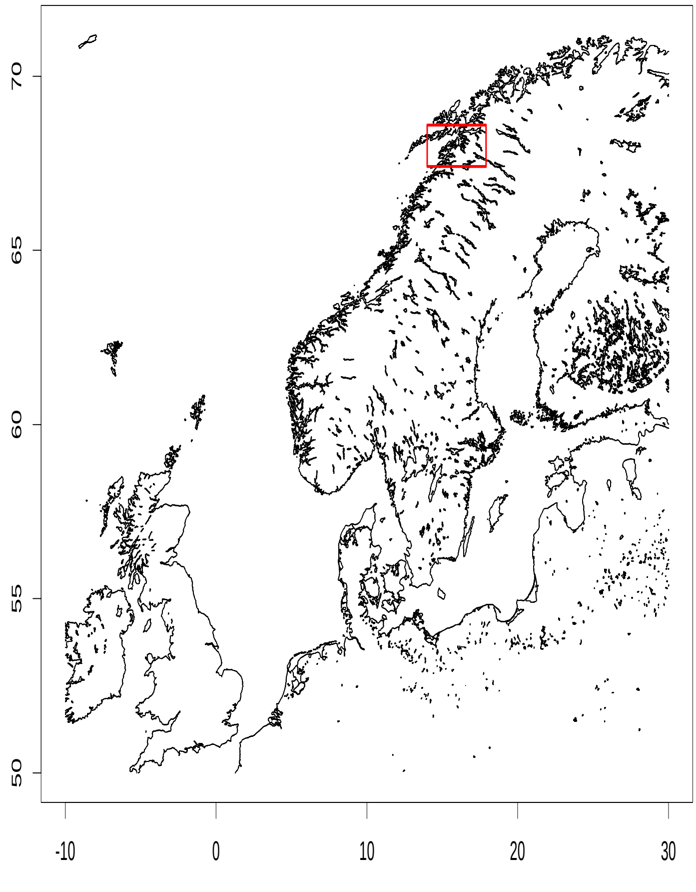
Spatial covariance modelling in a complex coastal domain by multidimensional scaling

Anders Løland* & Gudmund Høst
Norwegian Computing Center

June 2002

*anders.loland@nr.no

Norwegian Spring Spawning Herring



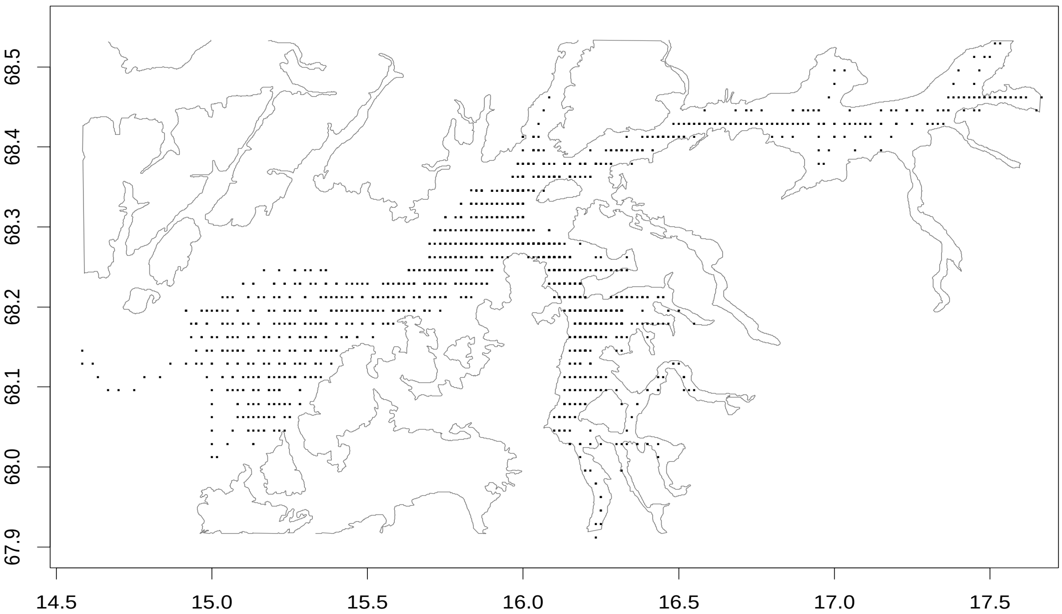
- The entire population of Herring over-winters in Vestfjord in northern Norway.

Norwegian Spring Spawning Herring

- Big population: approx. 3 herring per inhabitant of the earth!

Data

- An acoustic measure of the herring abundance was recorded at 742 data locations in December 1996.



- The spatial dependence often goes through the connecting water body.

- Problem: Complex geometry of the fjord system.

$\Rightarrow$ non-stationary spatial covariance

Water distance

- In aquatic studies, spatial interactions may be both easier to interpret and to quantify by using water distance (Rathbun 1998) than by using geographic distance.

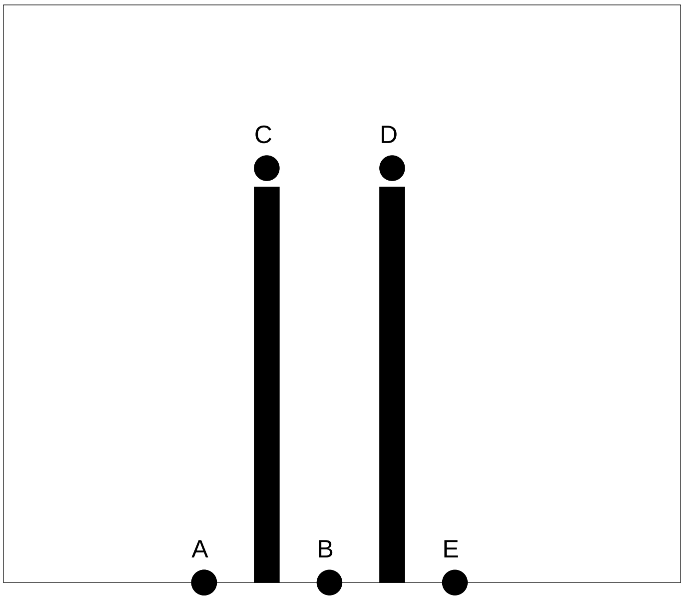
- Water distance: "the shortest path between those two sites that may be traversed entirely over water".

- Problem I: water distances may be non-Euclidean

  $\Rightarrow$ covariance and variogram functions may be invalid.

- Problem II: calculation of water distances for many spatial locations is computationally expensive.

Problem I: non-Euclidean water distance



Water distance metric and Gaussian covariance function with a particular parameter set

⇒ the matrix of covariances between these five locations has negative eigenvalues

⇒ covariance and variogram functions are not necessarily valid.

⇒ can not use for Kriging.

Spatial model

Assume measurements $Z(\boldsymbol{s}_i)$ of spatial random field $Z(\boldsymbol{s})$ at locations $\boldsymbol{s}_i \in \mathcal{G} \in \mathbb{R}^d$, $i = 1, \ldots, n$. We take $\mathsf{E}[Z(\boldsymbol{s}_i)]$ to be a constant.

Define the covariance function $C(\cdot)$ as

$$C(\boldsymbol{s}_i - \boldsymbol{s}_j) = \mathsf{COV}[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)].$$

This means that the covariance depends on the difference $\boldsymbol{s}_i - \boldsymbol{s}_j$ only and the process is *second-order stationary*. The *variogram* $2\gamma(\cdot)$ is

$$2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = \mathsf{VAR}[Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)].$$

Problem: find $\mathcal{G}$ and d, to approx. water distance space.

Water distance approximation

A Euclidean approximation to water distance
$\Rightarrow$ valid variograms and covariograms. Our method:

1. Decompose the coastal domain into a grid based on convex polygons.

2. Compute the water distances between all grid nodes, using the grid. $\Rightarrow$ Exact water distance in the grid.

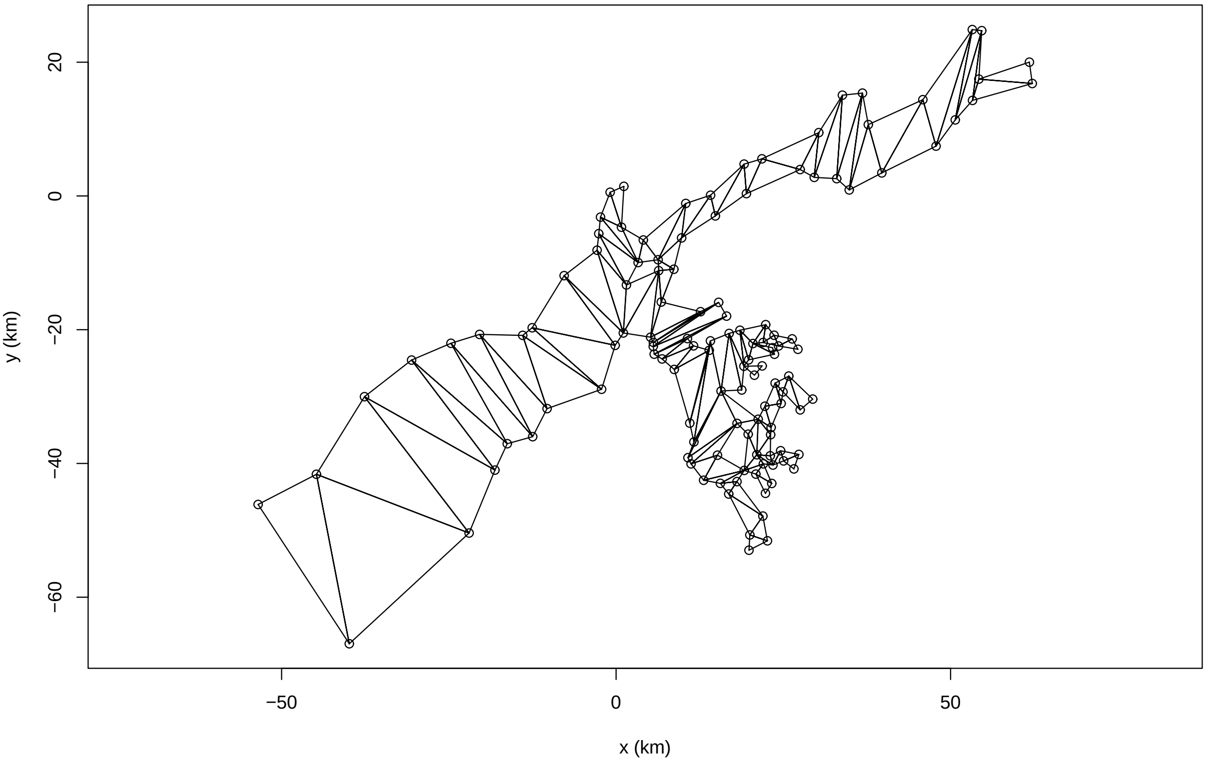
3. Apply Multidimensional scaling to the grid nodes using these distances. $\Rightarrow$ Approx. water distance.

4. Map the data locations from the original space to the new Euclidean one by linear interpolation.

Fast and easy to do, especially when you have the grid in advance.

For new data locations in the grid, you only need to do the mapping to the Euclidean space (4.).

Grid of convex polygons



Simplified triangular grid of the fjord system (in $\mathbb{R}^2$).

Problem II: Computational efforts

- With n data locations, $n(n-1)/2$ distances are needed.

- With our method we compute the distances in the grid and do the Multidimensional scaling once

  $\Rightarrow$ most of the computational burden is dependent on the size of the grid only.

- Typical herring survey in Vestfjord: $n \approx 20,000$ (each year).

- The exact sampling locations vary between years.

What is Multidimensional scaling?

Distance matrix $\boldsymbol{D}$ ($\dim(\boldsymbol{D}) = m \times m$); with all distances between the m grid locations.

$\boldsymbol{D}$ is symmetric with $d_{rr} = 0$ and $d_{rs} \geq 0$ for $r \neq s$.

Multidimensional scaling may be used for constructing a map of the data locations in a k-dimensional Euclidean space from $\boldsymbol{D}$, where $k \leq p \leq m$.

p is $\#$ of pos. eigenvalues of a transformation of $\boldsymbol{D}$.

How do we choose k?

- Ideally, we would like to use $k = p$
  $\Rightarrow$ the best Euclidean approximation of the space.
- In practise, $k = 2$ or $k = 3$ may be the easiest to work with and visualise.
- You can use measures based on the relative values of the eigenvalues.

Alternative approaches

- Sampson & Guttorp (1992) used non-metric multi-dimensional scaling of the spatial covariance matrix. Based on repeated observations.

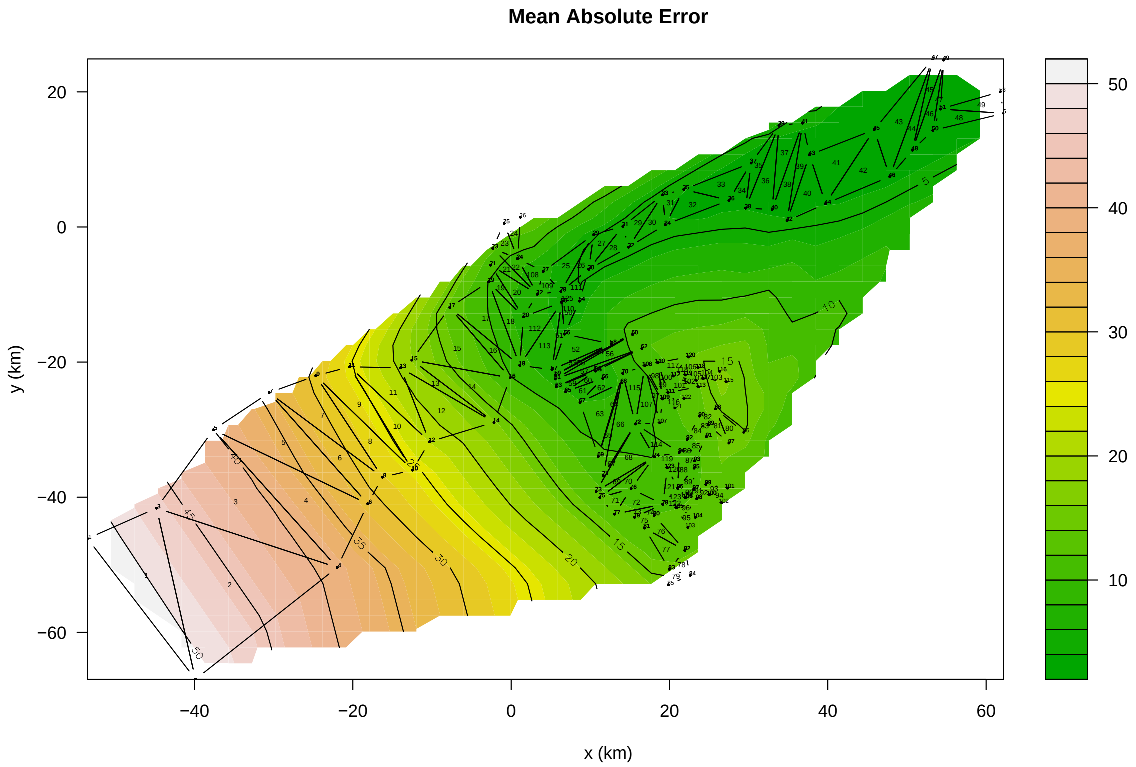
- Some non-stationary spatial models fitted to observations from a single realisation of a spatial process include
  - moving-window method of Haas (1995)
  - kernel smoothing approach of Higdon (1998) and Fuentes (2001)

Results

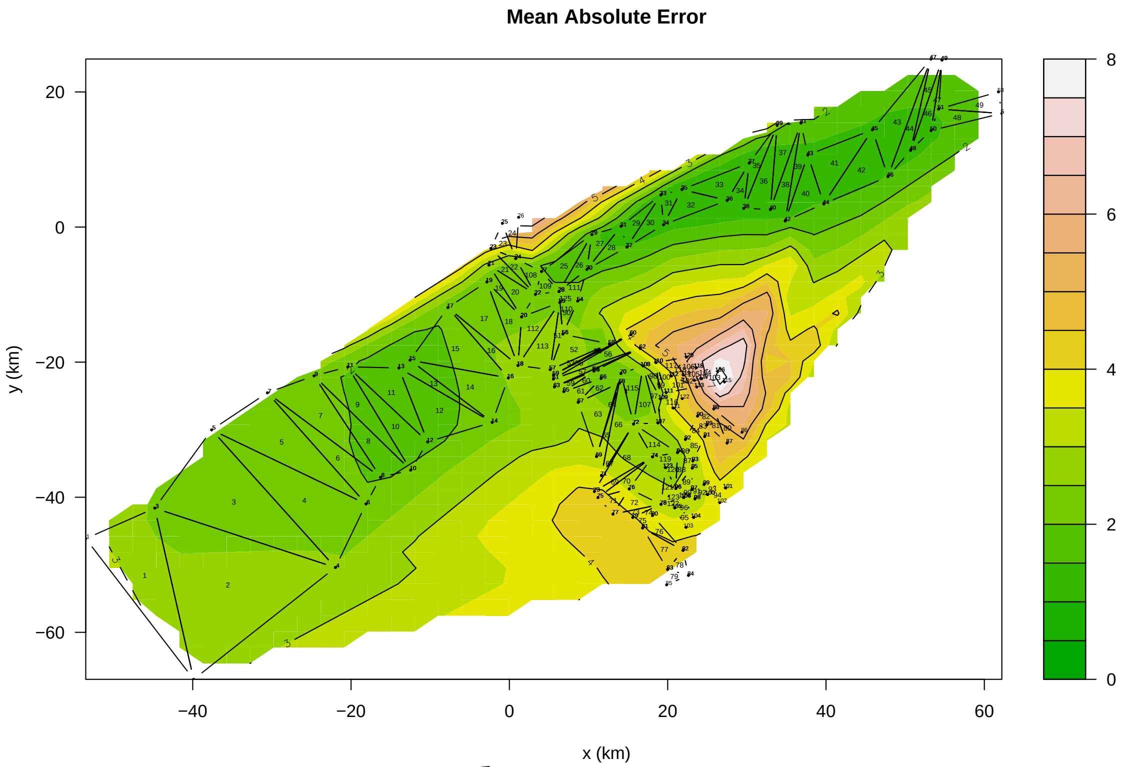
Define the sample mean absolute distance error $e(\boldsymbol{s}_i)$ as

$$e(\boldsymbol{s}_i) = \frac{\sum_{j=1, j\neq i}^{m} |d_{ij} - \hat{d}_{ij}|}{m-1}.$$

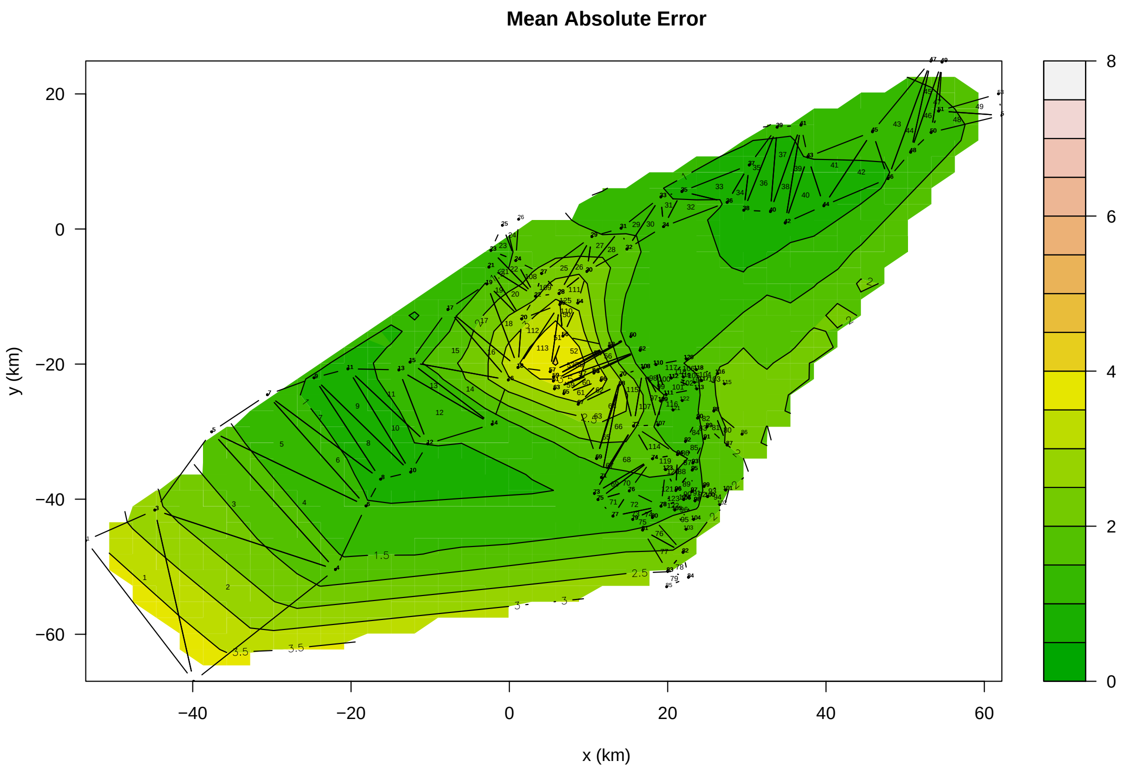
$$k = 1:$$



$k = 2$:

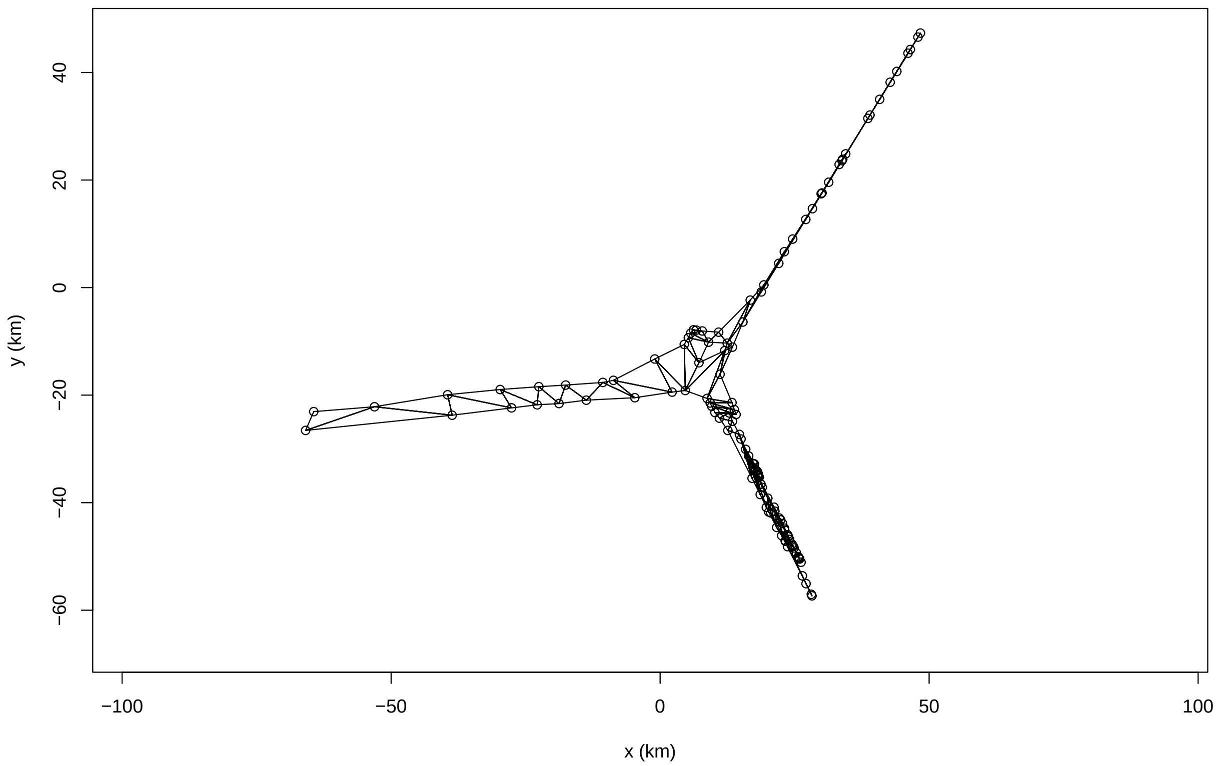

Mean Absolute Error
y (km)
x (km)
20
0
−20
−40
−60
−40
−20
0
20
40
60
8
6
4
2
0

$k = 10$:

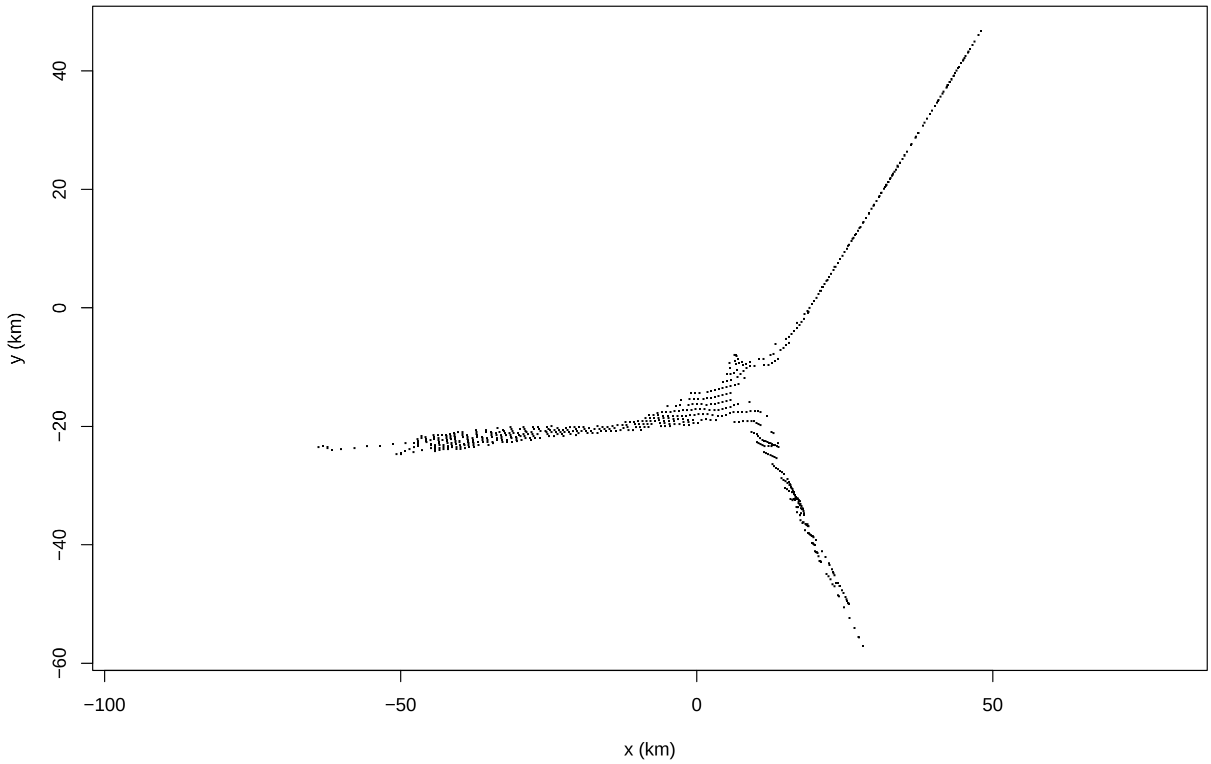

Mean Absolute Error
y (km)
x (km)
20
0
−20
−40
−60
−40
−20
0
20
40
60
8
6
4
2
0

Vestfjord

We found that $k = 2$ gave a quite good representation:



Transformed (triangular grid version of the) fjord.



Data locations in the transformed triangular grid.

Empirical variograms

Calculation of the exact water distances is demanding

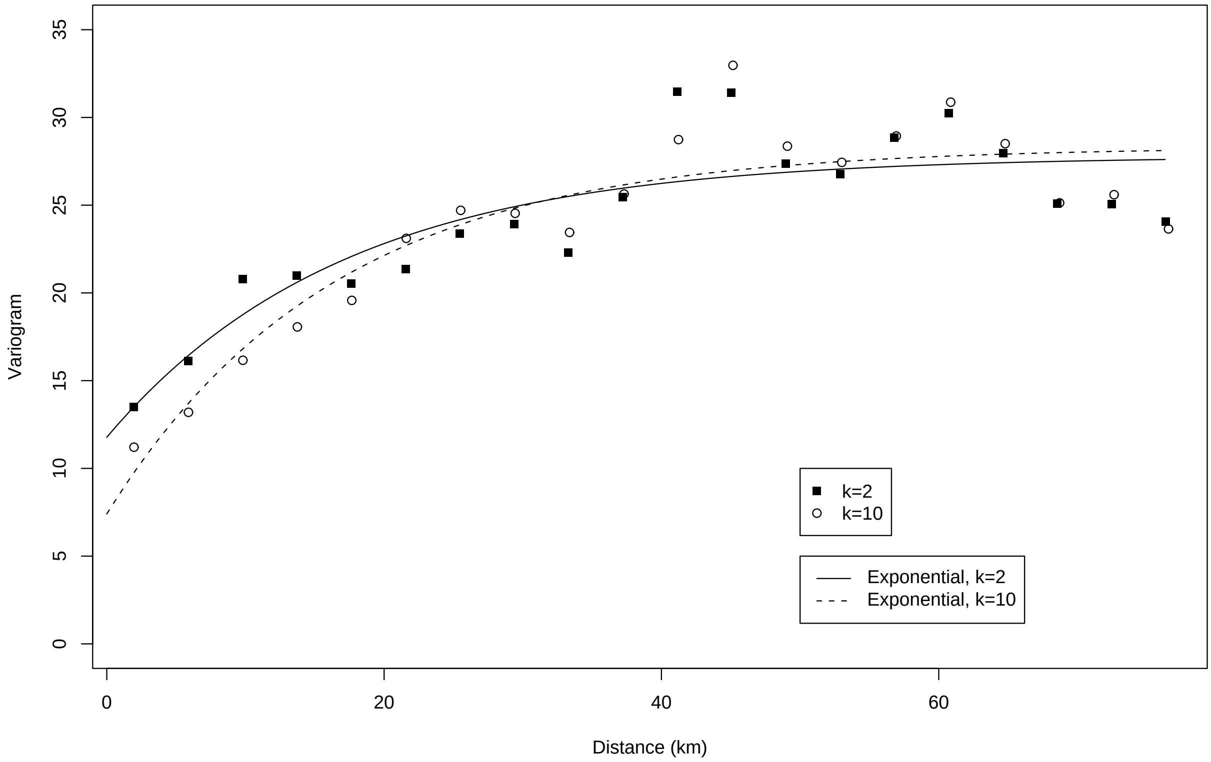
$\Rightarrow$ used only a subset of the data set in the example.

Denote by

- $2\hat{\gamma}_2$ the estimated variogram using our proposed approximate water distance with dimension $k = 2$
- $2\hat{\gamma}_{10}$ the estimated variogram using our proposed approximate water distance with dimension $k = 10$

We fitted variograms of the exponential class to the data with Ordinary Least Squares (OLS).

Empirical variograms



$2\hat{\gamma}_2$ and $2\hat{\gamma}_{10}$. The two lines represent variograms of the exponential class fitted to the data with OLS.

- Quite similar variograms.
- Largest difference between these two at the small lags.

  $\Rightarrow$ May be important in spatial prediction.

Conclusion

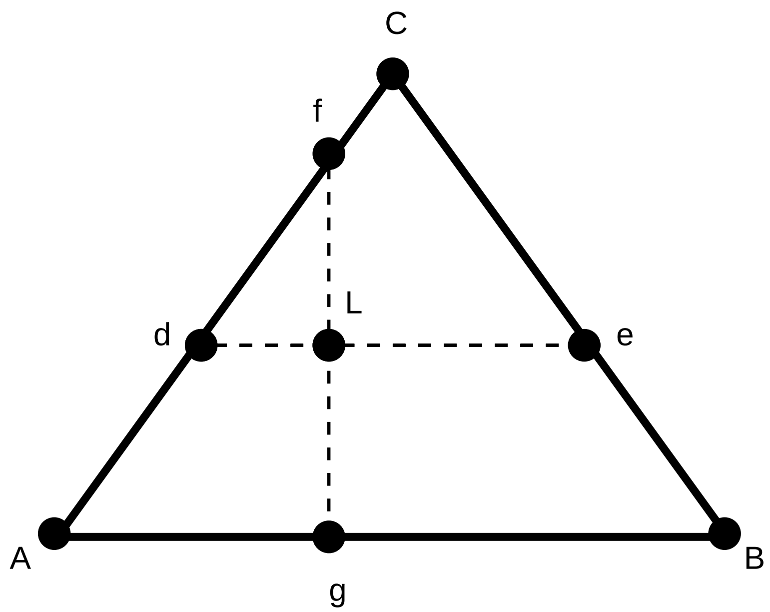
- Multidimensional scaling may be used as a tool for approximating the water distance metric.

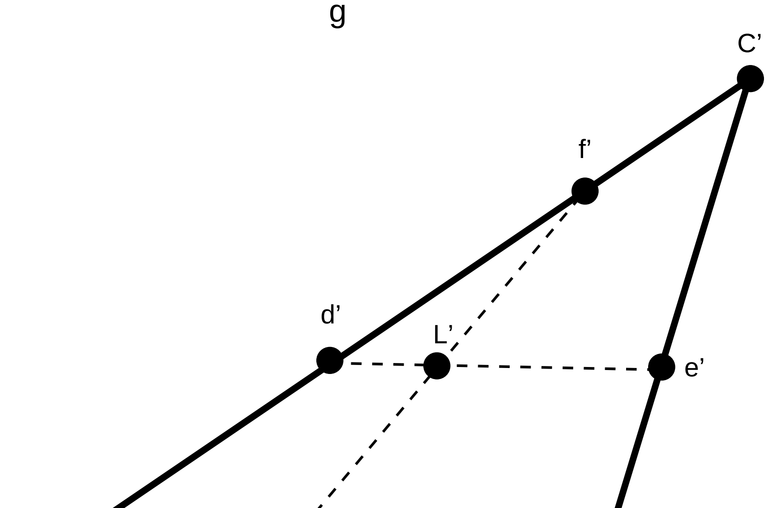
- Ensures valid spatial covariance models and may significantly increase computational efficiency.

- The water distance metric may be particularly important for problems with long-range dependence.

- The application to herring data from Vestfjord suggests that the geometry can be represented by a two-dimensional deformed space.

New locations





Covariance in one particular point, $k = 10$