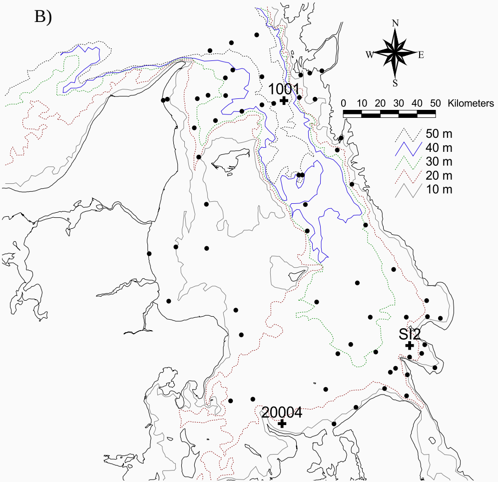
Geostatistical Design

Søren Lophaven

The Kattegat area

Location of monitoring stations



Sampling


The design problem

$$x_i \in A, \quad i = 1, \cdots, n$$

How should we choose sampling locations x_i:

- For optimal estimation of parameters in the geostatistical model
- For optimal prediction of $S(x)$

Different design situations:

- Retrospective: Add to, or delete from, an existing set of sampling locations
- Prospective: Choose optimal positions for a new set of sampling locations

Gaussian geostatistics

$$Y_i = S(x_i) + Z_i, \qquad i = 1, \cdots, n$$

- $S(x_i)$ is a Gaussian process with

$$\mathrm{E}[S(x)] = \mu$$
$$\mathrm{Var}[S(x)] = \sigma^2$$

and correlation function

$$\rho(u) = \mathrm{Corr}[S(x), S(x')]$$

- Z_i are IID with

$$Z_i \in N(0, \tau^2)$$

Parameter uncertainty is usually ignored in traditional

Plug-in prediction

The predictor that minimizes $\mathrm{E}[(\hat{S}(x) - S(x))^2]$ is called the kriging predictor. It can be shown that the kriging predictor for $T = S(x_0)$ is

$$\hat{T} = \mu + \sigma^2 r^T (\tau^2 I + \sigma^2 R)^{-1} (y - \mu I)$$

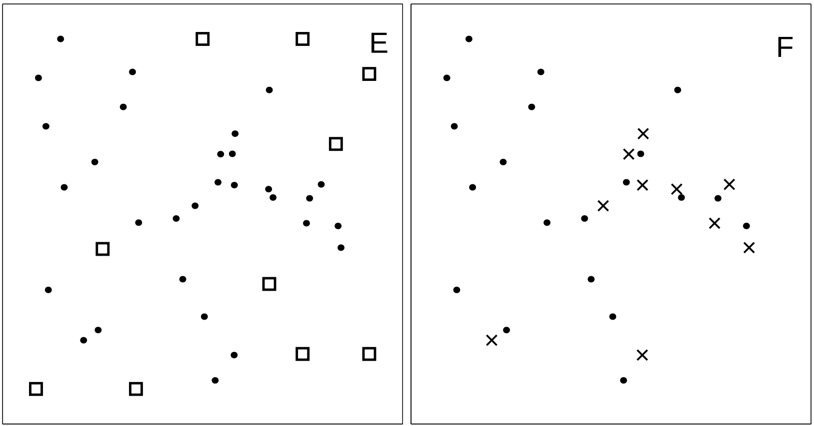
with prediction variance

$$\mathrm{Var}[T|y] = \sigma^2 - \sigma^2 r^T (\tau^2 I + \sigma^2 R)^{-1} \sigma^2 r$$

where

- R is a symmetric $n \times n$ matrix with elements $\rho(\| x_i - x_j \|)$
- r is a $n \times 1$ vector with elements $\rho(\| x_0 - x_i \|)$

Design - Spatial prediction



Parameter estimation - The semivariogram

- Definition:

$$\gamma(u) = \frac{1}{2}\mathrm{Var}[Y(x) - Y(x')] = \frac{1}{2}\mathrm{E}[(Y(x) - Y(x'))^2], \quad u = \| x - x' \|$$

- The relationship between the semivariogram and the correlation function:

$$\gamma(u) = \tau^2 + \sigma^2(1 - \rho(u))$$

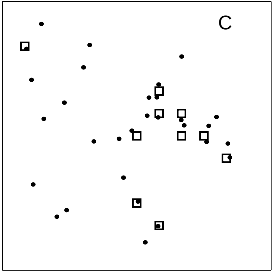
- The sample semivariogram:

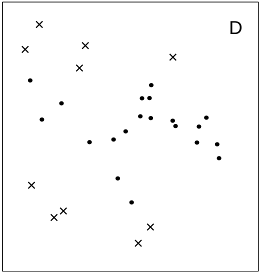
$$\hat{\gamma}(u) = \frac{1}{2N(u)} \sum_{i=1}^{N(u)} (Y(x) - Y(x'))^2$$

Design - Parameter estimation

Warrick, A. and Myers, D. (1987). Optimization of sampling locations for variogram calculations. *Water Resources Research*, **23**, 496–500.

Zimmermann, D.L. and Homer, K.E. (1991). A network design criterion for estimating selected attributes of the semivariogram. *Environmetrics*, **2**(4), 425-441.





Geostatistical design - A Bayesian approach

Idea: Design for prediction, but at the same time allow for parameter uncertainty. Thus, we work in the Bayesian paradigm and focus on minimizing the variance of the predictive distribution.

Given data y and prior distribution $pr(\theta) = pr(\beta, \sigma^2, \phi, \tau^2)$. The posterior distribution is:

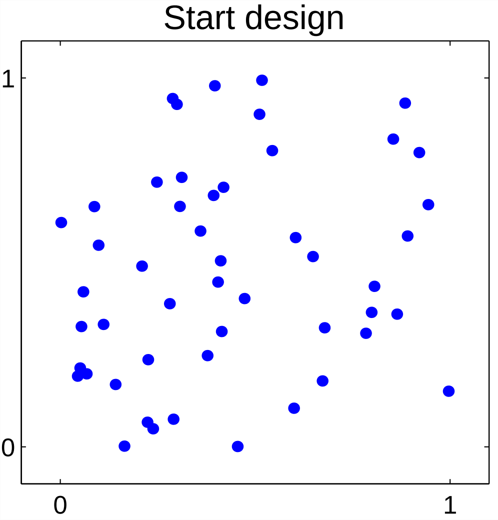
$$pr(\beta, \sigma^2, \phi, \tau^2 | y) \propto$$

$$pr(\beta, \sigma^2, \phi, \tau^2) |\tau^2 I + \sigma^2 R|^{\frac{1}{2}} \exp\left(\frac{1}{2}(y - F\beta)'(\tau^2 I + \sigma^2 R)^{-1}(y - F\beta)\right)$$

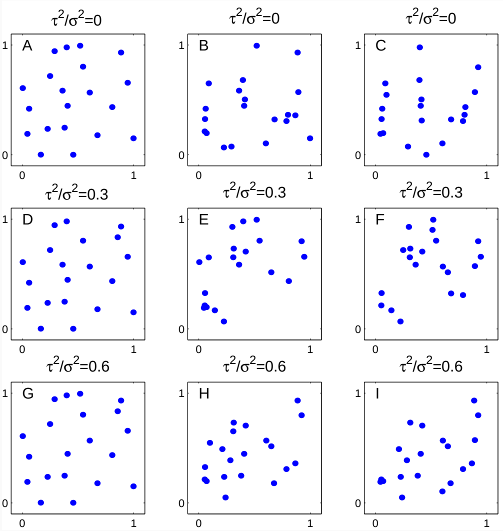
The predictive distribution $pr(y_0|y)$ is:

$$pr(y_0|y) = \int pr(y_0|y, \theta) pr(\theta|y) \mathrm{d}\theta$$

Retrospective design

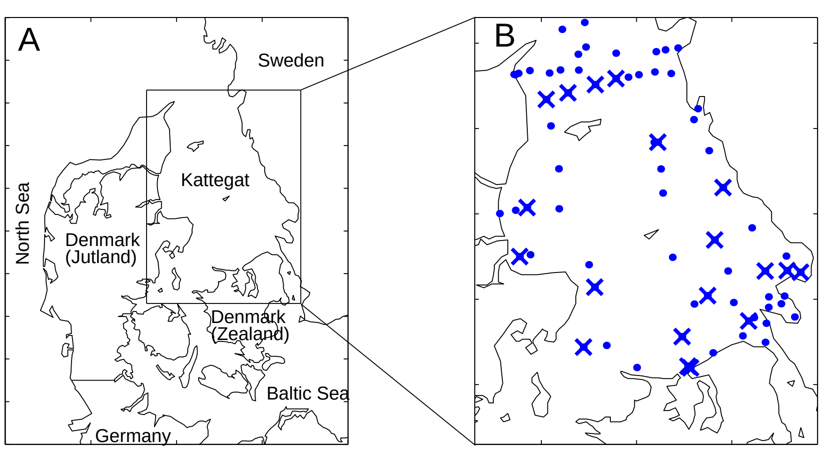


Retrospective design results

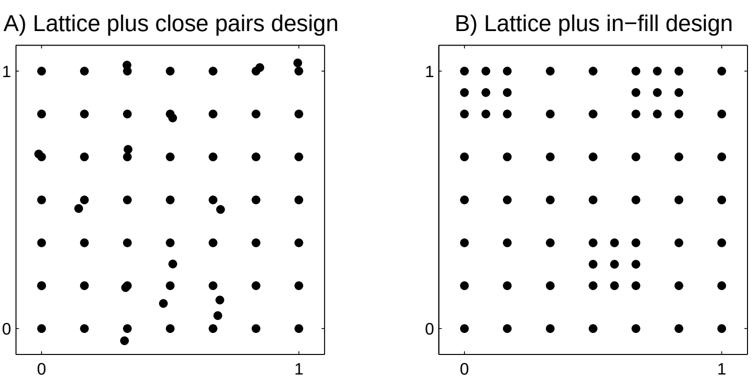


Environmental monitoring example

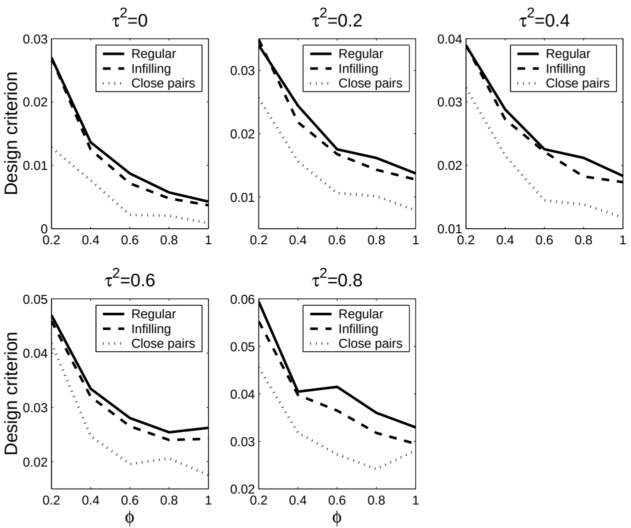
Fitted model has linear trend and exponential semivariogram



Prospective design



Prospective design results



Closing remarks

In general:

- Usually, the primary goal of a geostatistical analysis in spatial prediction
- Parameter estimation is usually not of direct interest
- But parameter uncertainty can have a material impact on prediction

About geostatistical designs:

- To estimate model parameters efficiently the design need to include a wide range of inter-point distances
- Therefore, spatially regular designs need to be tempered by inclusion of some close pairs of points