NRCSE

2. Covariances

Valid covariance functions

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?

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

Bochner's theorem: Every positive definite function C continuous at 0,0 can be written

$$C(t) = \int \exp(i < t, u > d\mu(u)$$

for a finite measure μ on R^2.
(Spectral representation)

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.

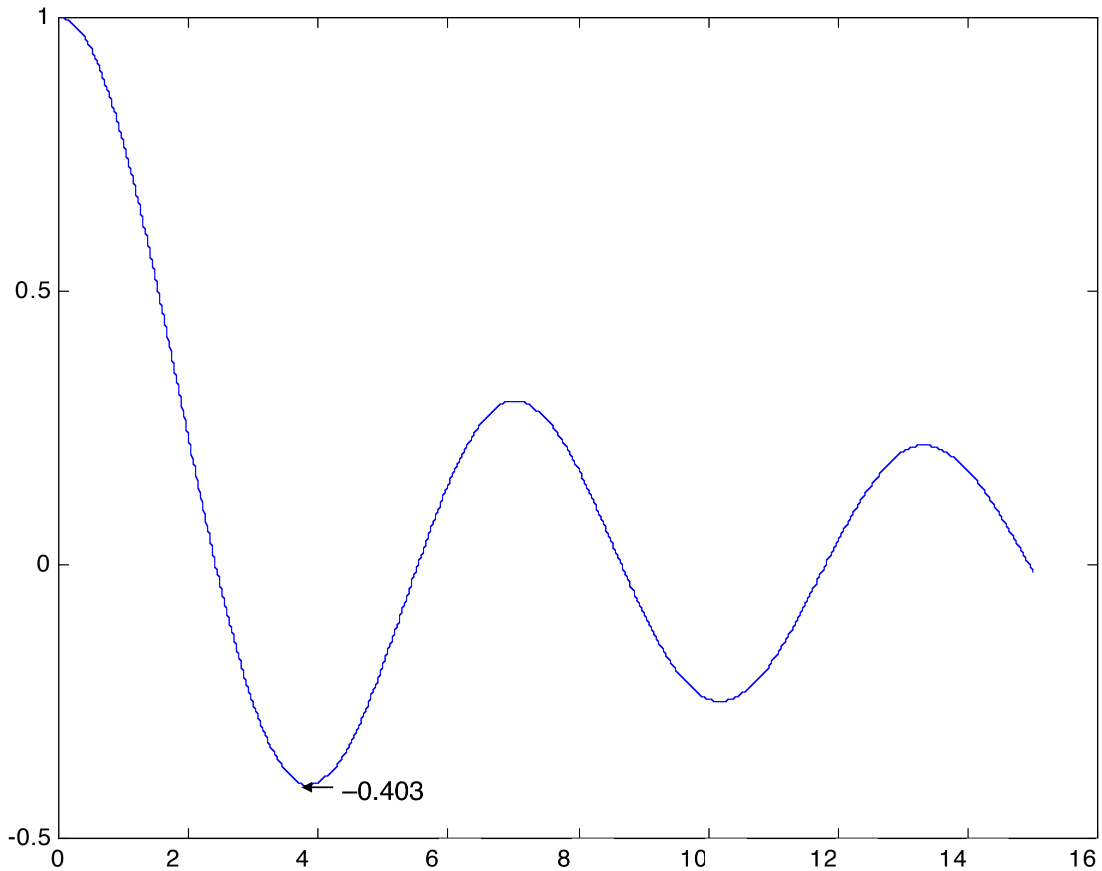
Hence

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

and in the case d=2

$$\rho(v) = \int_0^\infty J_0(sv)dG(s) \qquad \text{(Hankel transform)}$$

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,

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

The spherical

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

Corresponding variogram

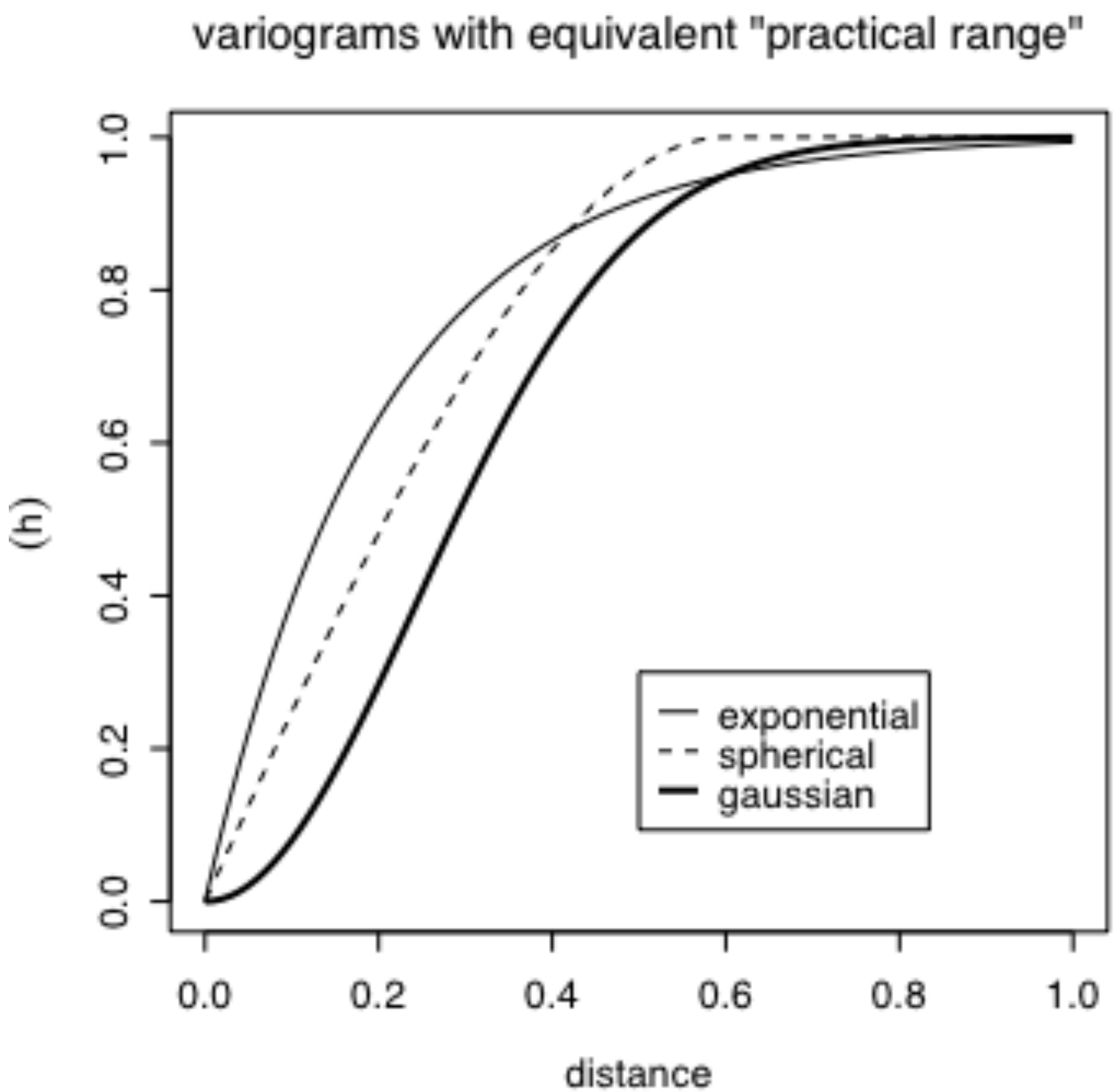
$$\underbrace{\boxed{\tau^2}}_{\text{nugget}} + \frac{\sigma^2}{2}\left(3\tfrac{t}{\phi} + \left(\tfrac{t}{\phi}\right)^3\right); \quad 0 \le t \le \phi$$

$$\underbrace{\boxed{\tau^2 + \sigma^2}}_{\text{sill}}; \quad t > \underbrace{\boxed{\phi}}_{\text{range}}$$

nugget → τ^2

sill → $\tau^2 + \sigma^2$

range → ϕ



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

The Matérn class

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

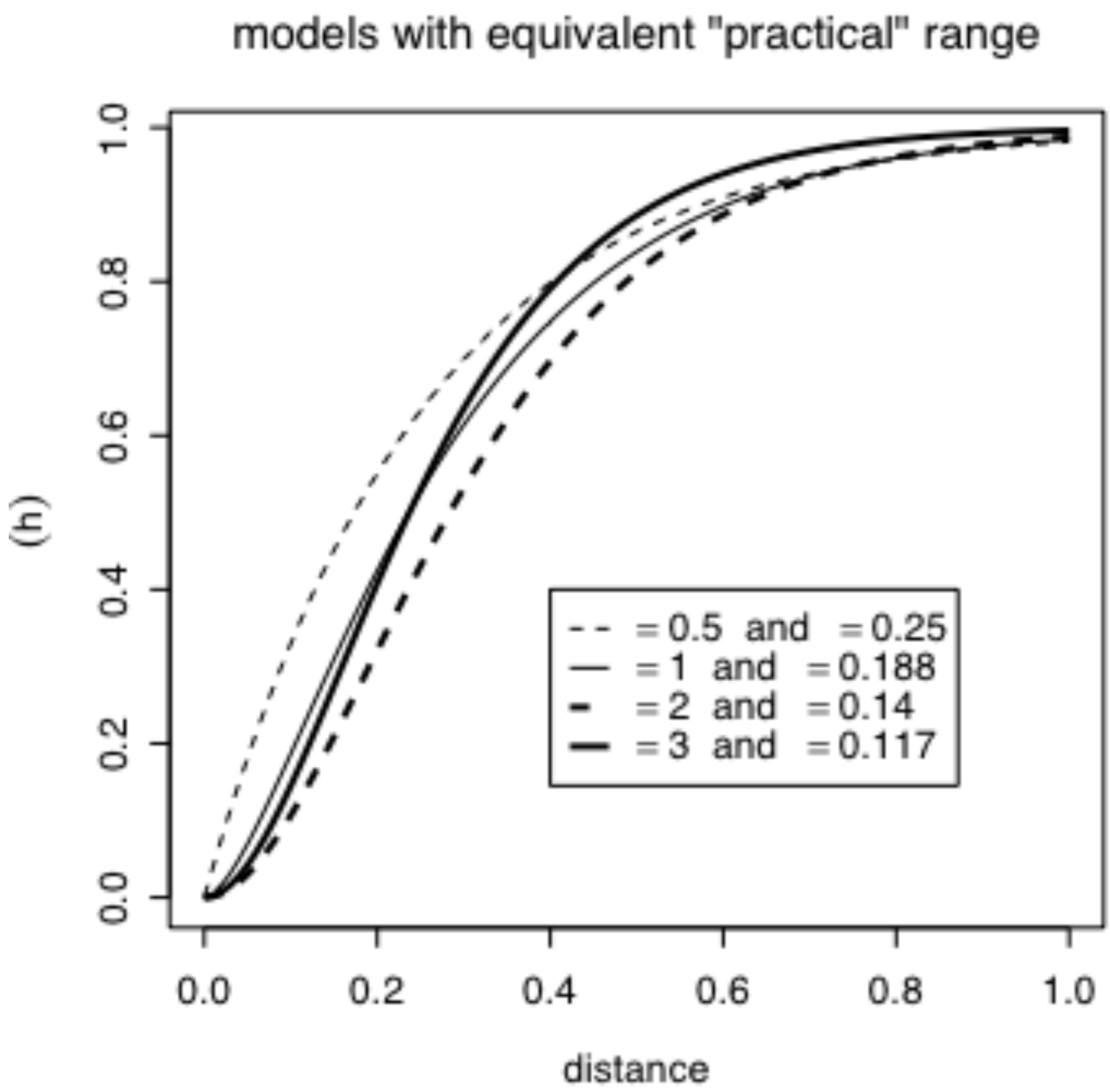
$$\rho(v) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{v}{\phi}\right)^{\kappa} K_\kappa\left(\frac{v}{\phi}\right)$$

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)
0.0
0.2
0.4
0.6
0.8
1.0
distance
= 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

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

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

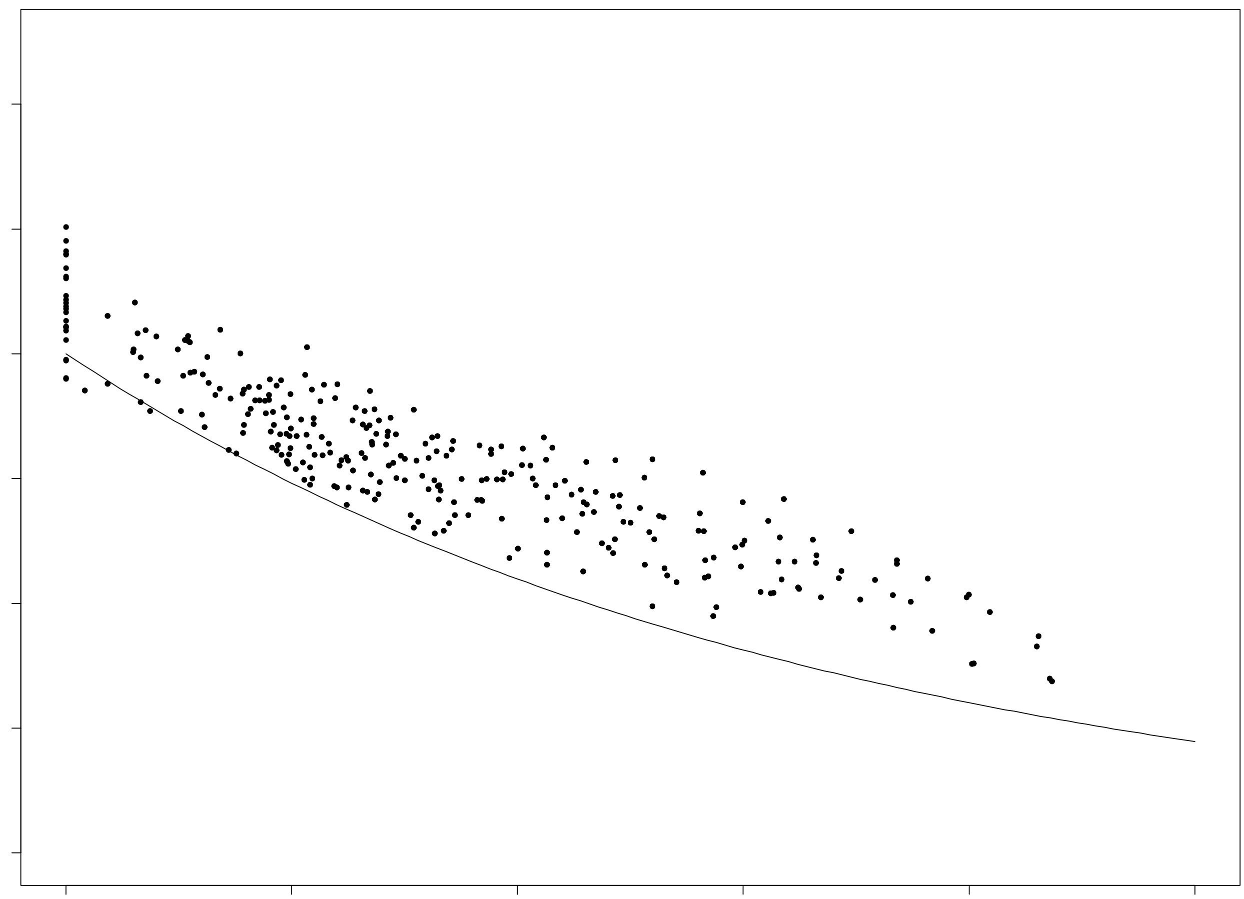
$$\ell(\mu, \alpha, \theta) = -\frac{n}{2}\log(2\pi\alpha) - \frac{1}{2}\log\det V(\theta) + \frac{1}{2\alpha}(Z - \mu)^T V(\theta)^{-1}(Z - \mu)$$

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

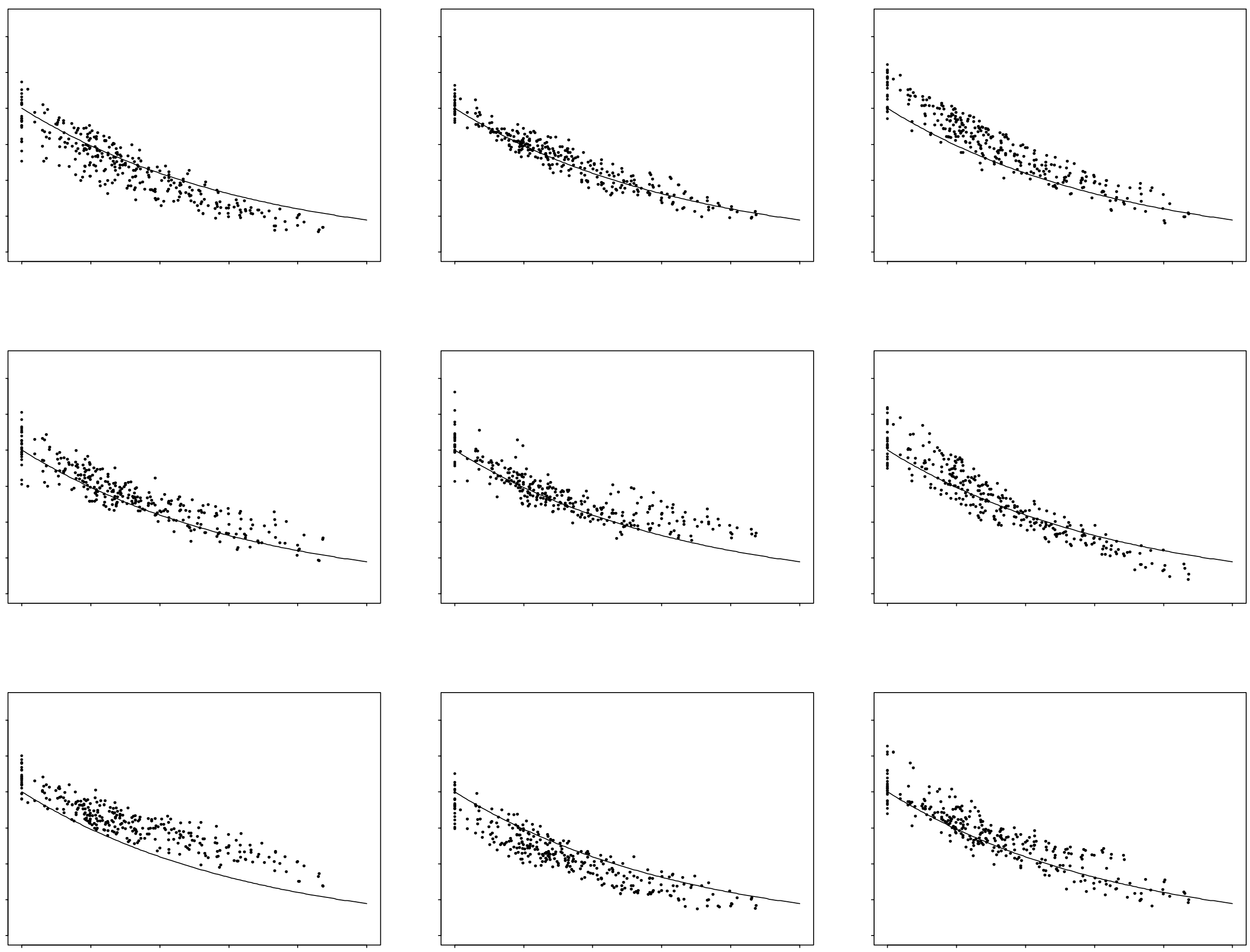
and θ maximizes the profile likelihood

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

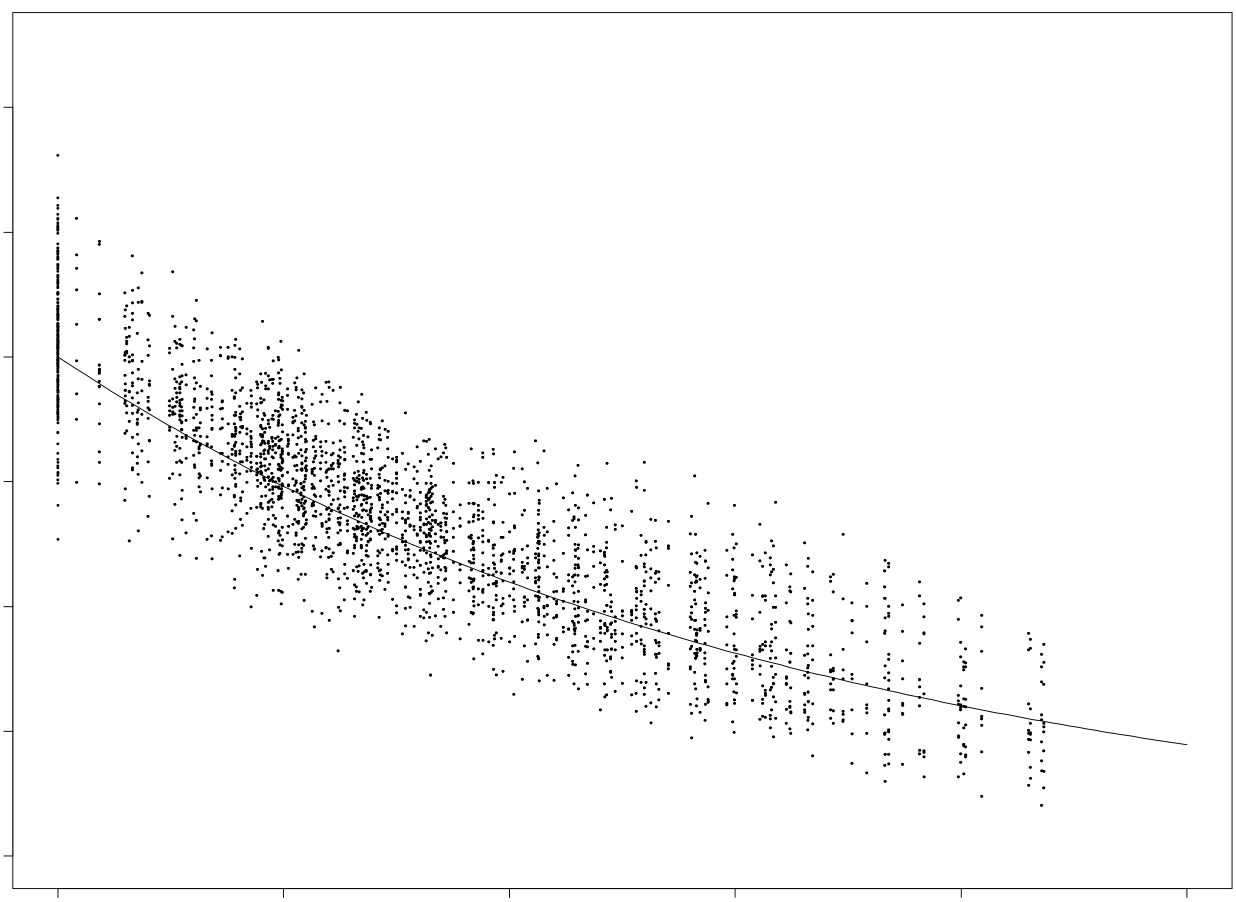
A peculiar ml fit



Some more fits



All together now...



Asymptotics

Increasing domain asymptotics: let region of interest grow. Station density stays the same

- Bad estimation at short distances, but effectively independent blocks far apart

Infill asymptotics: let station density grow, keeping region fixed.

- Good estimates at short distances. No effectively independent blocks, so technically trickier

Stein's result

Covariance functions C_0 and C_1 are *compatible* if their Gaussian measures are mutually absolutely continuous. Sample at $\{s_i, i=1,...,n\}$, predict at s (limit point of sampling points). Let $e_i(n)$ be kriging prediction error at s for C_i, and V_0 the variance under C_0 of some random variable.

If $\lim_n V_0(e_0(n))=0$, then

$$\lim_{n\to\infty} \frac{V_0(e_0(n))}{V_0(e_1(n))} = 1$$

The Fourier transform

$$g : R^d \rightarrow R$$

$$G(\omega) = \mathcal{F}(g) = \int g(s)\exp(i\omega^T s)ds$$

$$g(s) = \mathcal{F}^{-1}(G) = \frac{1}{(2\pi)^d}\int \exp(-i\omega^T s)G(\omega)d\omega$$

Properties of Fourier transforms

Convolution

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$$

Scaling

$$\mathcal{F}(f(a\bullet)) = \frac{1}{a}F(\omega / a)$$

Translation

$$\mathcal{F}(f(\bullet - b)) = \exp(ib)\mathcal{F}(f)$$

Parceval's theorem

$$\int f(s)^2 ds = \int \left| F(\omega) \right|^2 d\omega$$

Relates space integration to frequency integration. Decomposes variability.

Spectral representation

Stationary processes

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

Spectral process Y has stationary increments

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

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

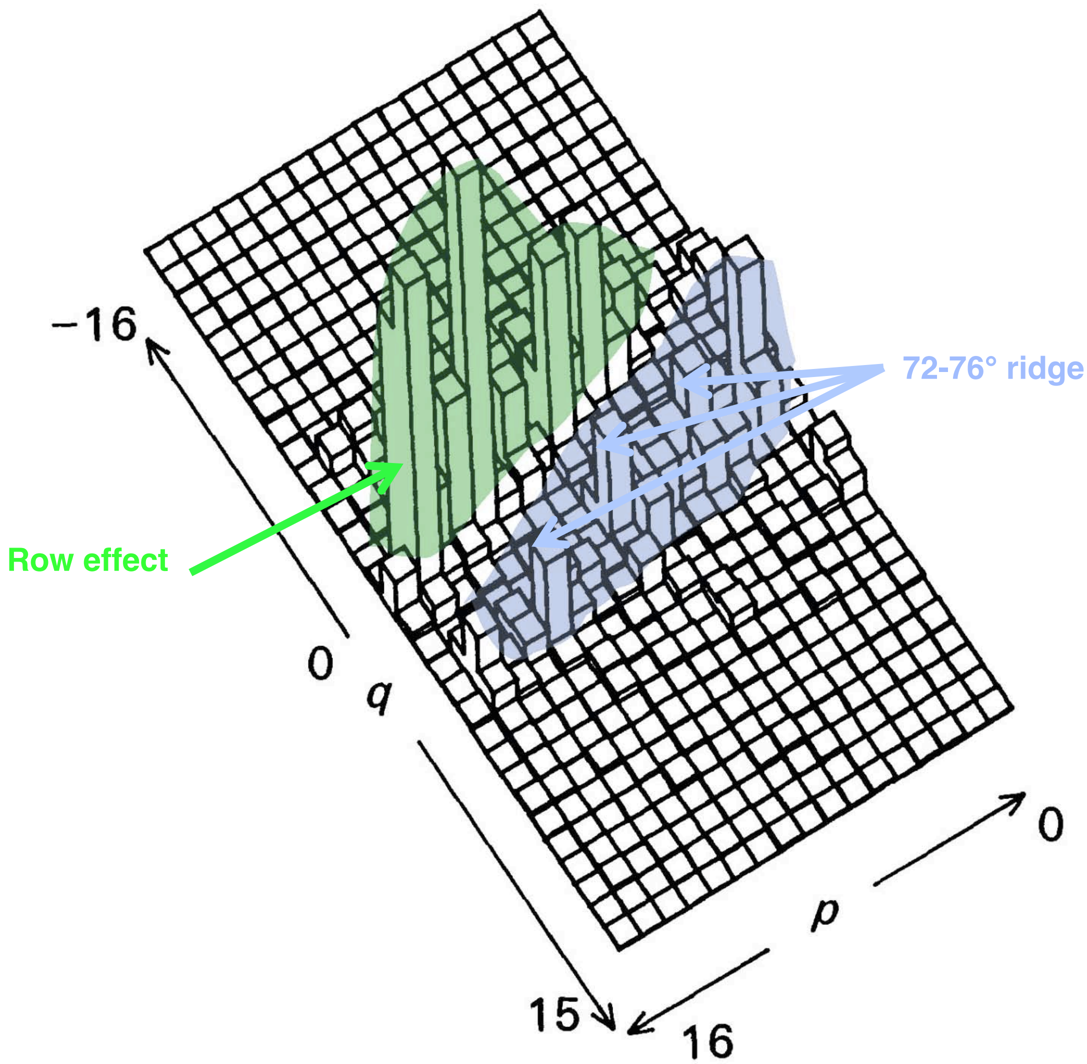
$$f(\omega)=\phi(\alpha^2+\|\omega\|^2)^{-\nu-1}$$

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

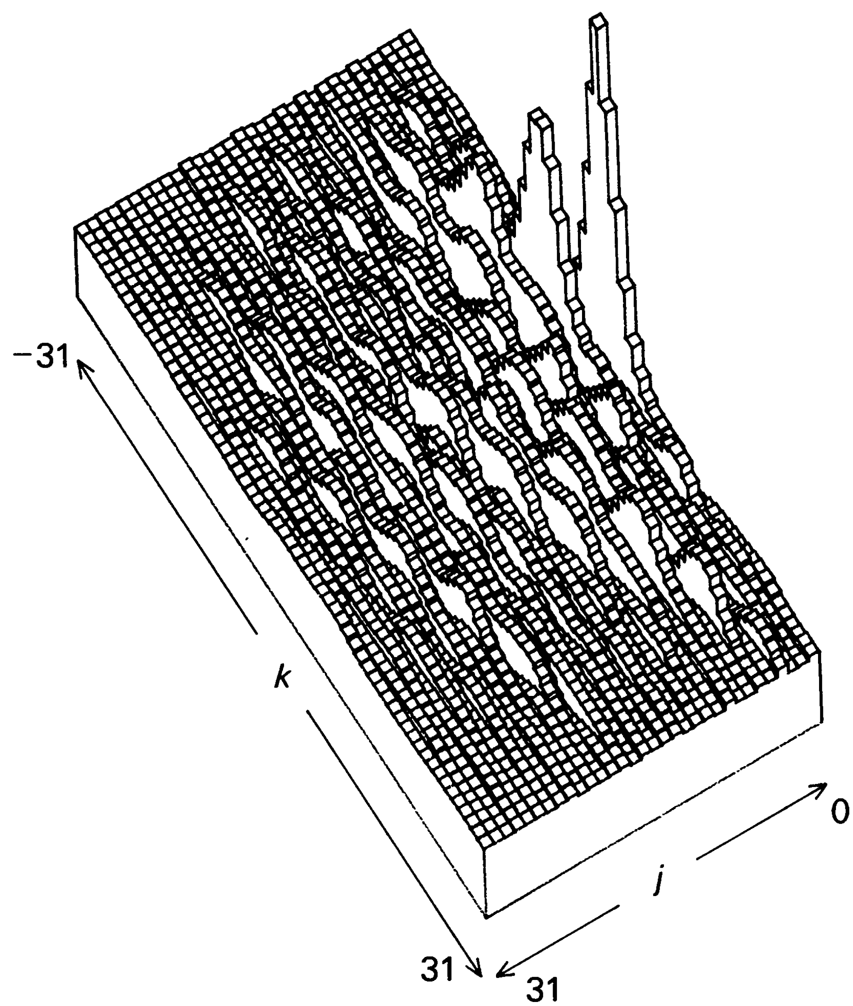
Thetford canopy heights

- 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



Correlation function



Global processes

Problems such as global warming require modeling of processes that take place on the globe (an oriented sphere). Optimal prediction of quantities such as global mean temperature need models for global covariances.

Note: spherical covariances can take values in [-1,1]–not just imbedded in R^3.

Also, stationarity and isotropy are identical concepts on the sphere.

Isotropic covariances on the sphere

Isotropic covariances on a sphere are of the form

$$C(p,q) = \sum_{i=0}^{\infty} a_i P_i(\cos\gamma_{pq})$$

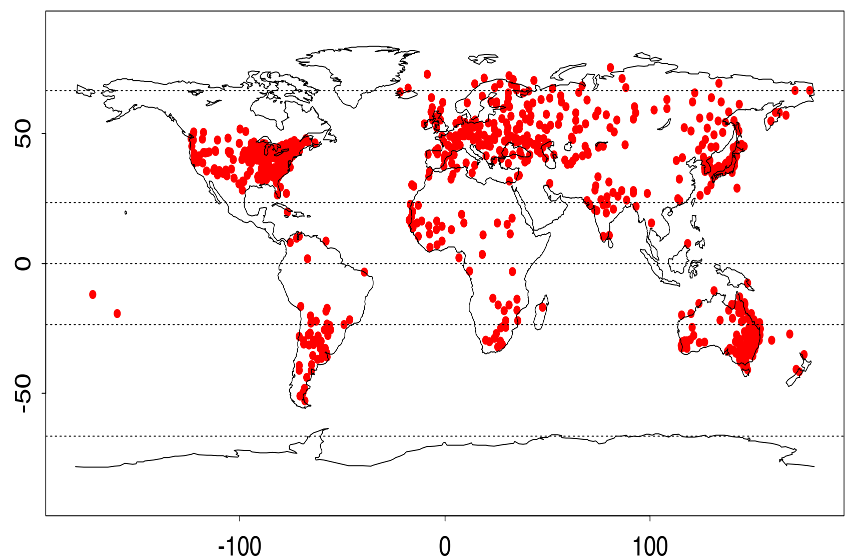
where p and q are directions, γ_{pq} the angle between them, and P_i the Legendre polynomials.

Example: $a_i=(2i+1)\rho^i$

$$C(p,q) = \frac{1-\rho^2}{1-2\rho\cos\gamma_{pq}+\rho^2} - 1$$

Global temperature

Global Historical Climatology Network 7280 stations with at least 10 years of data. Subset with 839 stations with data 1950-1991 selected.



Isotropic correlations