Hierachical statistical models (HM)

Data model: $[Z|Y,\theta]$

Process model: $[Y|\theta]$

Parameter model: $[\theta]$

For precipitation (Thea):

Data model: $[Z|Y,\theta]$ Distribution for observations given true precipitation Independent $Z(s_i) = N(Y(s_i), \sigma_d^2)$

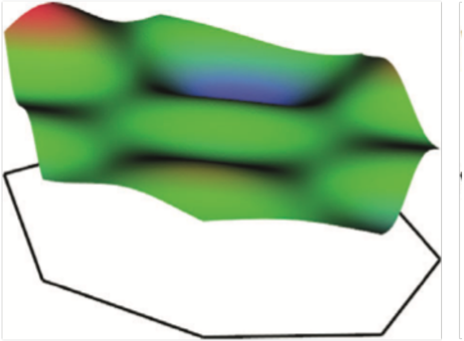
Process model: $[Y|\theta]$ Distribution for precipitation. $Y(s) \sim GRF(\theta_p)$

Parameter model: $[\theta]$ Prior for parameters. $[\theta] = [\sigma_d^2][\theta_p]$

- Bayesian HM (BHM): θ considered random variable, given prior
- Empirical HM (EHM): θ considered fixed, but unknown

Part 1: Geostatistical processes

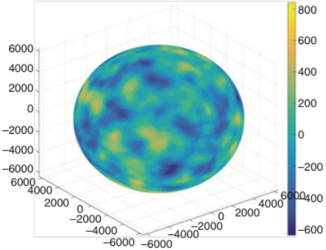
We need a stochastic models for random variables that are defined for a domain (in space) for the process model.

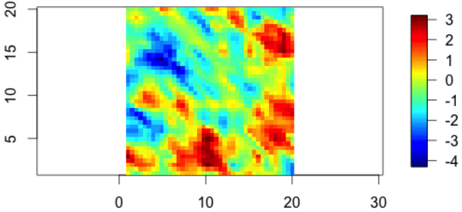


Gaussian Random Field (GRF)

The random field $Y(s)$, $s \in D$ (f.ex. in R^2) is a Gaussian Random field if, for any n, and any set of locations $s_1, s_2, \ldots, s_n$, all finite collections $(Y(s_1), Y(s_2), \ldots Y(s_n))$ are multivariate Normal distributed.

(a)





Covariance function

- Need a way to specify the covariance matrix Σ for any set of locations $(s_1, s_2, \ldots, s_n)$
- Σ has to be positive (semi) definite.
- Want locations close to have higher correlation that locations further away.

One valid correlation function:

Exponential correlation function

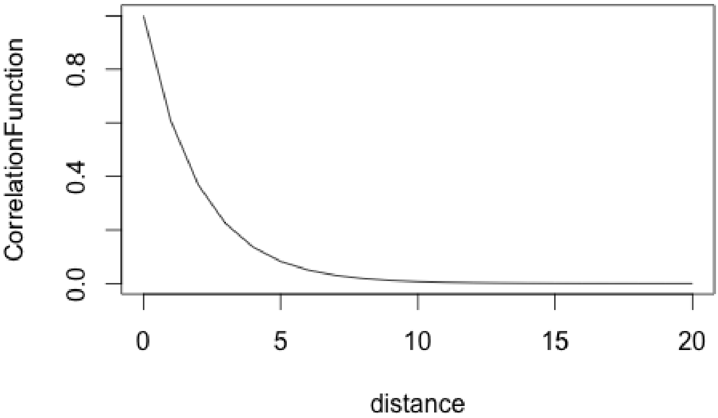
$$Corr(Y(s_1), Y(s_2) = \exp(-d(s_1, s_2)/\theta_1)$$

where $d(s_1, s_2)$ is the (Euclidian) distance between s_1 and s_2.

Exponential covariance function :

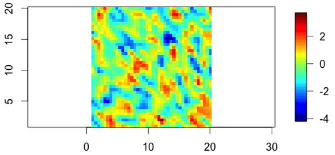
$$C(Y(s_1), Y(s_2)) = \sigma_1^2 Corr(Y(s_1), Y(s_2))$$

Exponential correlation function, $\theta = 2$

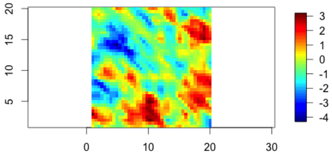


Examples samples from GRFs

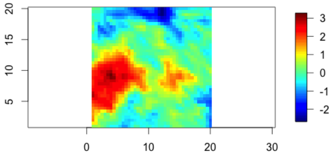
$\theta_1 = 0.5$ and $\sigma_1^2 = 2$



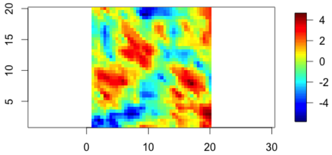
$\theta_1 = 2$ and $\sigma_1^2 = 2$



$\theta_1 = 10$ and $\sigma_1^2 = 2$



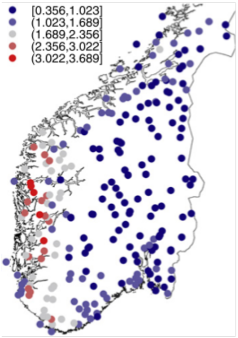
$\theta_1 = 2$ and $\sigma_1^2 = 4$



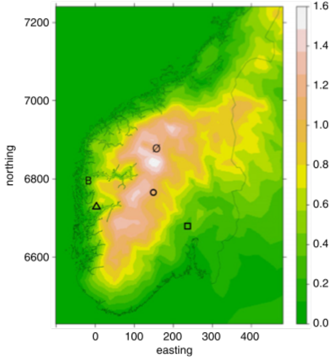
Annual precipitation in Southern Norway

Can this be a sample from a GRF (as previous page)?

R. Ingebrigtsen et al. / Spatial Statistics 8 (2014) 20–38 23



(a) Annual precipitation (m).



(b) Smoothed elevation (km).

Fig. 1. The precipitation data are from the Norwegian Meteorological Institute and consist of observations from 233 weather stations in southern Norway. The smoothed elevation model is based on GLOBE and computed on a triangulation of the domain. The locations of the three weather stations Kvamskogen (△), Hemsedal (∘), and Hønefoss (□) are indicated on the elevation map in (b). These stations will receive closer examination in Section 5.1. Also indicated on the map is Brekke (B) and Øygarden (Ø), these two weather stations have, respectively, the maximum and minimum annual normal in Norway. The coordinate reference system is UTM33 in km.

Today

Random Fields:

- Stationary
- Isotopic

Covariance functions:

- Power exponential
- Matern

Stationarity

Second order (weakly) stationary

The random field $Y(s)$ for $s \in D$ is second order stationary if

- $E(Y(s)) = \mu$ for all $s \in D$
- $Cov(Y(s+h), Y(s)) = C_Y(h)$ for all $s, s+h \in D$
- I.e. the covariance only depends on the vector difference between the locations.

Strong stationarity

Let $F(Y(s_1), Y(s_2), \dots, Y(s_n))$ be the cdf of $Y(s_1), Y(s_2), \dots, Y(s_n)$. The random field $Y(s)$ is strongly stationary if $F(Y(s_1), Y(s_2), \dots, Y(s_n)) = F(Y(s_1+h), Y(s_2+h), \dots, Y(s_n+h))$ for all s_i and $s_i + h \in D$

- General: Strong stationarity $\Rightarrow$ second order stationarity.
- For GRF: Strong stationarity $\Leftrightarrow$ second order stationarity

Isotropy

A subclass of weakly stationary covariance functions are:

Isotropic covaraince function

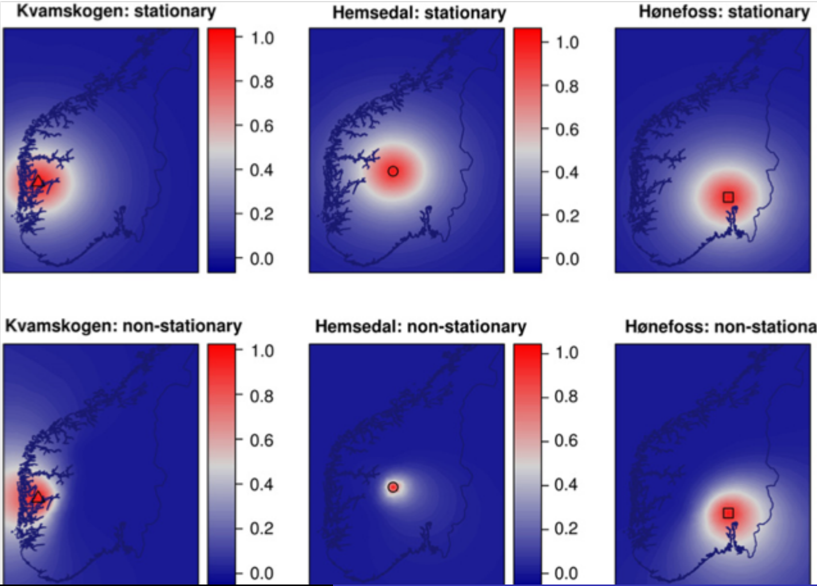
A covariance function is isotropic if it only depends on the distance between the locations:

$$C_Y(s, s+h) = \|h\|$$

- Not isotropic = anisotropic
- The exponential covariance function is isotropic.

Correlation for 3 sites with two models

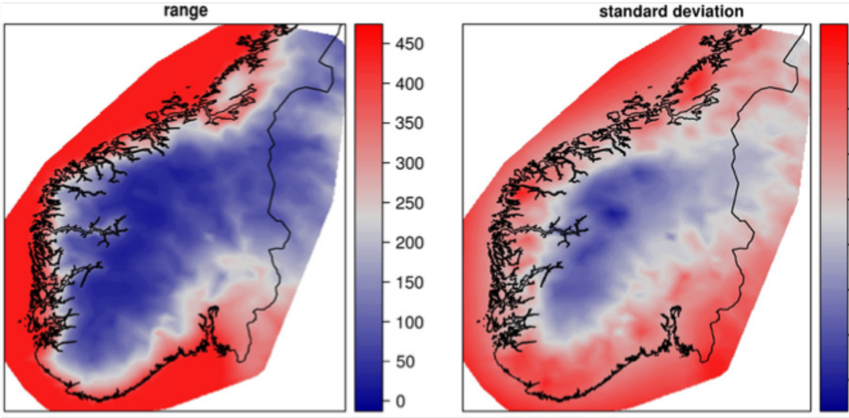
R. Ingebrigtsen et al. / Spatial Statistics 8 (2014) 20–38



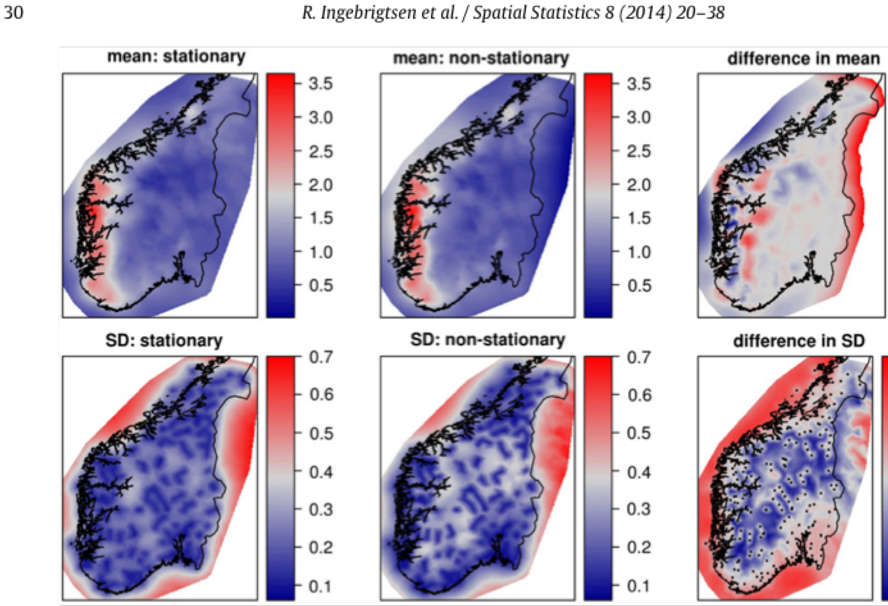
Range and standard deviation

Elevation is used to model the spatial dependency:

R. Ingebrigtsen et al. / Spatial Statistics 8 (2014) 20–38



For predictions of $Y(s)$



Variogram

- An alternative to covariance functions measuring spatial dependency.

Stationary variogram

$$\gamma_Y(h) = \frac{1}{2} Var(Y(s+h) - Y(s))$$

- Stationary covariance function $\Rightarrow$ stationary variogram (pg 128-).

Common covariance functions

d is distance

- Power exponential: $C_Y(d) = \sigma_1^2 \exp(-(d/\theta_1)^{\theta_2})$
- Matern: $C_Y(d) = \sigma_1^2(2^{\nu-1}\Gamma(\nu))^{-1}(d/\theta_1)^{\nu} K_{\nu}(d/\theta_1)$
  - A Gaussian process with Matérn covariance has sample paths that are $\nu - 1$ times differentiable.
  - In Cressie & Wikle: $\nu \to \theta_2$
- θ_1: Range
- θ_2: Smoothness

Two spatial processes Z(s) and Y(s)

For a while we have assumed

- Known parameters, i.e. mean μ and covariance function / Σ
- Perfect observations (observe $Y(s)$) $\Rightarrow Y(s) = Z(s)$

What if we do not have perfect observations?

- Much of the discussion about HM and kriging is related to this.

Micro-scale variation

Note: Covariance functions (and variograms) in Chp 4.1 are (often) written as:

$$C_Y(d) = \sigma_0^2 I(d = 0) + \mathrm{Our}\, C_Y(d)$$

Microscale variation: $\sigma_0^2 I(d = 0)$ is very small scale variation in $Y(s)$

Nugget effect: $\sigma_0^2 + \sigma_\epsilon^2$: Due to both microscale variation and measurement uncertainty.

Covariance function: With microscale variation, when $h \to 0$: $Corr(Y(s), Y(s + h)) < 1$ (see e.g. figure 4.2 and 4.3)

Before Tuesday

- Read 4.1.1 and 4.1.2