Scale-Space Methods and GMRFs with applications

Fred Godtliebsen
Department of mathematics and Statistics
University of Tromsø

Outline of talk

- Scale-space methods
  - ▷ Motivation
  - ▷ Scale-space examples
- Gaussian Markov Random Fields
- Research in eVITA project

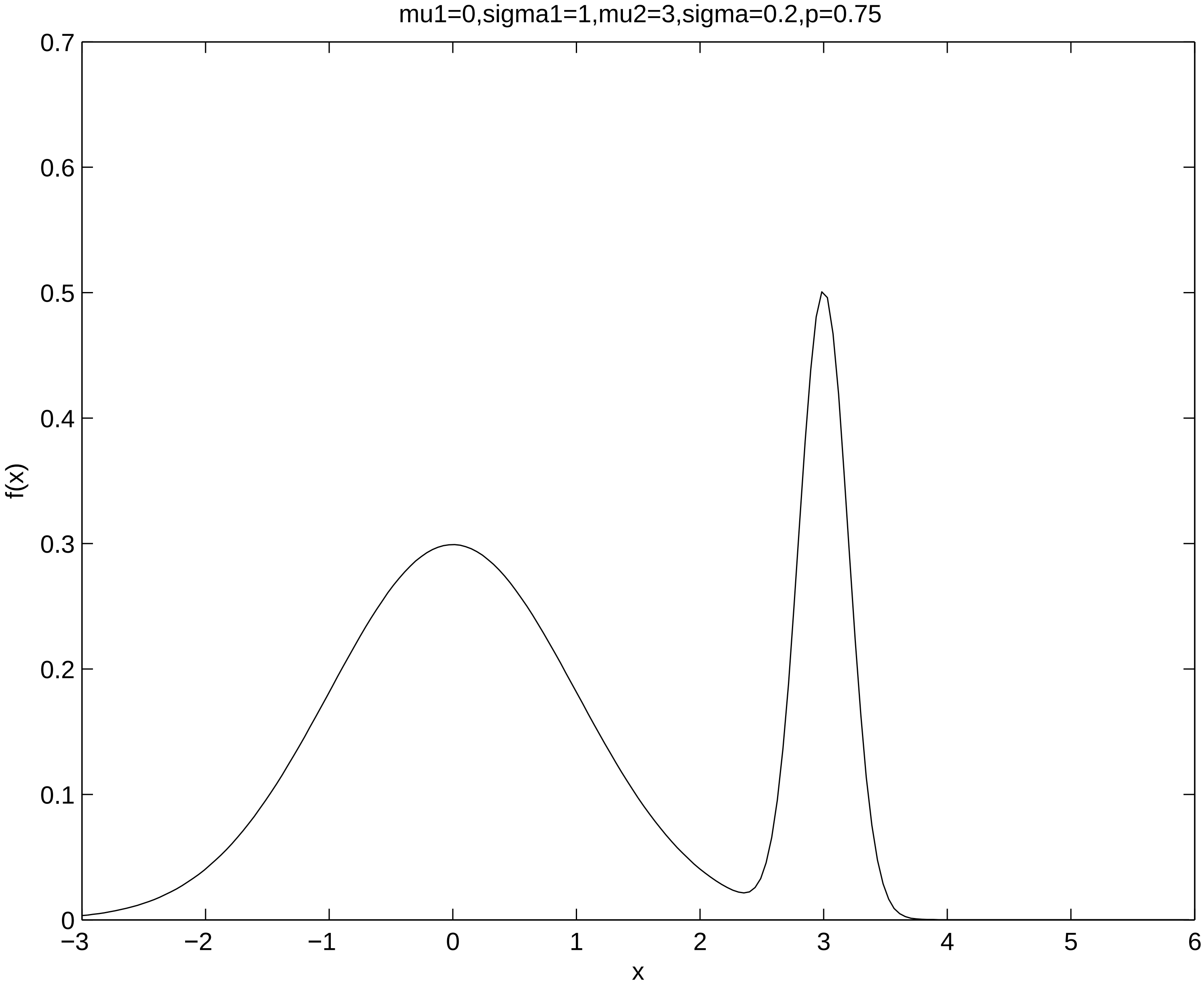
1. Scale-Space methods

MOTIVATION : Suppose we sample $X_1, X_2, \ldots, X_n$ from

$$f(x) = p\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + (1-p)\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

Plot of f when we have:

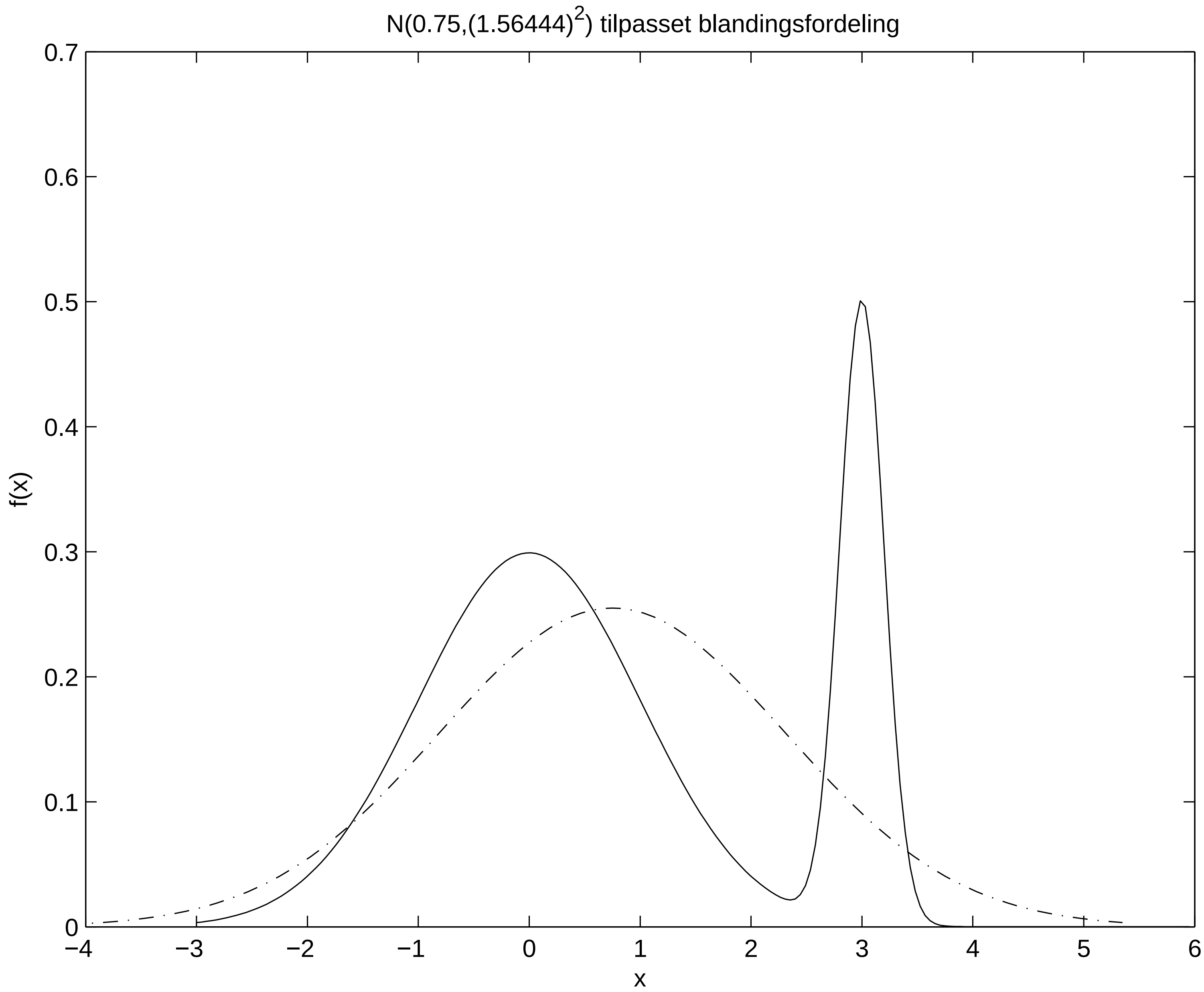
$$\begin{aligned} p &= 0.75 \\ \mu_1 &= 0 \\ \sigma_1 &= 1 \\ \mu_2 &= 3 \\ \sigma_2 &= 0.2 \end{aligned}$$


mu1=0,sigma1=1,mu2=3,sigma=0.2,p=0.75
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
f(x)
−3
−2
−1
0
1
2
3
4
5
6
x

- **Interesting question :** What happens if we assume our model is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty?$$

- **Bad fit :** See figure on next page.



N(0.75,(1.56444)^2) tilpasset blandingsfordeling
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
f(x)
−4
−3
−2
−1
0
1
2
3
4
5
6
x

- Better estimate obtained by kernel density estimator

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

where e.g.

$$K(\frac{x - X_i}{h}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}}$$

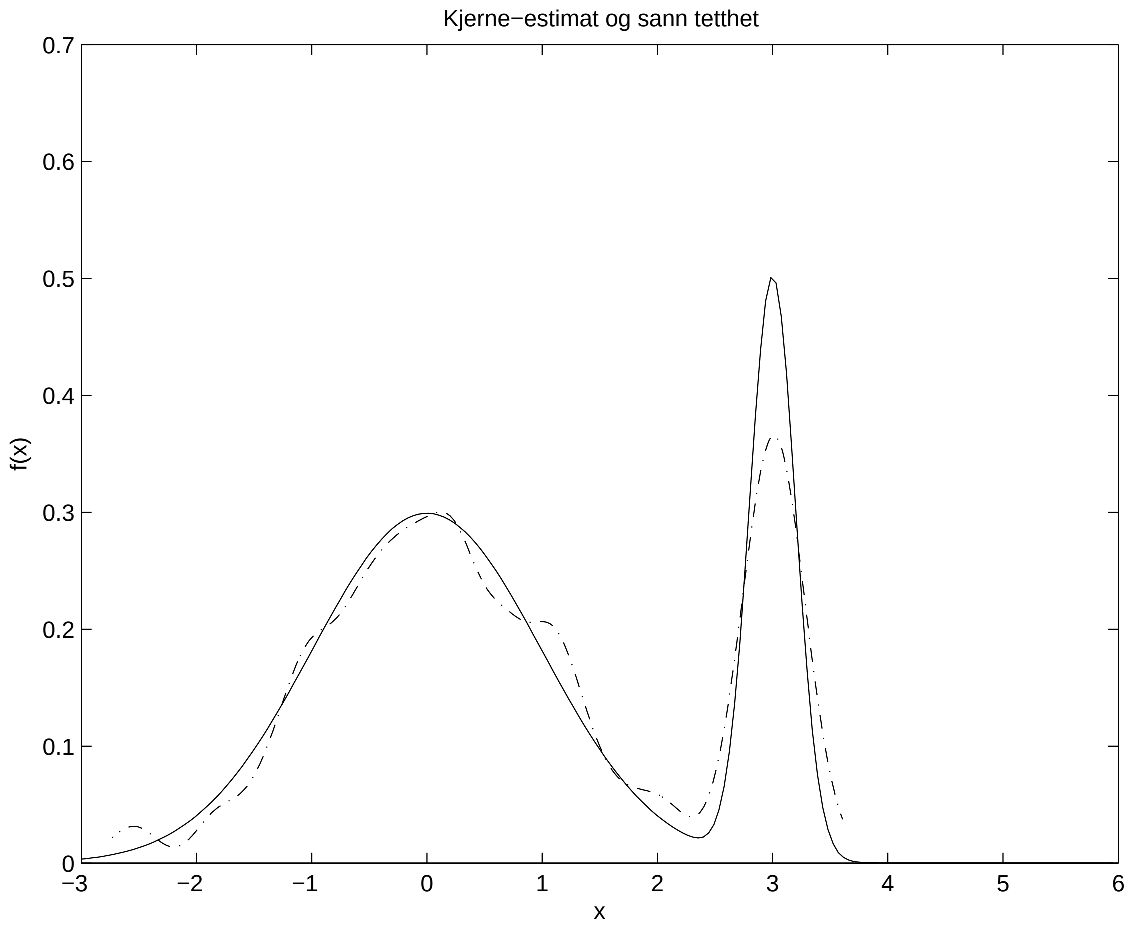
and h is a *smoothing parameter.*

- Easy to show that

$$\begin{aligned} \mathrm{E}[\hat{f}_h(x)] &= f(x) + h^2 \frac{f''(x)}{2} \int_{-\infty}^{\infty} K(z) z^2 dz + o(h^2) \\ \mathrm{Var}[\hat{f}_h(x)] &= \frac{1}{nh} \int_{-\infty}^{\infty} K(z)^2 dz + o(\frac{1}{nh}). \end{aligned}$$

- **Means :** $\hat{f}_h(x)$ is not an unbiased estimator of $f(x)$.

Kernel estimate of normal mixture :

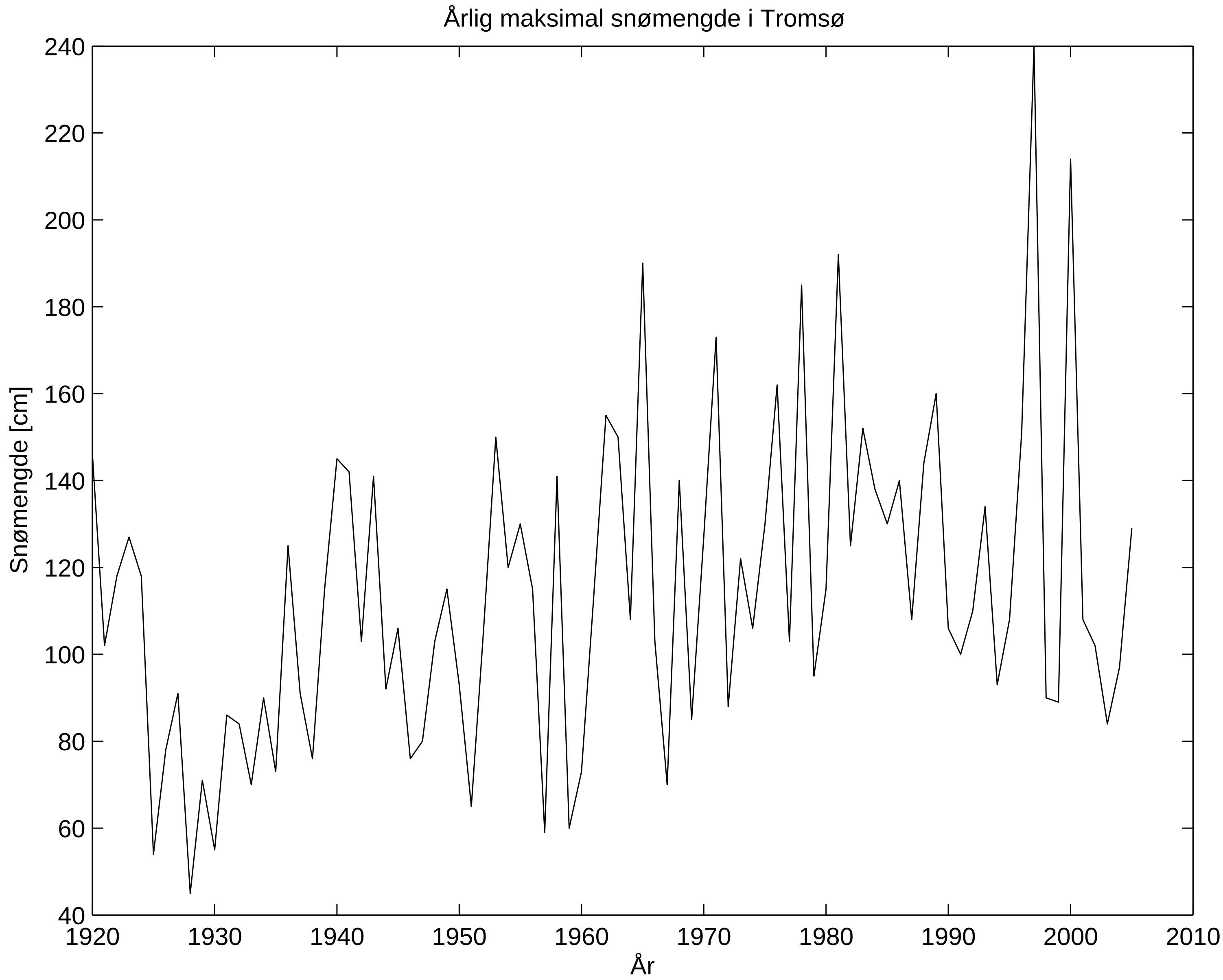


- **Interpretation :**
  - **Large** h **:** Gives smooth $\hat{f}_h$, but bad estimate of $f(x)$ where $f''(x)$ is big (in absolute value).
  - **Small** h **:** Gives rough curve.
  - **Important compromise :** Between bias and variance.
- Much research on similar situations has been performed.

What about other types of curves?

- We have considered estimation of densities.
- Suppose now we have observed something that can be modeled as a function of another variable. (i.e. a regression situation)
- **Example :** Look at annual maximum snow amount in Tromsø as function of year, see figure next page.

Annual maximum snow amount in Tromsø :



How could we estimate curve here ?

- **Parametric model :** In our example it would be natural to use the model:

$$y_i = a + bt_i + e_i, \quad i = 1, ..., n.$$

where

$$\begin{aligned} y_i &= \text{observed maximum snow amount in year } i \\ t_i &= \text{year number } i \\ e_i &= \text{random noise number } i \end{aligned}$$

where a and b are parameters that need to be estimated.

- **Nonparametric model :**

$$y_i = m(t_i) + e_i, \quad i = 1, ..., n.$$

where

$$\begin{aligned} y_i &= \text{observed maximum snow amount in year } i \\ m(t_i) &= \text{expected snow amount in year nr } i \\ e_i &= \text{random noise nr } i \end{aligned}$$

Here, we seek an estimate of m.

- Parametric and nonparametric models for such curves can be handled in the same way as we did for densities.

Problem with Confidence Intervals :

- **Recall interpretation of CI:** Suppose we have an iid sample $Y_1, ..., Y_n$ from a $\mathrm{N}(\mu, \sigma^2)$ distribution and we want to construct a 95 % CI for μ. The CI is then

$$[\bar{Y} - t_{0.025,n-1}\frac{S}{\sqrt{n}}, \bar{Y} + t_{0.025,n-1}\frac{S}{\sqrt{n}}].$$

If we repeat this procedure 100 times, i.e. we have 100 such samples available, then we expect about 95 of the constructed CIs to cover the true parameter μ.

- Since $\hat{f}_h(x)$ is a biased estimator for $f(x)$,

$$\frac{\hat{f}_h(x) - f(x)}{\widehat{\mathrm{SD}[\hat{f}_h(x)]}}$$

  cannot be used to get a 'correct' CI for $f(x)$. (Consider e.g. a situation where $f''(x)$ is large.)

- Similar 'story' for $f'(x)$

- An excellent discussion of this phenomenon is given in Section 6.2 of the Chaudhuri and Marron (1999) paper.

Scale-Space idea :

- Leave the search for underlying true curve f (or f').
- I.e. do not seek CI for $f'(x)$.
- **Instead :** Seek CI for scale-space version (i.e. smooth version of $f'(x)$) :

$$f_h'(x) = \mathrm{E}[\hat{f}_h'(x)] = (K_h^{'} \star f)(x)$$

- **Two advantages :**
  - $\hat{f}'_h(x)$ is an unbiased estimator of $f'_h(x)$ so CI for $f'_h(x)$ is easily found from

$$\frac{\hat{f}'_h(x) - f'_h(x)}{\widehat{\mathrm{SD}[\hat{f}'_h(x)]}} \approx \mathrm{N}(0,1).$$

  - All h values are used, so no need to search for an optimal h.

Scale-Space Examples

- For each (x, h) location, test

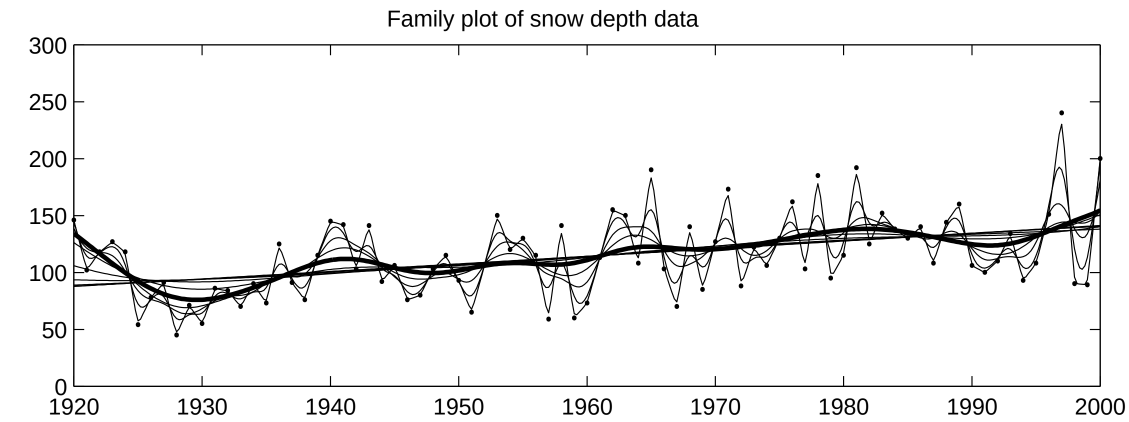
$$\mathrm{H}_0 : f_h'(x) = 0 \quad \text{against} \quad \mathrm{H}_1 : f_h'(x) \neq 0.$$

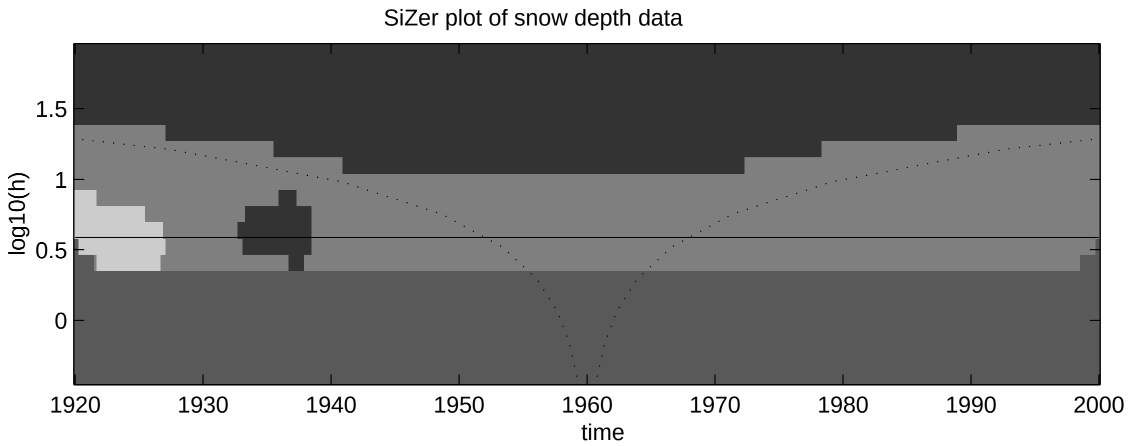
- Test is based on CI for $f_h'(x)$, i.e. :

$$[\hat{f}_h'(x) - q \cdot \widehat{\mathrm{SD}[\hat{f}_h'(x)]}, \hat{f}_h'(x) + q \cdot \widehat{\mathrm{SD}[\hat{f}_h'(x)]}]$$

- (x, h)-location is
  - ▷ Significant increasing (black) when CI above 0.
  - ▷ Significant decreasing (white) when CI under 0.
  - ▷ Not significant (gray) when CI covers 0.

Example 1 : SiZer for snow data

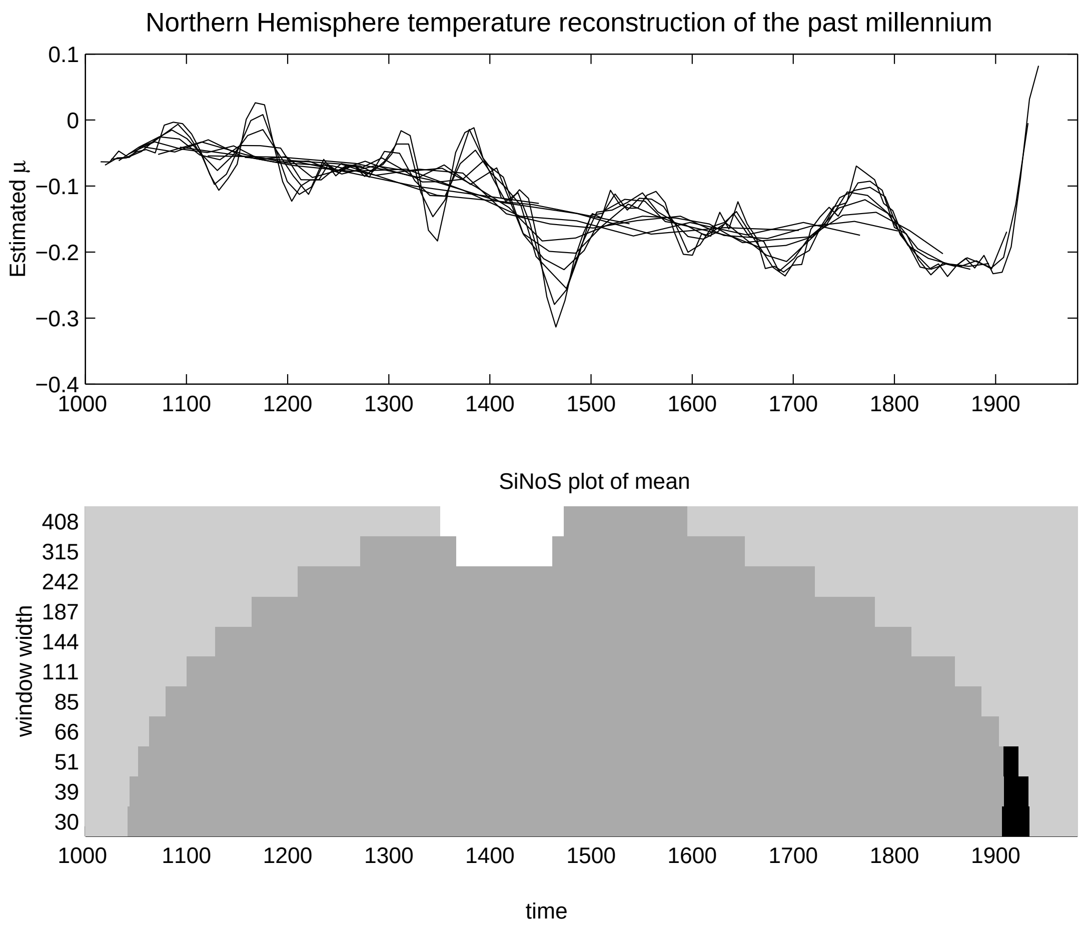




SiNos

- This technique detects areas where there are changes in the statistical properties of the signal.
- Assumes underlying model is a stationary Gaussian process.
- Testing for departures from this model.

Example 2 : Temperature



Example 3 : Liver transplantation

- Collaboration with surgeons at UNN.
- Research concerning liver transplantation.
- Surgery performed on pigs.
- Several parameters (like blood pressure) are measured every 4th second for 9 hours.
- SiNos can be used to decide whether observed structures are significant.

Example 4 : Albedo application

Albedo : Proportion of light or radiation reflected by surface.

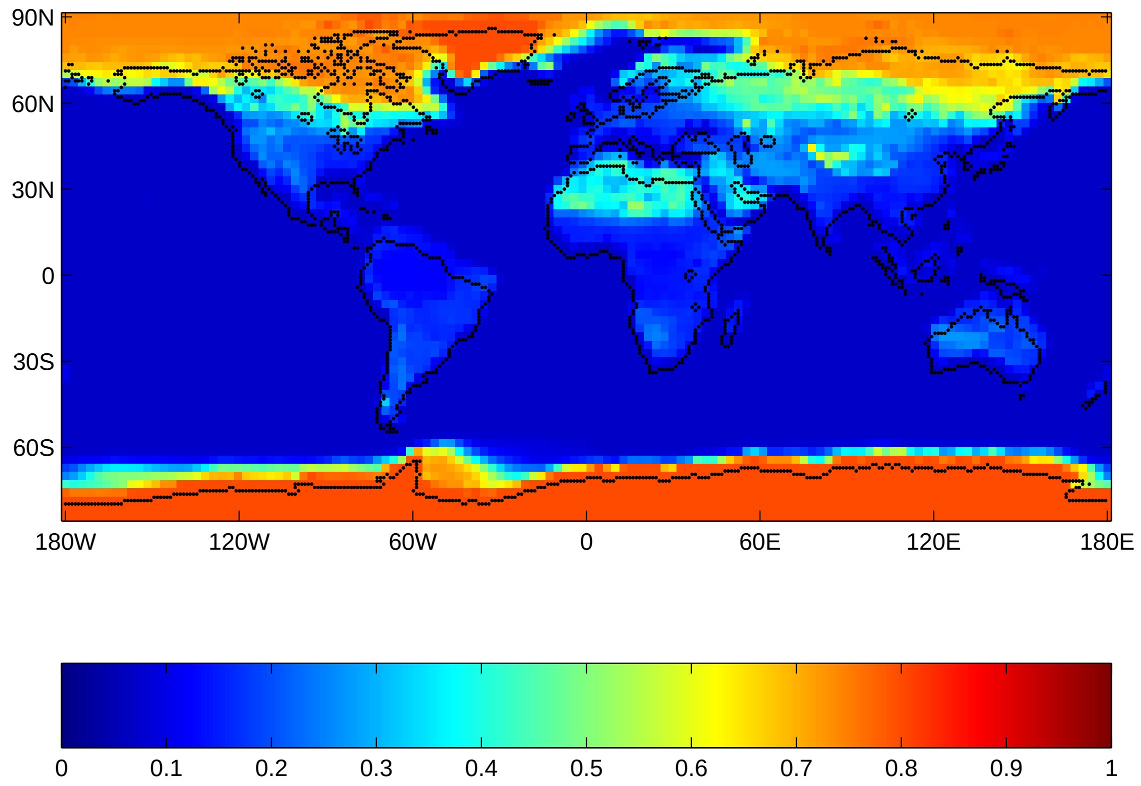


Figure 1: ECHAM5 model simulation of March mean

Example 4 : Albedo application

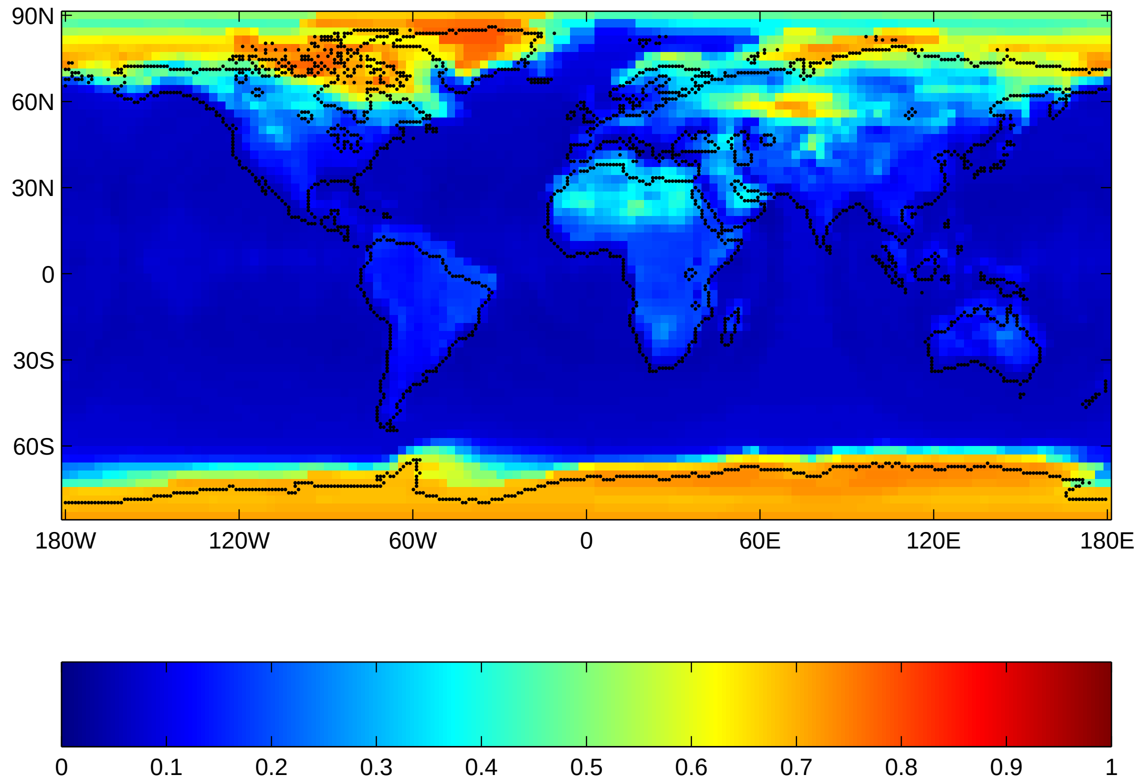


Figure 2: PINKER observations of March mean

Example 4 : Albedo application

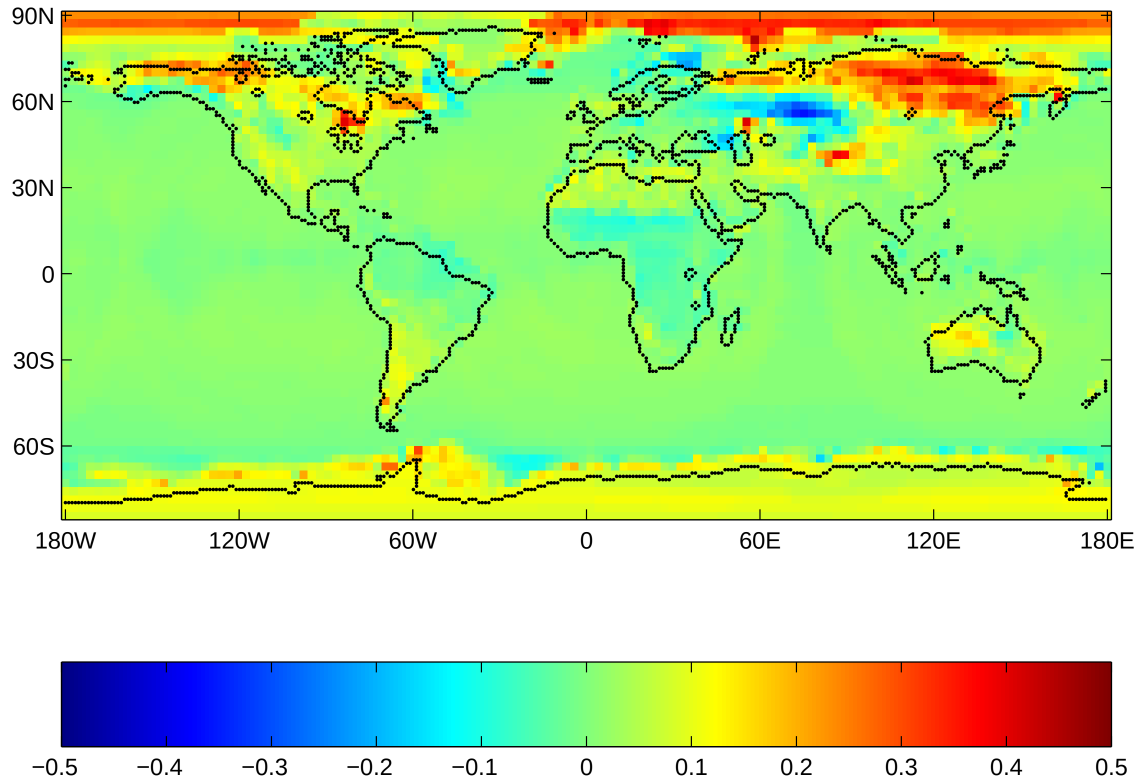


Figure 3: ECHAM5 - PINKER of March mean

Example 4 : Albedo application

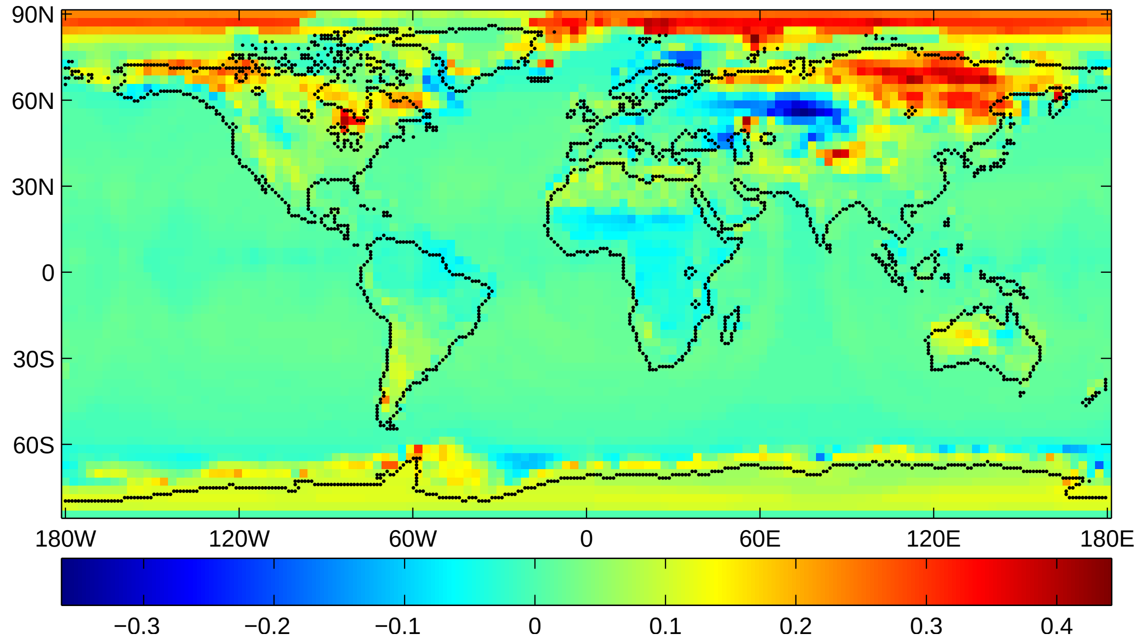


Figure 4: Smoothing with scale of 280km

Example 4 : Albedo application

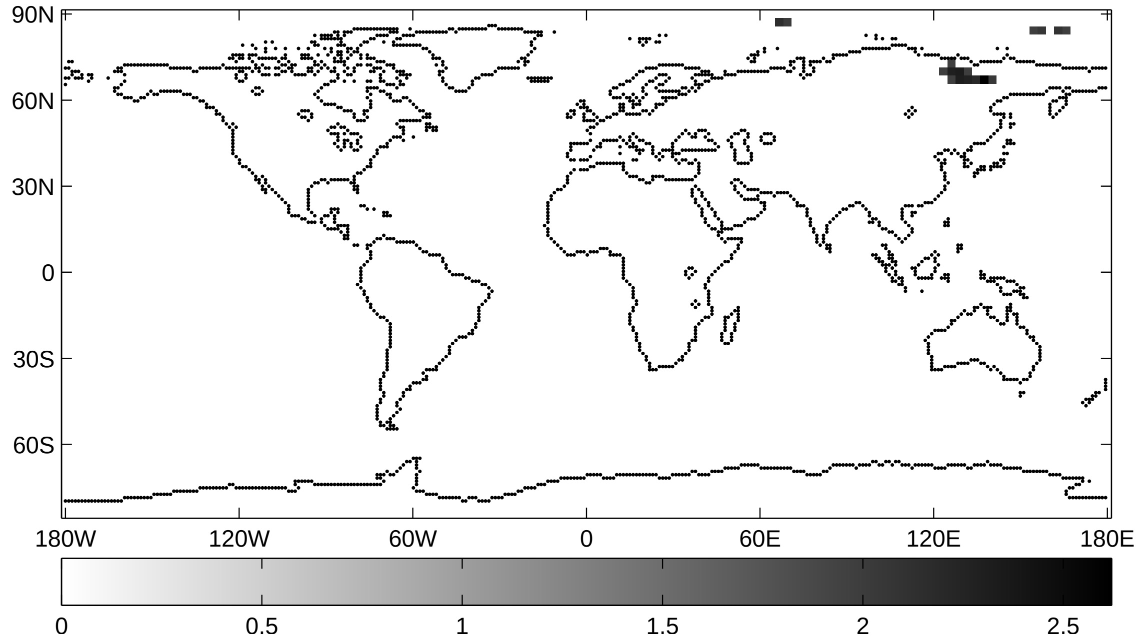


Figure 5: Significance plot for scale of 280km

Example 4 : Albedo application

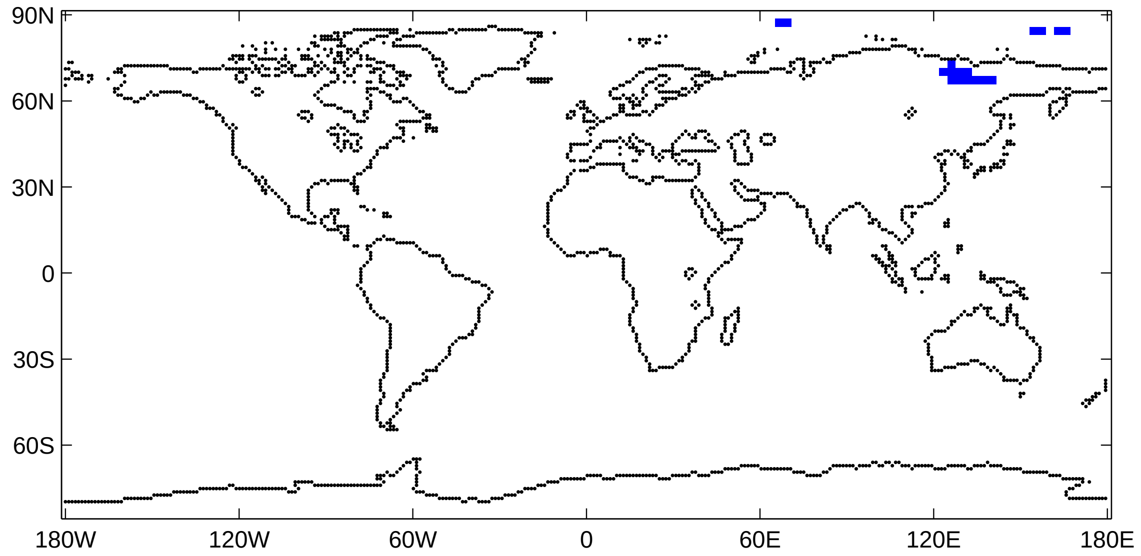


Figure 6: Feature map for scale of 280km

Example 4 : Albedo application

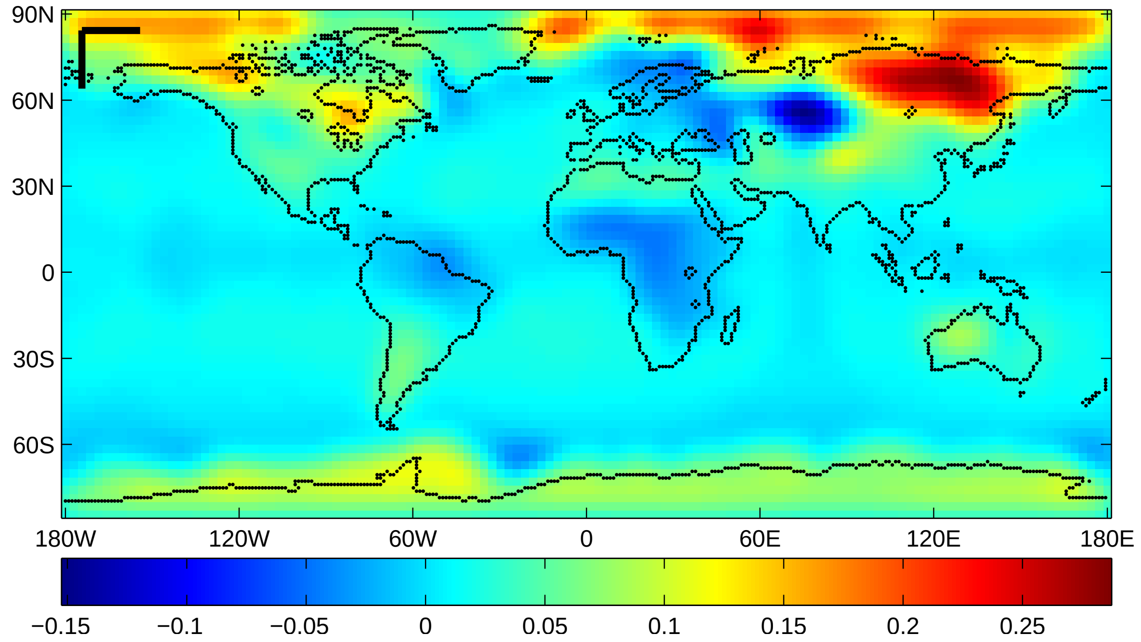


Figure 7: Smoothing with scale of 2000km

Example 4 : Albedo application

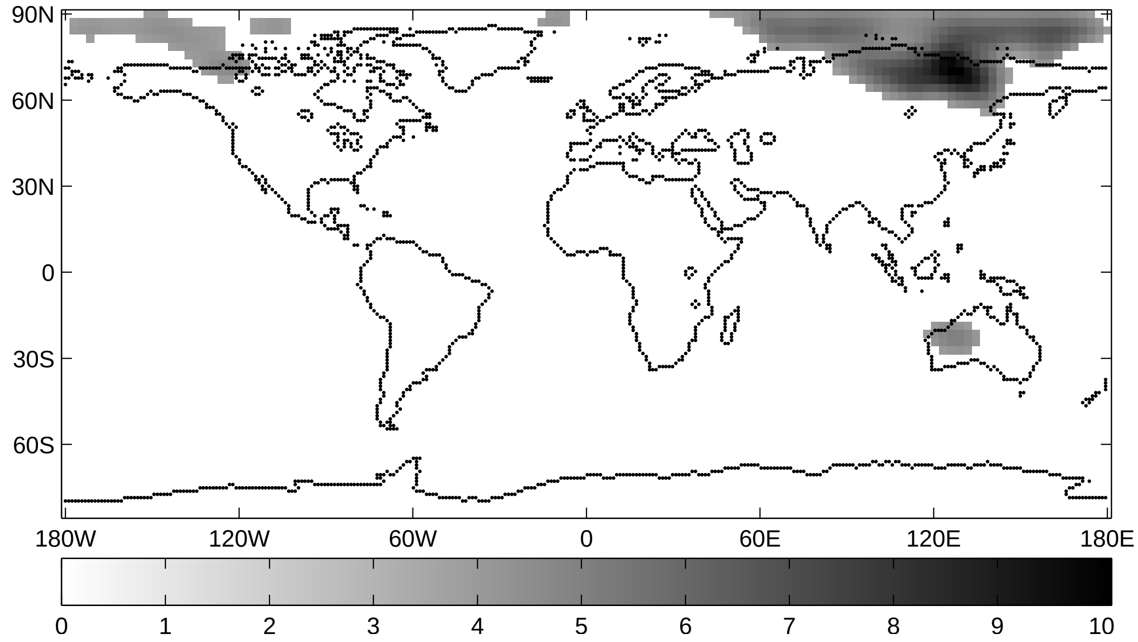


Figure 8: Significance plot for scale of 2000km

Example 4 : Albedo application

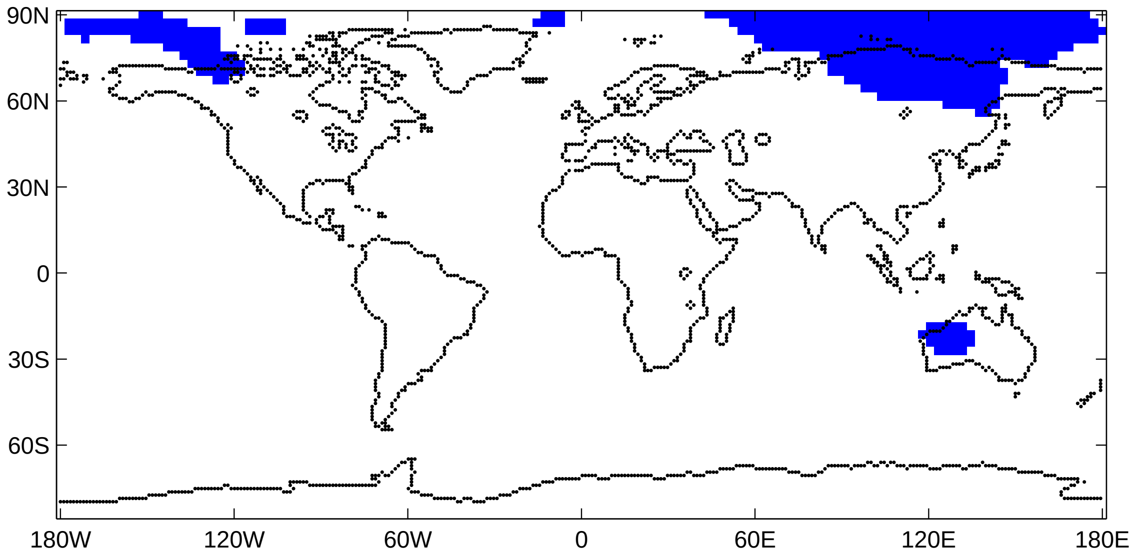


Figure 9: Feature map for scale of 2000km

Example 4 : Albedo application

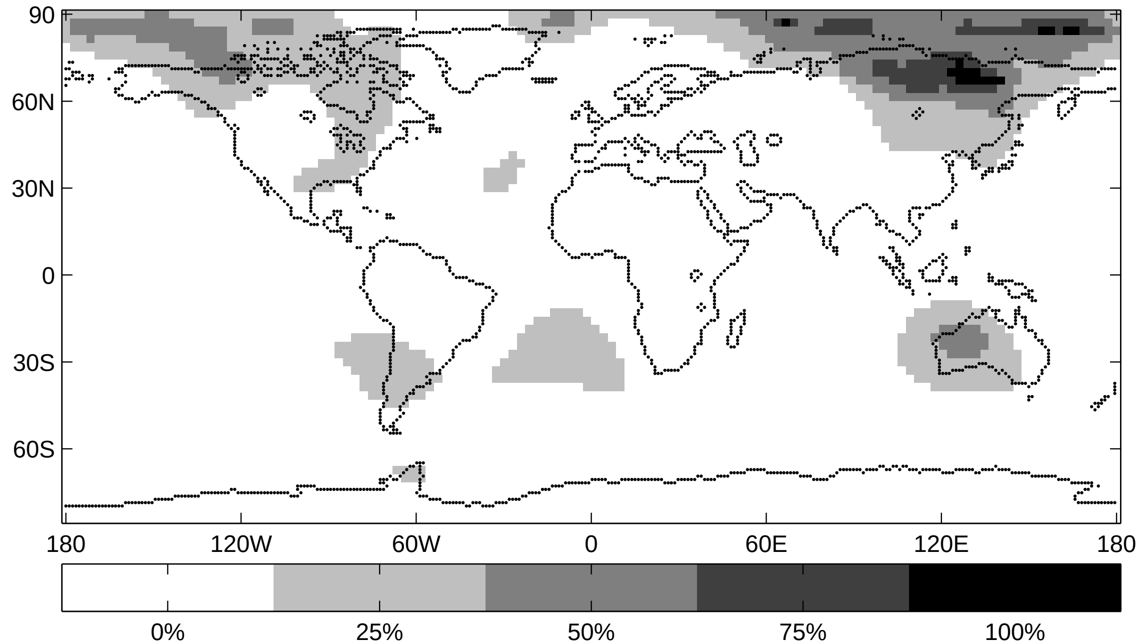
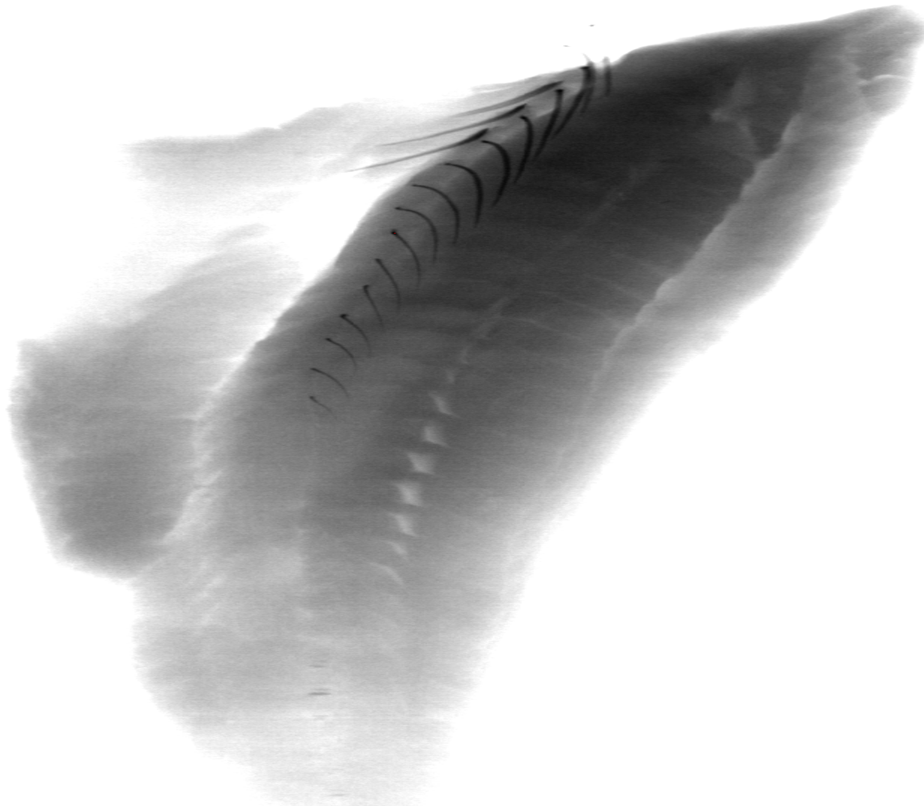


Figure 10: Significant features for all scales

Example 5 : Detection of fish bones

- Fillets often contain remaining bones after fish has gone through fillet machine.
- Would like to have exact position of bones so that bones can be automatically removed.
- This equipment is now in use in fish factories.

Example 5 : Detection of fish bones



Example 5 : Detection of fish bones

Gaussian Markov Random Fields

- $\mathbf{x} = (x_1, \ldots, x_n)^T \sim \mathrm{MN}(\mu, \mathbf{\Sigma})$.
- Labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, ..., n\}$ and $\mathcal{E}$ are the set of edges $\{i, j\}$, where $i, j \in \mathcal{V}$ and $i \neq j$.
- Furthermore, denote $\mathbf{Q}(= \Sigma^{-1})$ by the precision matrix.
- The random vector $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbf{R}^n$ is called a GMRF with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean μ and precision matrix $\mathbf{Q} > 0$ if and only if its density has the form

$$\pi(\mathbf{x}) \quad = \quad (2\pi)^{-n/2}|\mathbf{Q}|^{1/2}\exp[-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{Q}(\mathbf{x}-\mu)]$$

and

$$Q_{ij} \neq 0 \Longleftrightarrow \{i,j\} \in \mathcal{E} \text{ for all } i \neq j$$

- Realistic modelling will often produce a sparse **Q** matrix and then we really benefit from the speedup achieved.
- Rue and Held (2005) is an excellent reference concerning GMRF.

Research in eVITA project

3 MAIN APPLICATION AREAS:

- fMRI (1 researcher, 2.5 years)
  - Migraine, face recognition, lingvistics.
- Telemedicine (1 PhD student)
  - Monitoring of patients, Computer-aided diagnostics
  - Connected to SFI in telemedicine
- Climatology (1 postdoc in 2.5 years)
  - Unevenly sampling, multivariate time series
  - Spatio-temporal problems

ADDITIONAL GOAL:

- Develop GMRF expertise at Department of Mathematics and Statistics, UiT.
- Håvard Rue has prof II position at UiT in the project period, i.e. 5 years.

Example of advantage by using GMRF

Fishery example :

- Method must take less than 1 second.
- Original scale-space method using MCMC took several hours.
- Same problem solved by GMRF takes just a few seconds.
- This speedup is typical by utilizing GMRF.

Collaborators in eVITA project

- Prof. Probal Chaudhuri (ISI, Calcutta)
- Prof. Lasse Holmström (Univ. Oulo, Finland)
- Dr. Jörg Polzehl (Wias, Berlin)
- Prof. Arvid Lundervold (Univ. Bergen)
- Prof. Håvard Rue (NTNU, Trondheim)
- Norwegian Polar Institute
- UNN
- NST