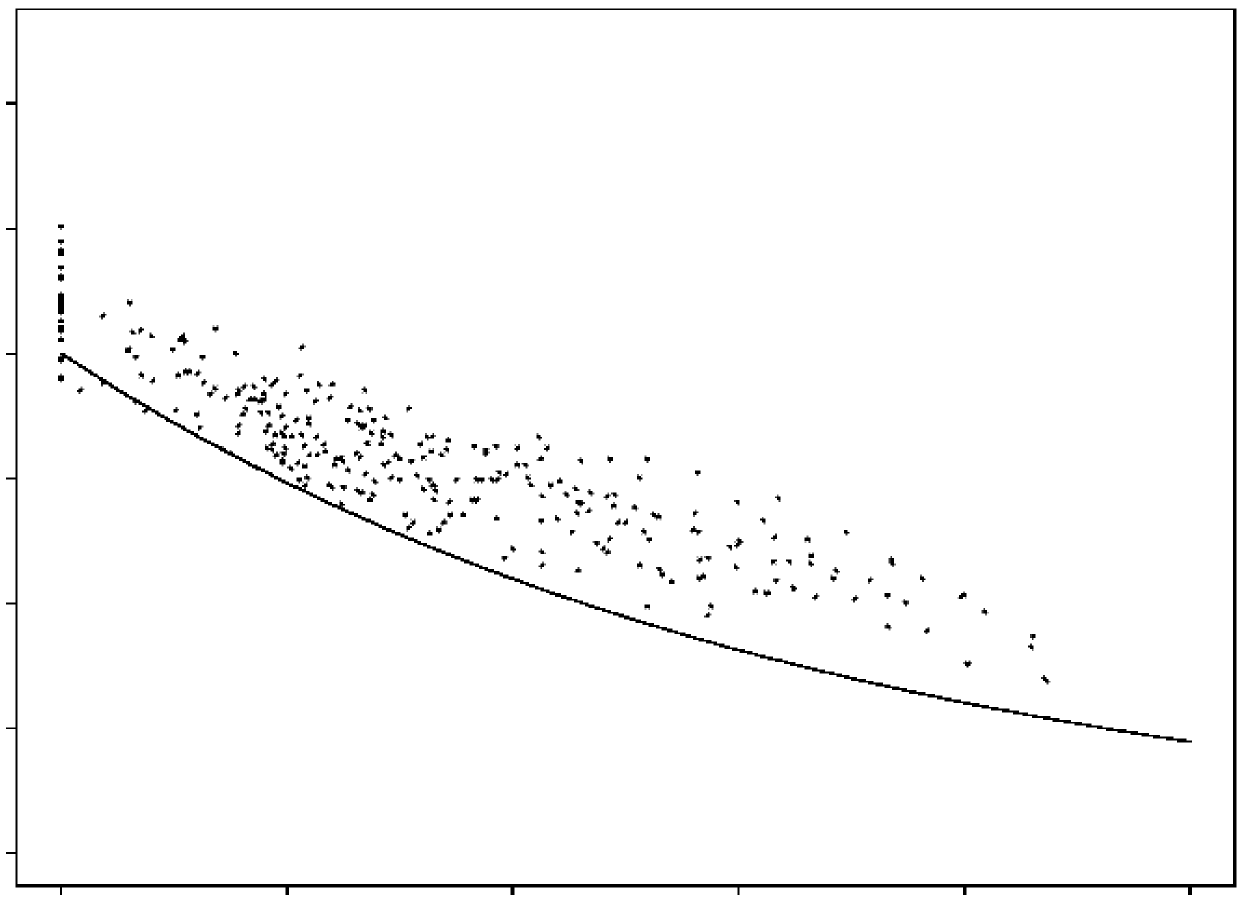
and θ maximizes the profile likelihood

$$\ell^*(\theta) = -\frac{n}{2}\log\frac{G^2(\theta)}{n} - \frac{1}{2}\log\det V(\theta)$$

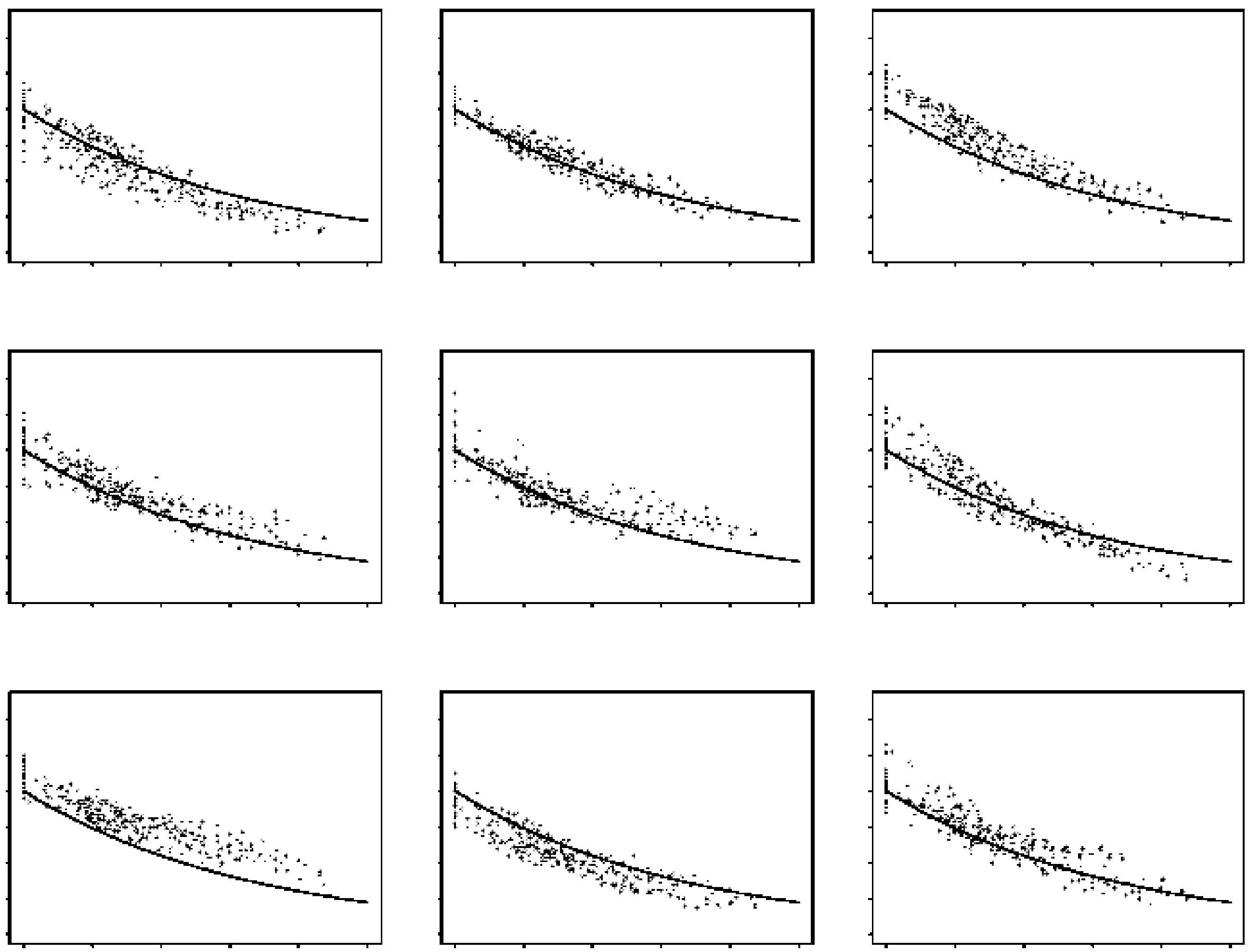
Parana data



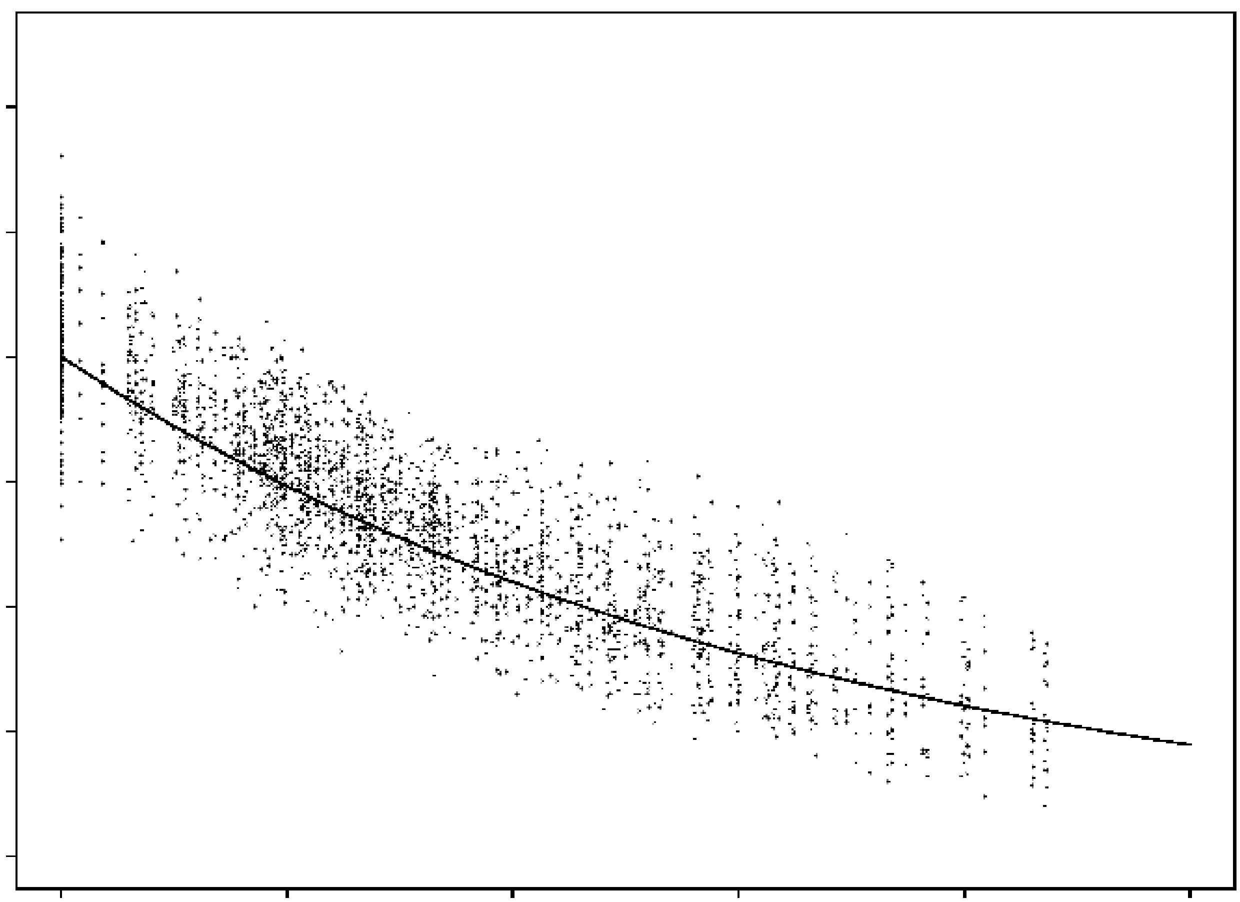
A peculiar ml fit



Some more fits



All together now...



Bayesian kriging

Instead of estimating the parameters, we put a prior distribution on them, and update the distribution using the data.

Model: $(Z \mid \theta) \sim N(\beta, \sigma^2 C(\phi) + \tau^2 I)$

Prior: $f(\theta) = f(\beta) f(\sigma^2) f(\phi) f(\tau^2)$

Posterior:

$$f(\beta \mid Z = z) \propto f(\beta) \iiint f(z \mid \theta) f(\sigma^2) f(\phi) f(\tau^2) \, d\sigma^2 \, d\phi \, d\tau^2$$

geoR

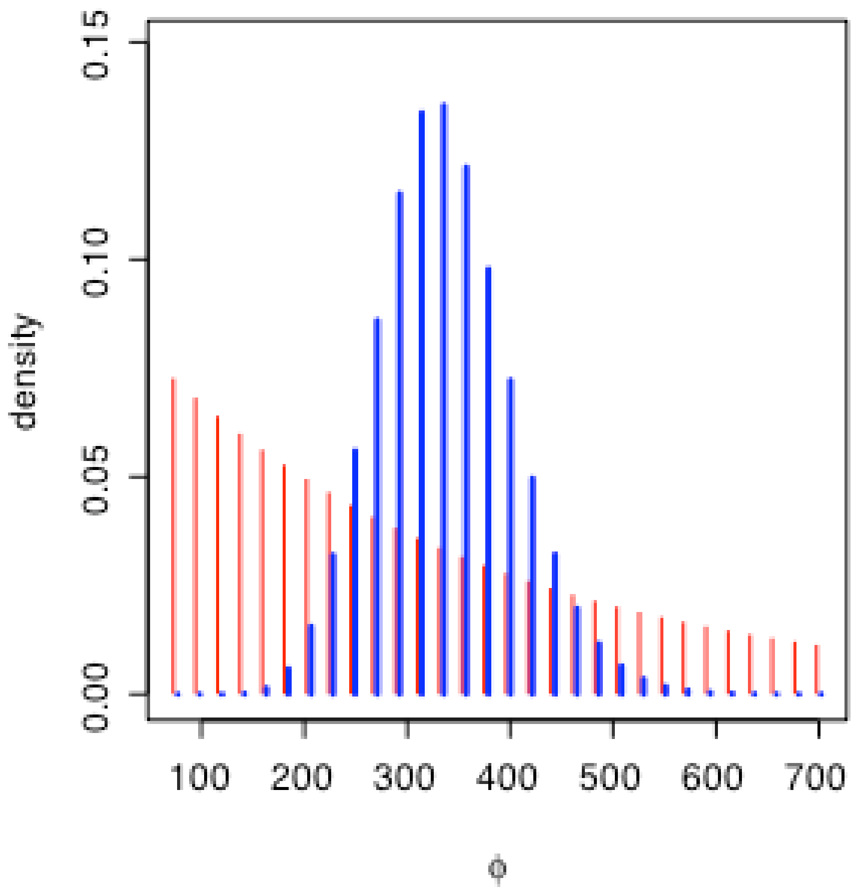
Prior is assigned to ϕ and τ/σ. The latter assumed zero unless specified.

The distributions are discretized.

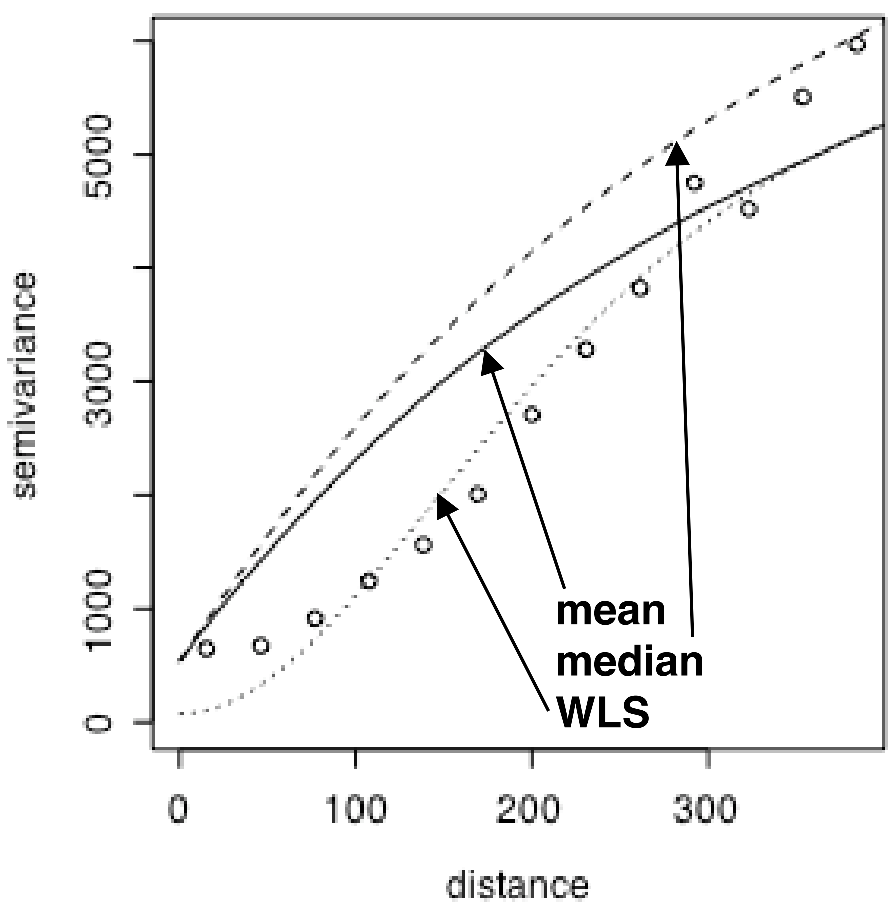
Default prior on mean β is flat (if not specified, assumed constant).

(Lots of different assignments are possible)

Prior/posterior of ϕ

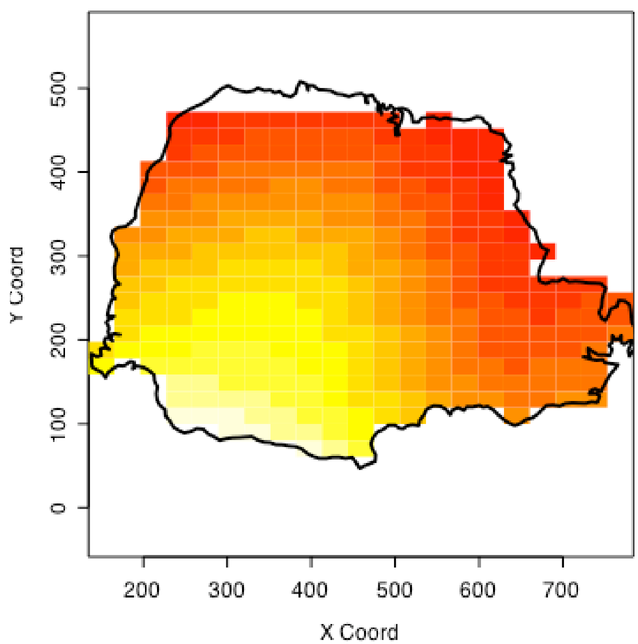


Variogram estimates

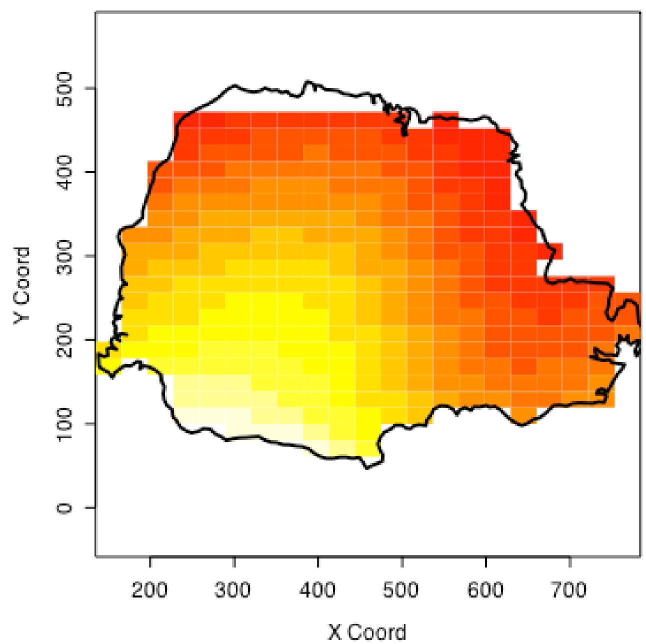


Bayes vs universal kriging

Bayes predictive mean



Universal kriging



Spectral representation

Stationary processes

$$Z(s) = \int_{R^d} \exp(is^T\omega)dY(\omega)$$

Spectral process Y has stationary increments

$$E|dY(\omega)|^2 = dF(\omega)$$

If F has a density f, it is called the spectral density.

$$Cov(Z(s_1),Z(s_2)) = \int_{R^2} e^{i(s_1-s_2)^T\omega} f(\omega)d\omega$$

Estimating the spectrum

For process observed on nxn grid, estimate spectrum by *periodogram*

$$I_{n,n}(\omega) = \frac{1}{(2\pi n)^2} \left| \sum_{j \in J} z(j) e^{i\omega^T j} \right|^2$$

$$\omega = \frac{2\pi j}{n};\ J = \left\{ \lfloor (n-1)/2 \rfloor, \ldots, n - \lfloor (n-1)/2 \rfloor \right\}^2$$

Equivalent to DFT of sample covariance

Properties of the periodogram

Periodogram values at Fourier frequencies (j,k)π/Δ are

- **uncorrelated**
- **asymptotically unbiased**
- **not consistent**

To get a consistent estimate of the spectrum, smooth over nearby frequencies

Some common isotropic spectra

Squared exponential

$$f(\omega)=\frac{\sigma^2}{2\pi\alpha}\exp(-\|\omega\|^2/4\alpha)$$

$$C(r)=\sigma^2\exp(-\alpha\|r\|^2)$$

Matérn

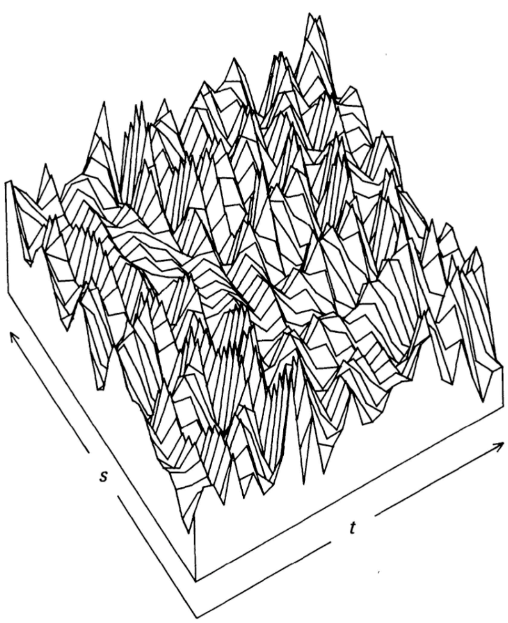
$$f(\omega)=\phi(\alpha^2+\|\omega\|^2)^{-\nu-1}$$

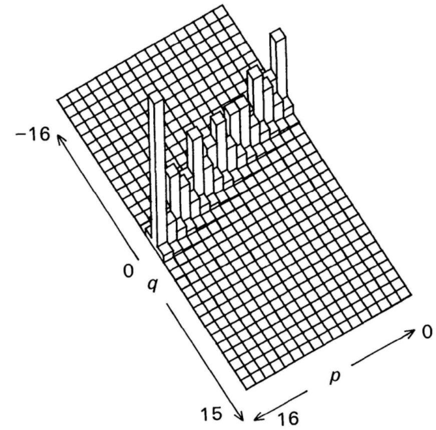
$$C(r)=\frac{\pi\phi(\alpha\|r\|)^{\nu}\mathcal{K}_{\nu}(\alpha\|r\|)}{2^{\nu-1}\Gamma(\nu+1)\alpha^{2\nu}}$$

A simulated process

$$Z(s) = \sum_{j=0}^{15} \sum_{k=-15}^{15} g_{jk} \cos\left(2\pi \left[\frac{js_1}{m} + \frac{ks_2}{n} \right] + U_{jk} \right)$$

$$g_{jk} = \exp(-|j + 6 - k \tan(20°)|)$$

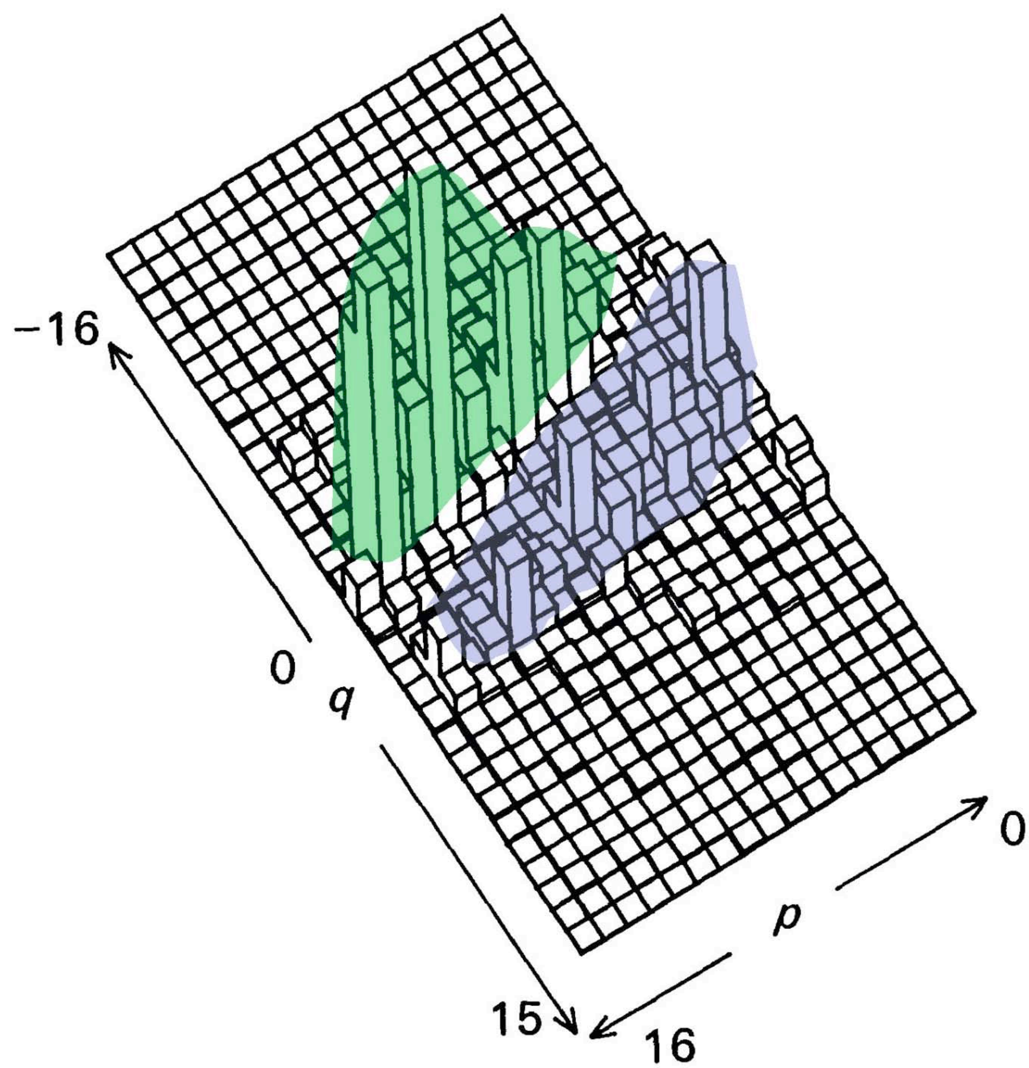




Thetford canopy heights

- 39-year thinned commercial plantation of Scots pine in Thetford Forest, UK
- Density 1000 trees/ha
- 36m x 120m area surveyed for crown height
- Focus on 32 x 32 subset

Spectrum of canopy heights



Whittle likelihood

Approximation to Gaussian likelihood using periodogram:

$$\ell(\theta) = \sum_{\omega} \left\{ \log f(\omega;\theta) + \frac{I_{N,N}(\omega)}{f(\omega;\theta)} \right\}$$

where the sum is over Fourier frequencies, avoiding 0, and f is the spectral density

Takes O(N logN) operations to calculate instead of $O(N^3)$.

Using non-gridded data

Consider

$$Y(x) = \Delta^{-2} \int h(x-s)Z(s)ds$$

where

$$h(x) = 1(|x_i| \leq \Delta / 2,\ i = 1,2)$$

Then Y is stationary with spectral density

$$f_Y(\omega) = \frac{1}{\Delta^2} |H(\omega)|^2 f_Z(\omega)$$

Viewing Y as a lattice process, it has spectral density

$$f_{\Delta,Y}(\omega) = \sum_{q \in Z^2} \left| H(\omega + \frac{2\pi q}{\Delta}) \right|^2 f_Z(\omega + \frac{2\pi q}{\Delta})$$

Estimation

Let $Y_{n^2}(x) = \frac{1}{n_x} \sum_{i \in J_x} h(s_i - x) Z(s_i)$

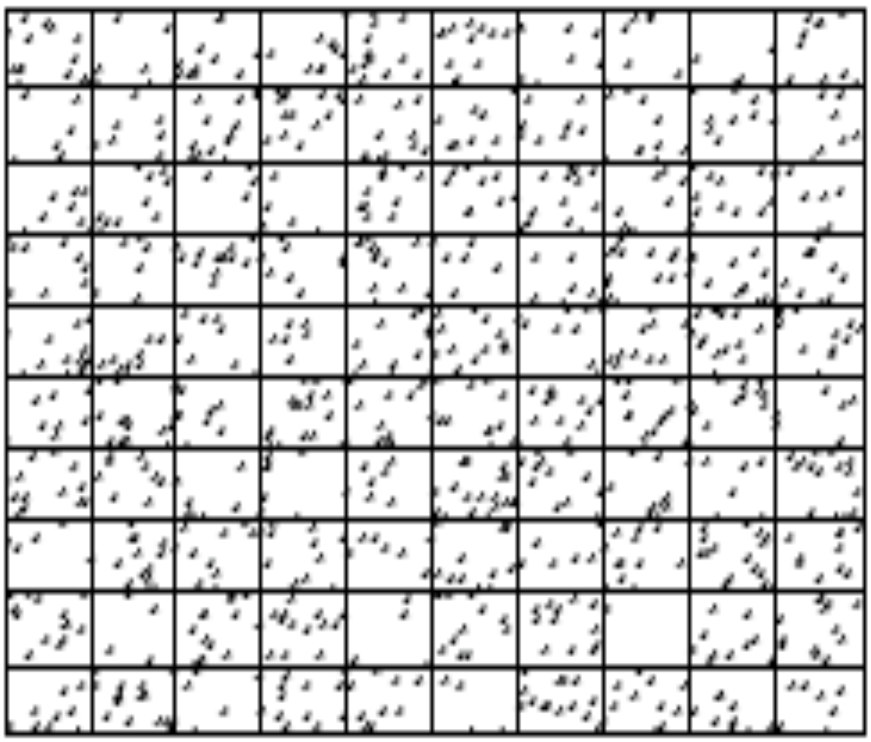
where J_x is the grid square with center x and n_x is the number of sites in the square. Define the tapered periodogram

$$I_{g_1 Y_{n^2}}(\omega) = \frac{1}{\sum g_1^2(x)} \left| \sum g_1(x) Y_{n^2}(x) e^{-ix^T\omega} \right|^2$$

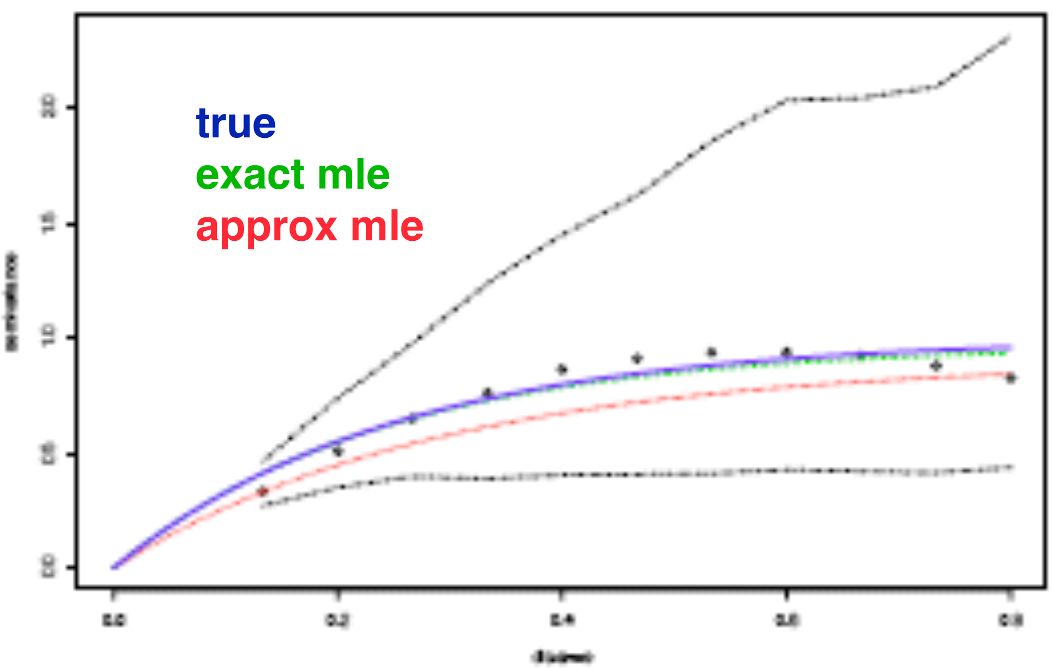
where $g_1(x) = n_x / \bar{n}$. The Whittle likelihood is approximately

$$\ell_Y = \frac{n^2}{(2\pi)^2} \sum_j \left\{ \log f_{\Delta,Y}(2\pi j / n) + \frac{I_{g_1, Y_{n^2}}(2\pi j / n)}{f_{\Delta,Y}(2\pi j / n)} \right\}$$

A simulated example

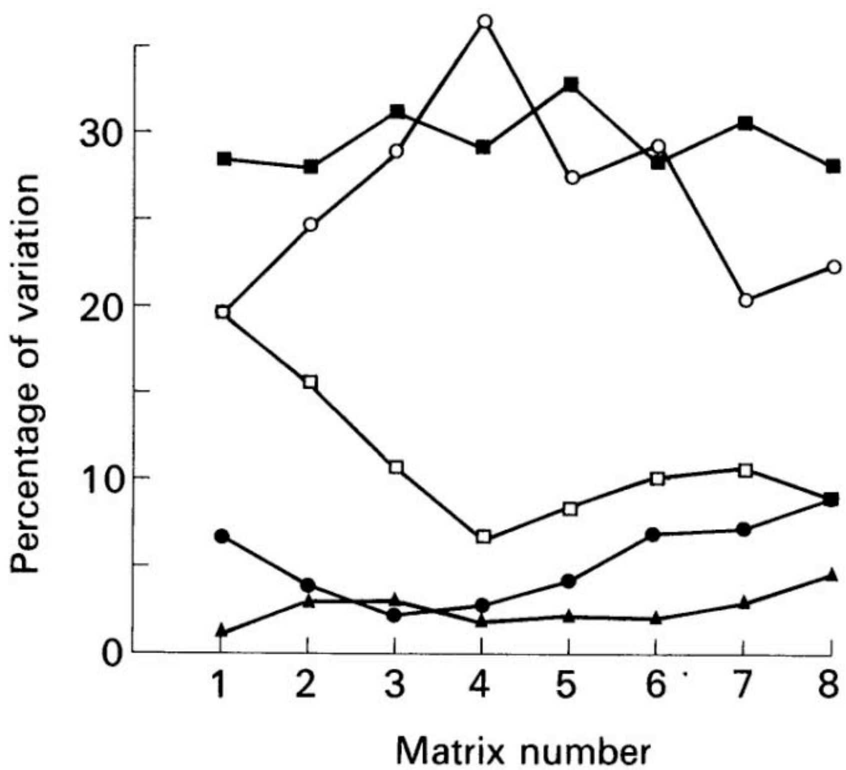


Estimated variogram



Thetford revisited

Features depend on spatial location



Some references

Bertil Matern; *Spatial Variation*. Medd. Statens Skogsforskningsinst.49: 5, Ch 2-3.

(Reprinted in Springer Lecture Notes In Statistics, vol. 36)

Cressie: ch. 2.3.1, 2.4, 2.6.

P. Guttorp, M. Fuentes and P. D. Sampson (2006): Using transforms to analyze space-time processes.

http://www.nrcse.washington.edu/pdf/trs80.pdf (to appear,SemStat 2004 proceedings, Chapman & Hall)

Banerjee, S., Carlin, B. P. and Gelfand, A. E.: *Hierarchical Modeling and Analysis for Spatial Data,* Chapman and Hall, 2004. pp. 129-135.