Design and Analysis of Environmental Monitoring Programs

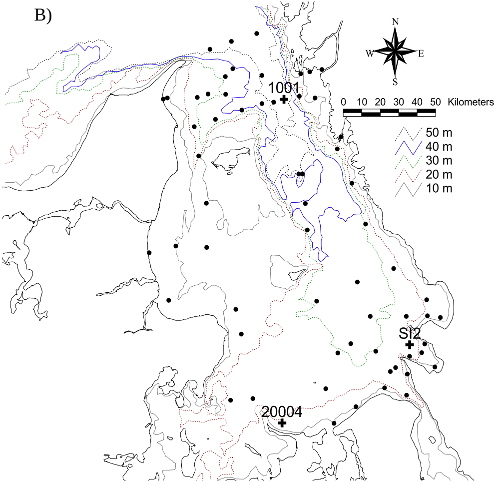
Søren Lophaven

Outline

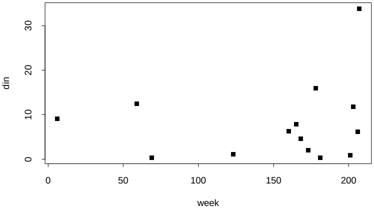
- Introduction
    - Study area
    - Data presentation
    - Background
    - The aim of the thesis
- Geostatistical design
- Space-time modelling
- Conclusions

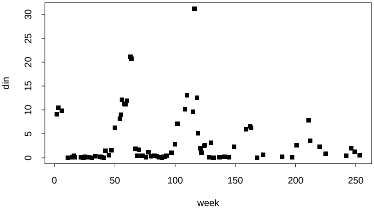
The Kattegat area

Location of monitoring stations

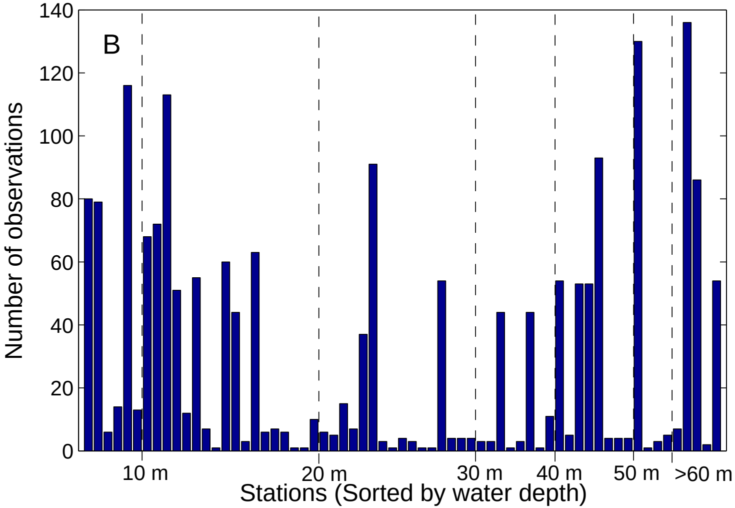


Sampling DIN at individual stations

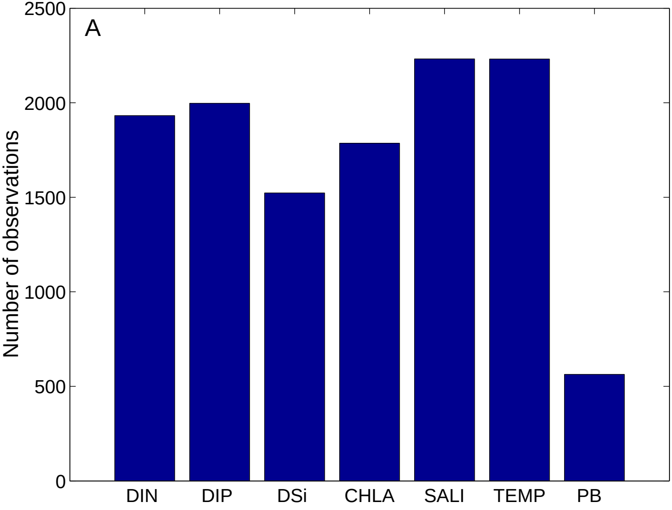




Number of DIN samples at stations



The total number of observations



Sampling


Background

- Oxygen deficiency, covering large areas of the Danish estuaries
- Implementation of the Action Plan on the Aquatic Environment in 1987
- In connection to that, a national monitoring program was established

Monitoring objectives:

- Increase the knowledge on the processes in the marine environment
- Provide better assessment of effects of anthropogenic stresses

The aim of the thesis

Improve reporting the state of the environment in the Kattegat by means of statistical methods.

- Design of monitoring networks
- Modelling of space-time phenomena

Focus has been on:

- Simple methods
- Methods that are able to handle missing values
- Methods that are generally applicable

The geostatistical model

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

$S(x_i)$ is a Gaussian process with

$$\mathrm{E}[S(x)] = \mu = F\beta$$

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

and correlation function

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

Z_i are IID with

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

This implies that:

$$Y \in N(F\beta, \sigma^2 R + \tau^2 I)$$

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 $\rho(\| x_0 - x_i \|)$

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

Geostatistical design - A Bayesian approach

The idea is to use a Bayesian formulation of the geostatistical model. This incorporates parameter uncertainty in 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:

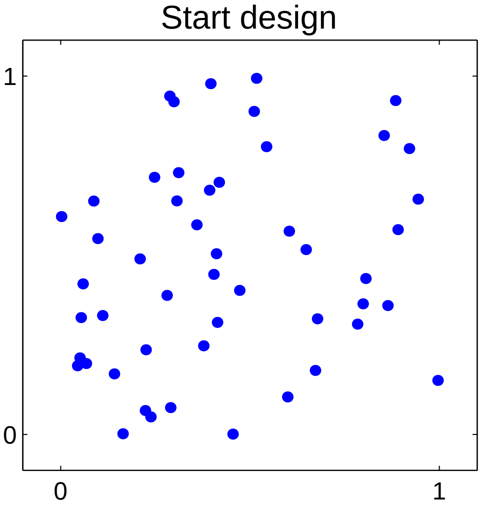
$$pr(\beta, \sigma^2, \phi, \tau^2 | y) \propto$$

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

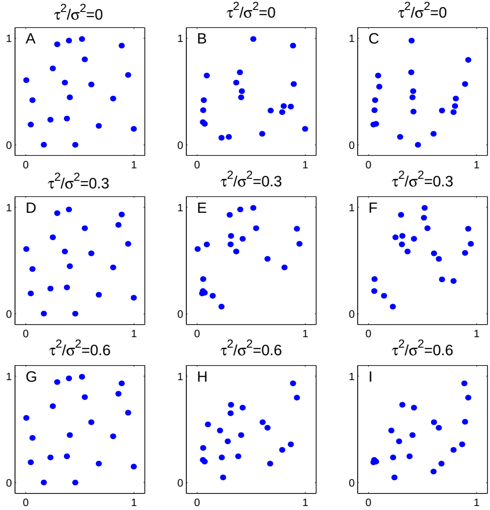
The predictive distribution $pr(y_0|y)$ is:

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

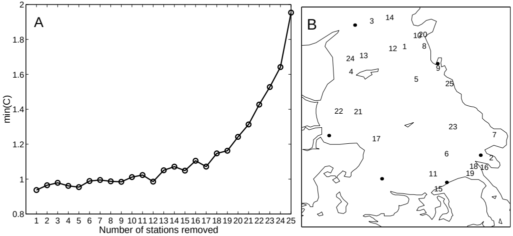
Retrospective design results

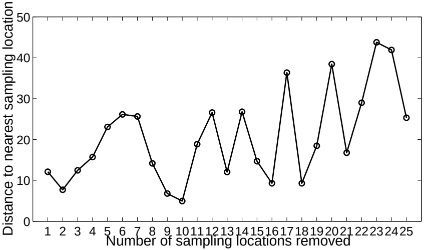


Retrospective design results



Design - Kattegat





Space-time modelling of DIN - Approach 1

Model:

$$Z_{kl}(s,t) = station_k + week_l + \varepsilon(s), \qquad \varepsilon(s) \sim \mathrm{N}(0, \sigma^2 \boldsymbol{\Sigma})$$

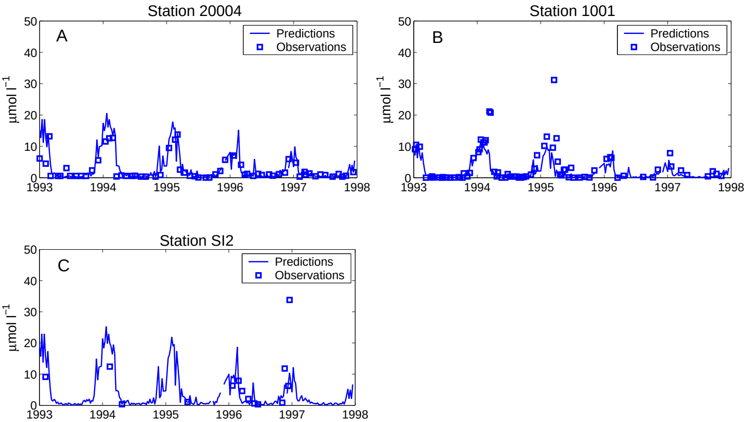
Prediction:

$$\hat{Z}(s,t) = \boldsymbol{x_p}^T \hat{\beta} + \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Z} - \boldsymbol{X}\hat{\beta})$$

Prediction variance:

$$\mathrm{V}(\hat{Z}(s,t)) =$$
$$\sigma^2 - \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c} + (\boldsymbol{x_p} - \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c})^T (\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} (\boldsymbol{x_p} - \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c})$$

Space-time modelling of DIN - Approach 1



Space-time modelling of DIN - Approach 2

Model:

$$Z(s,t) = \mu(t) + \varepsilon(s,t)$$

Mean component:

$$\mu_{ij}(t) = \alpha_i + \beta_j + \delta_1 \sin\left(\frac{2\pi t}{52} + \varphi_1\right) + \delta_2 \sin\left(\frac{2\pi t}{26} + \varphi_2\right)$$

Residual component:

$$\mathrm{E}(\varepsilon(s,t)) = 0$$

$$\mathrm{C}_{st}(h_s, h_t) = \mathrm{Cov}(\varepsilon(s+h_s, t+h_t), \varepsilon(s,t))$$

Space-time modelling of DIN - Approach 2

The product model:

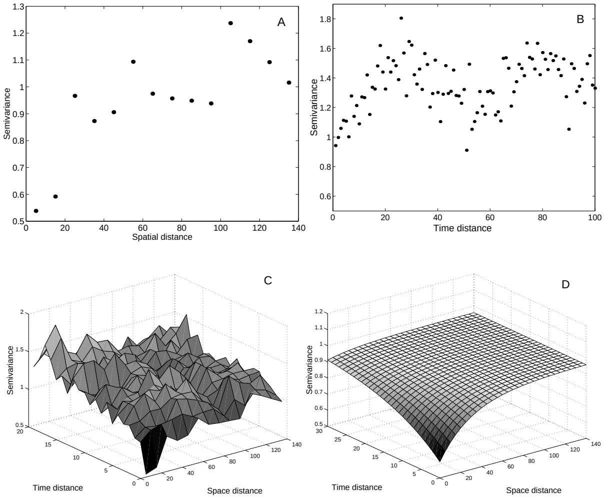
$$C_{st}(h_s, h_t) = C_s(h_s)C_t(h_t)$$

$$\gamma_s(h_s) = C_s(0) - C_s(h_s)$$

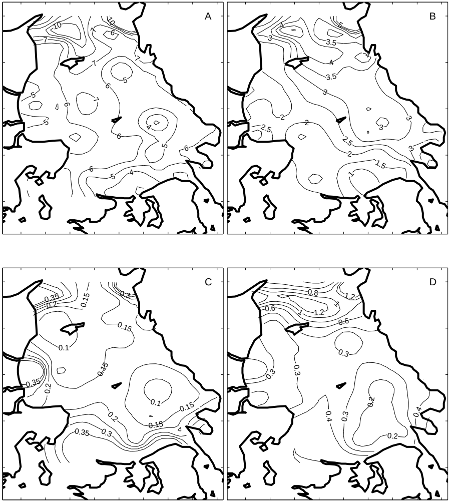
$$\gamma_t(h_t) = C_t(0) - C_t(h_t)$$

$$\gamma_{st}(h_s, h_t) = C_s(0)\gamma_t(h_t) + C_t(0)\gamma_s(h_s) - \gamma_s(h_s)\gamma_t(h_t)$$

Space-time modelling of DIN - Approach 2



Space-time modelling of DIN - Approach 2



Conclusions

- The analysis indicates that the number of monitoring stations in the Kattegat could be reduced
- A reduction in the number of stations should be accompanied by an increase in the sampling frequency at the remaining stations, and intensive spatial sampling in some periods of the year
- The modelling results could be physically interpreted, and used for improving reporting the state of the marine environment in the Kattegat
- The results could also be used in combination with hydrodynamic ecosystem models, and thereby advance the knowledge of the biochemical processes in the marine environment, and reduce uncertainties of regional nutrient and carbon budgets