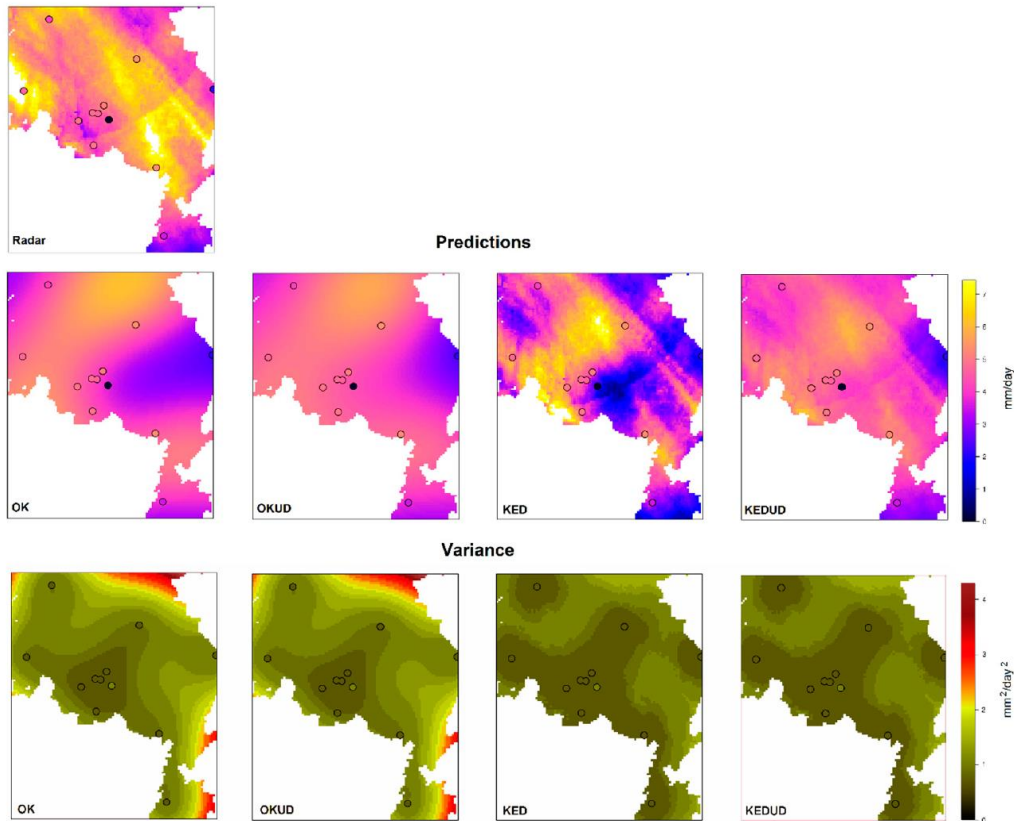
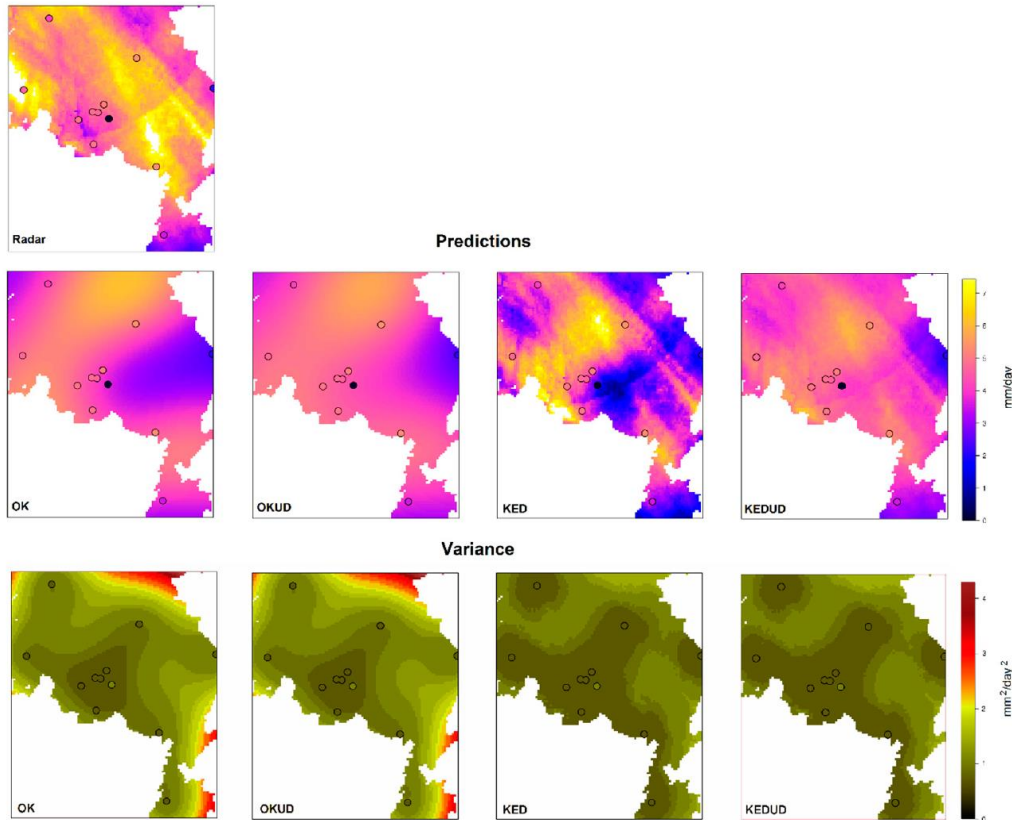


- Kriging (Simple, Ordinary, Universal)
- Conditional Simulation



Cecinati et al.: Considering Rain Gauge Uncertainty Using Kriging for Uncertain Data, Atmosphere (2018)

- Kriging (Simple, Ordinary, Universal)
- ~~Conditional Simulation~~



Cecinati et al.: Considering Rain Gauge Uncertainty Using Kriging for Uncertain Data, Atmosphere (2018)

Kriging

The measurement Z_i at location $x_i \in \mathbb{R}^n$ is interpreted as a **random variable** with **expected value**

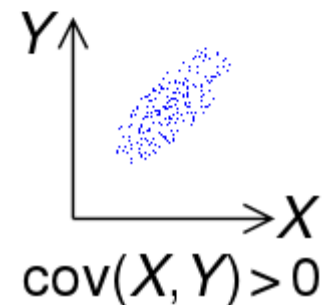
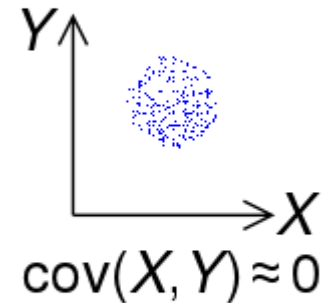
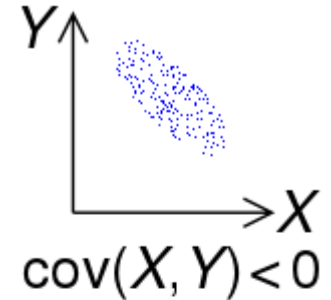
$$E[Z_i] = \int_{\Omega} z d\mu_{Z_i}(z)$$

and **variance**

$$Var(Z_i) = E[(Z_i - E[Z_i])^2].$$

The **covariance** of two random variables is given by

$$Cov(Z_i, Z_j) = E[(Z_i - E[Z_i])(Z_j - E[Z_j])].$$



Kriging

The measurement Z_i at location $x_i \in \mathbb{R}^n$ is interpreted as a **random variable** with **expected value**

$$E[Z_i] = \int_{\Omega} z d\mu_{Z_i}(z)$$

and **variance**

$$Var(Z_i) = E[(Z_i - E[Z_i])^2].$$

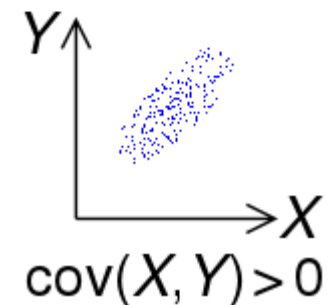
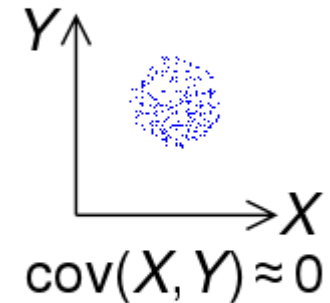
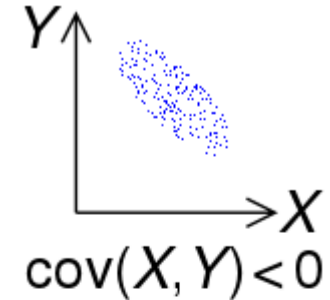
The **covariance** of two random variables is given by

$$Cov(Z_i, Z_j) = E[(Z_i - E[Z_i])(Z_j - E[Z_j])].$$

A crucial assumption in geostatistics is that of **stationarity***

$$E[Z(x)] = m, \quad Cov(Z(x+h), Z(x)) = C(h).$$

The covariance depends only on the increment $x - x_i$ between to locations, not the location itself!



* Stationarität oder auch Homogenität

Kriging

An approximation Z^* of Z is given by

$$Z^*(x) = \lambda_0 + \sum_{i=1}^N \lambda_i(x) Z_i,$$

with unknown weights $\lambda_i(x)$ such that

$$Z^*(x_i) = z_i, \quad i = 1, \dots, N.$$

An approximation Z^* of Z is given by

$$Z^*(x) = \lambda_0 + \sum_{i=1}^N \lambda_i(x) Z_i,$$

with unknown weights $\lambda_i(x)$ such that

$$Z^*(x_i) = z_i, \quad i = 1, \dots, N.$$

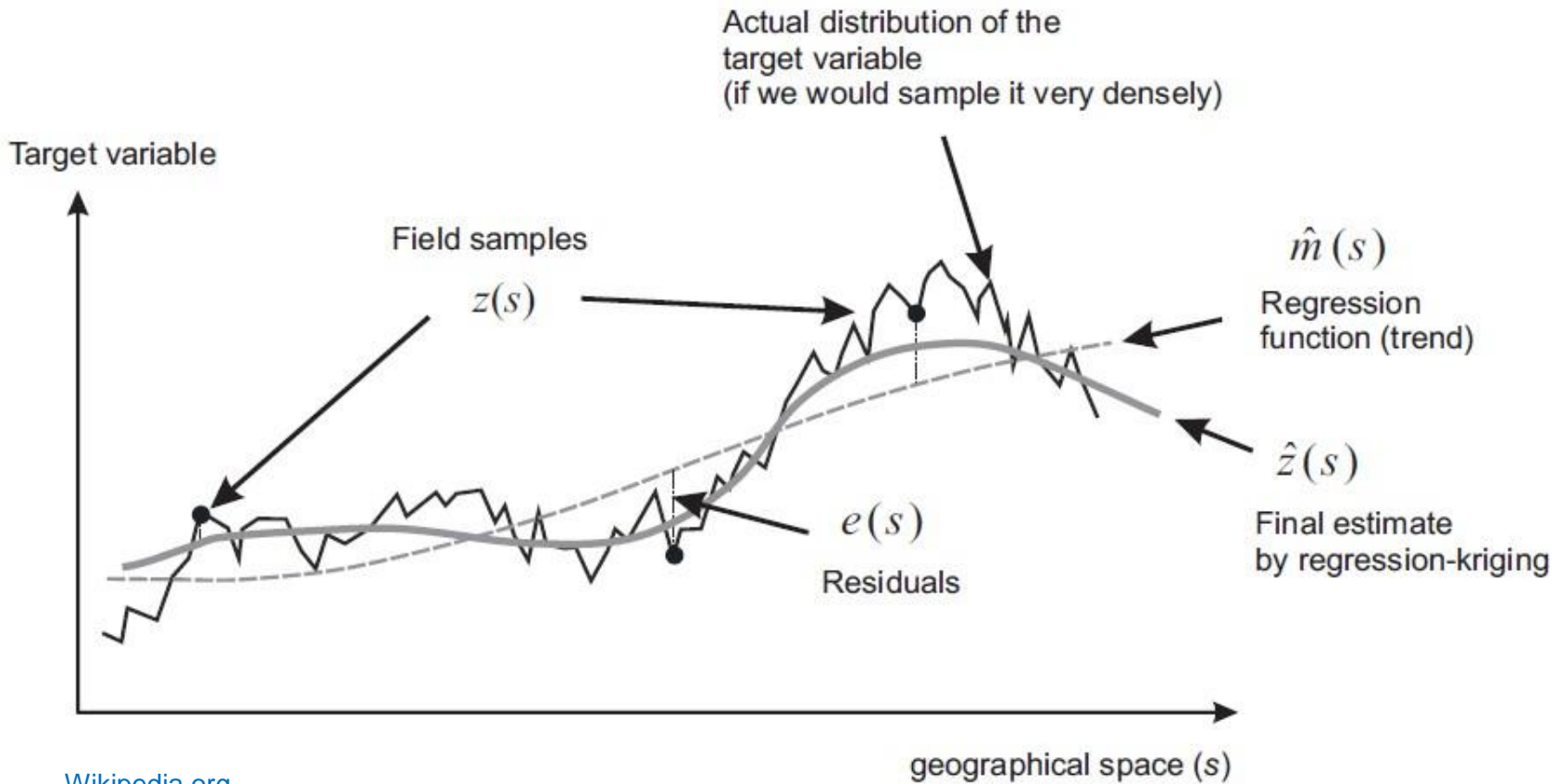
How to find the weights $\lambda_i(x)$? **Unbiasedness**

$$E[Z^*(x)] = E[Z(x)]$$

and **variance minimizing**

$$\min_{Z^*} \text{Var}(Z^*(x) - Z(x)).$$

Kriging



Wikipedia.org

Kriging

Simple Kriging: Expectation $m_0 = E[Z(x)]$ is known: Unbiasedness yields

$$\lambda_0 = m_0 - \sum_{i=1}^n \lambda_i m_0.$$

Kriging

Simple Kriging: Expectation $m_0 = E[Z(x)]$ is known: Unbiasedness yields

$$\lambda_0 = m_0 - \sum_{i=1}^n \lambda_i m_0.$$

Variance minimization leads to

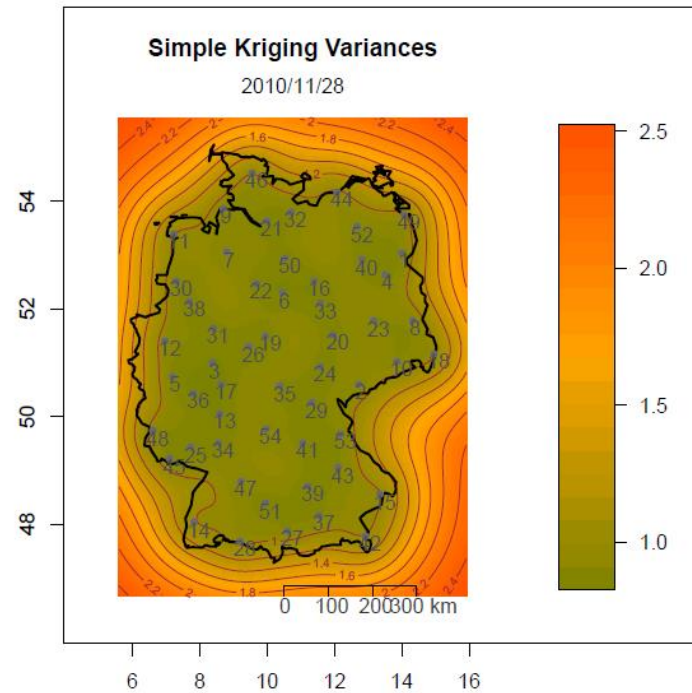
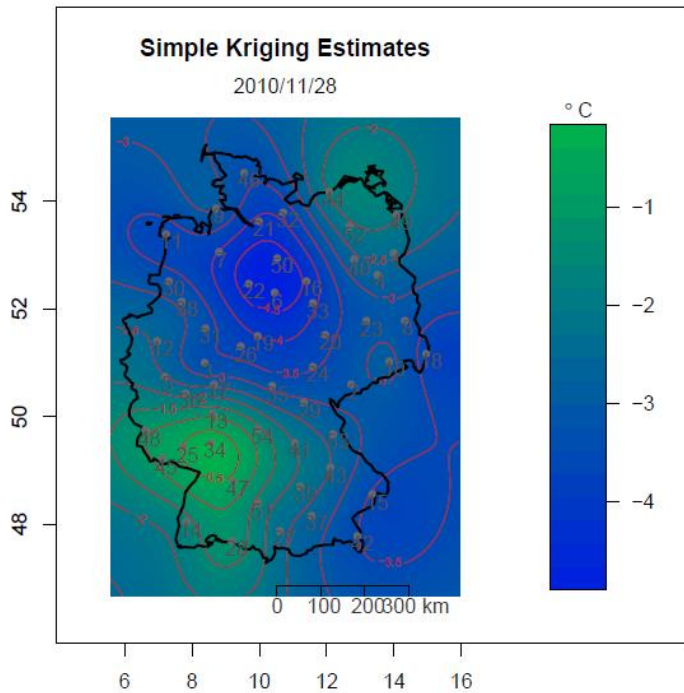
$$\sum_{i=1}^n \lambda_i(x) \text{Cov}(Z_i, Z_k) = \text{Cov}(Z_k, Z(x)), \quad k = 1, \dots, n.$$

Note that stationarity implies $\text{Cov}(Z_k, Z(x)) = C(x_k - x) \rightarrow$ Relation to radial basis functions.

Kriging

Simple Kriging: Expectation $m_0 = E[Z(x)]$ is known: Unbiasedness yields

n



Example from the Bachelor thesis "Kriging methods in spatial statistics", Andreas Lichtenstern, TU München

Kriging

In general, $Cov(Z(x_i), Z(x_j))$ is unknown and needs to be estimated or modelled.

A crucial role is given to the **variogram**

$$\gamma(h) = \frac{1}{2}Var(Z(x) - Z(x + h)) = C(0) - C(h).$$

It can be estimated from the data (under certain conditions).

Kriging

In general, $Cov(Z(x_i), Z(x_j))$ is unknown and needs to be estimated or modelled.

A crucial role is given to the **variogram**

$$\gamma(h) = \frac{1}{2} Var(Z(x) - Z(x+h)) = C(0) - C(h).$$

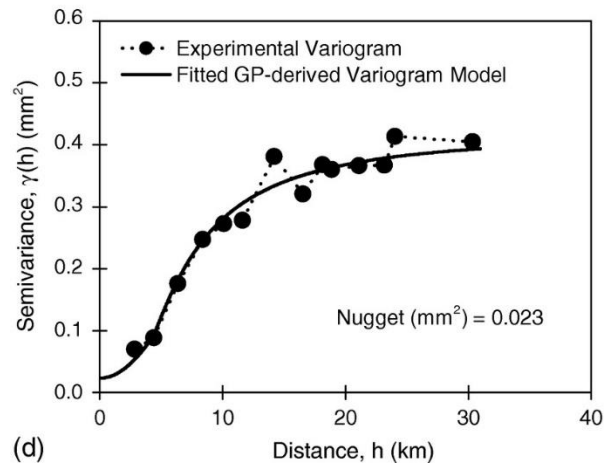
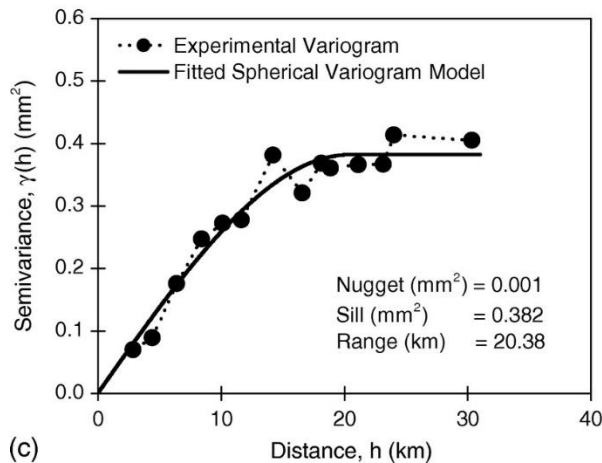
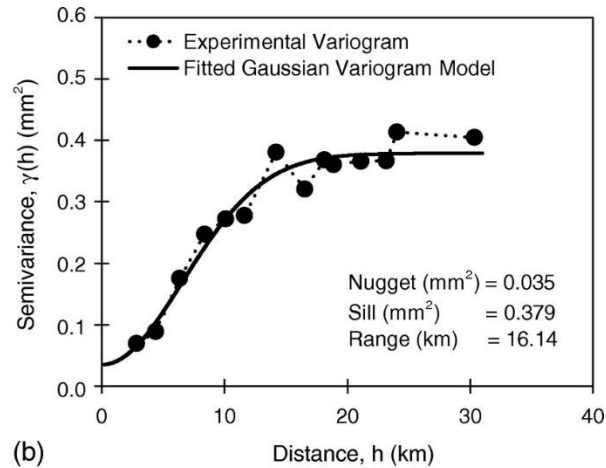
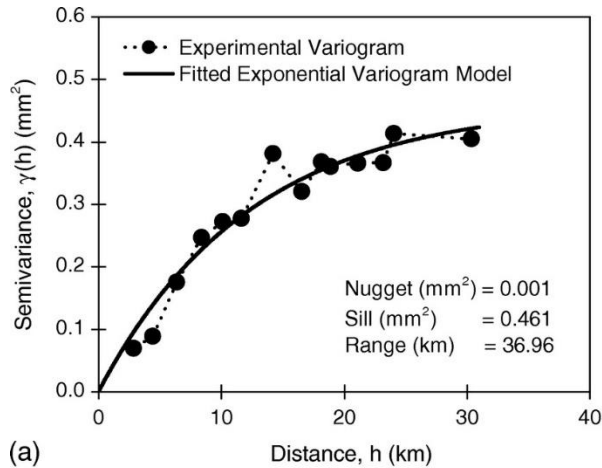
It can be estimated from the data (under certain conditions).

Empirical variogram

$$\hat{\gamma}(h) = \frac{1}{N_h} \sum_{x_i - x_j = h} (Z_i - Z_j)^2.$$

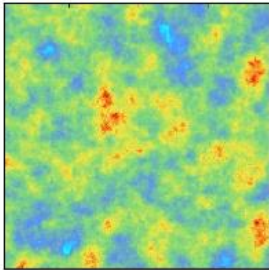
After the empirical variogram has been computed, a theoretical model needs to be fit to it. This model needs to satisfy certain properties, e.g. conditional negative definiteness.

Kriging

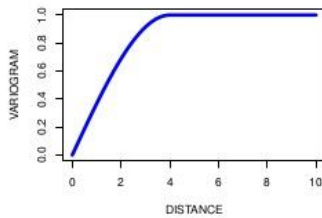


Kriging

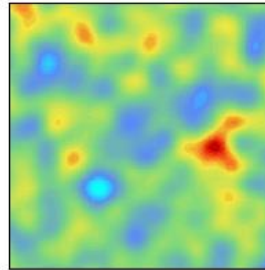
Rough



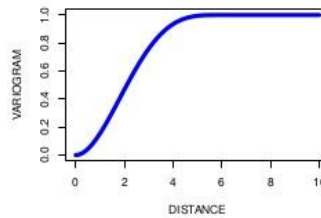
Spherical model



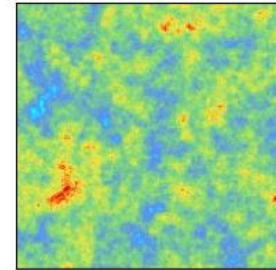
Smooth



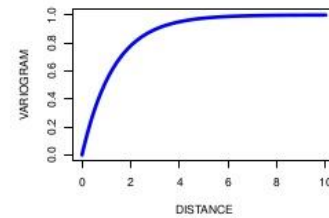
Cubic model



Rough

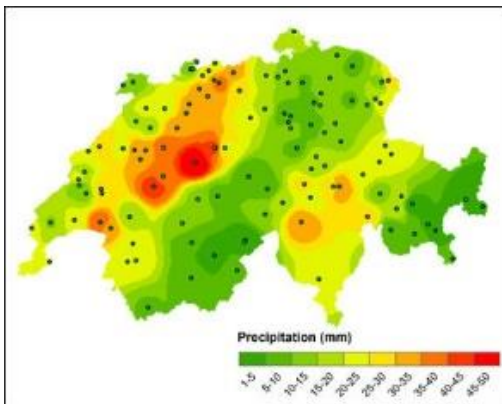
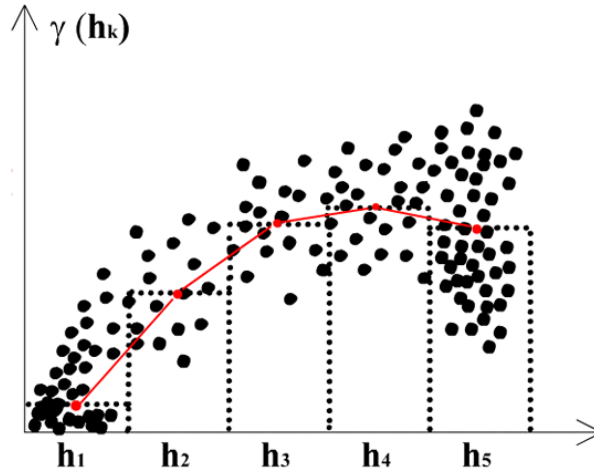


Exponential model

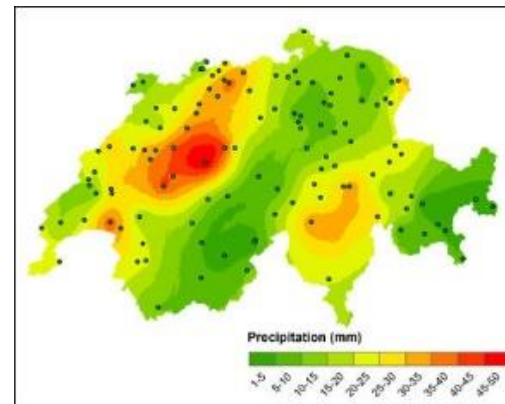


From a presentation by H. Wackernagel

Example IDW, Kriging

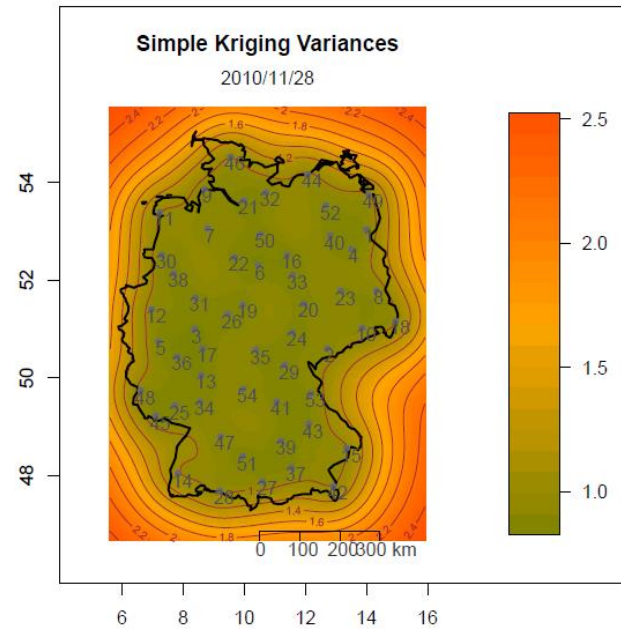
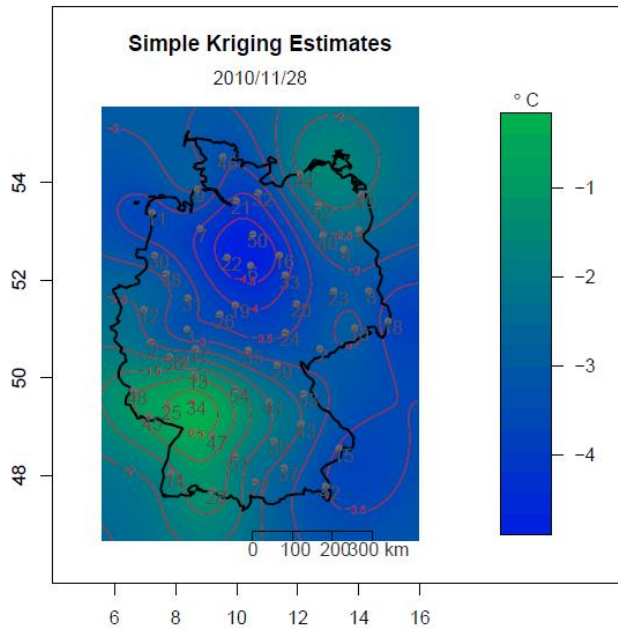


IDW



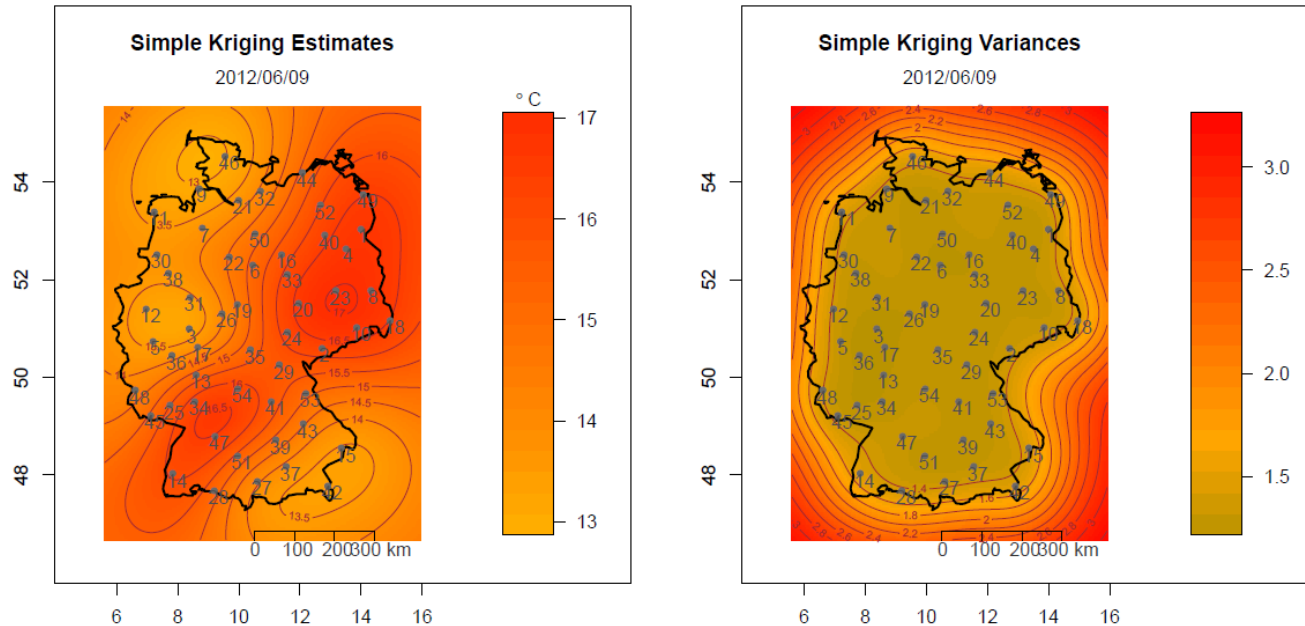
Kriging

Example Kriging



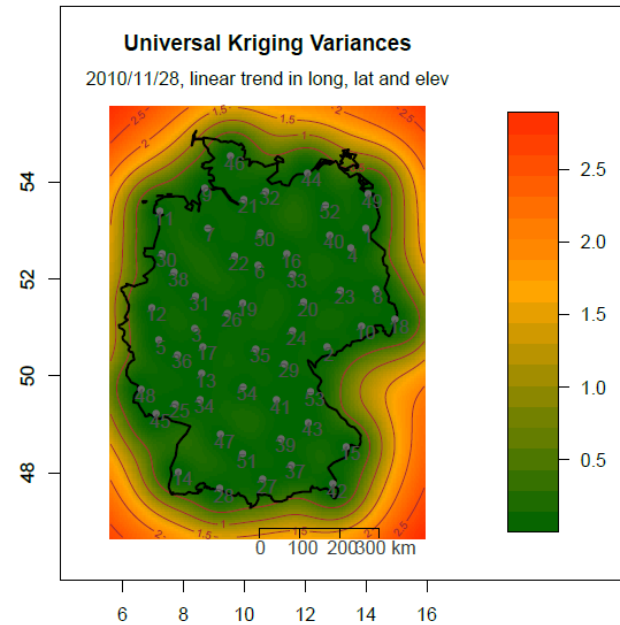
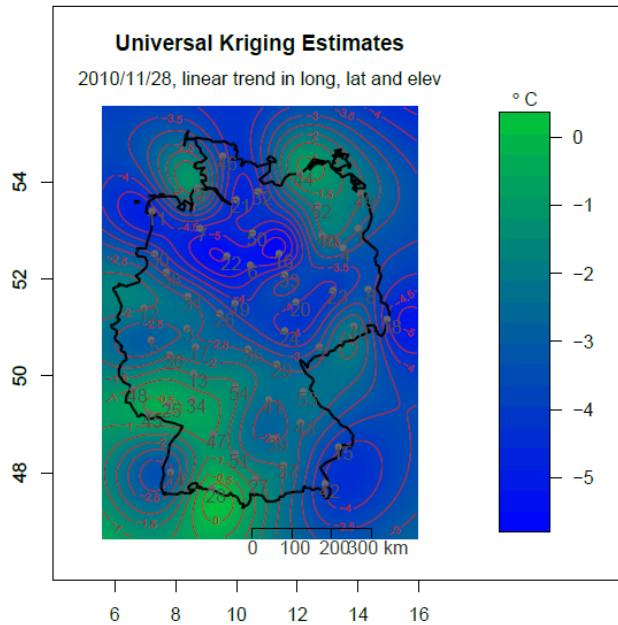
Example from the Bachelor thesis "Kriging methods in spatial statistics", Andreas Lichtenstern, TU München

Example Kriging



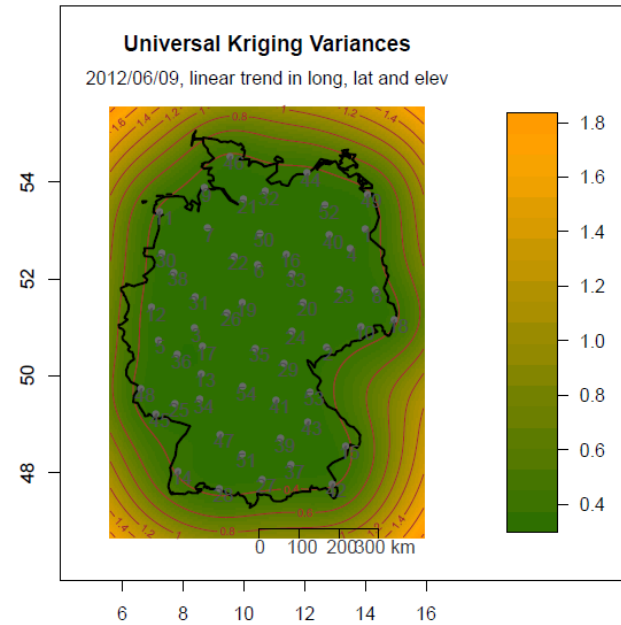
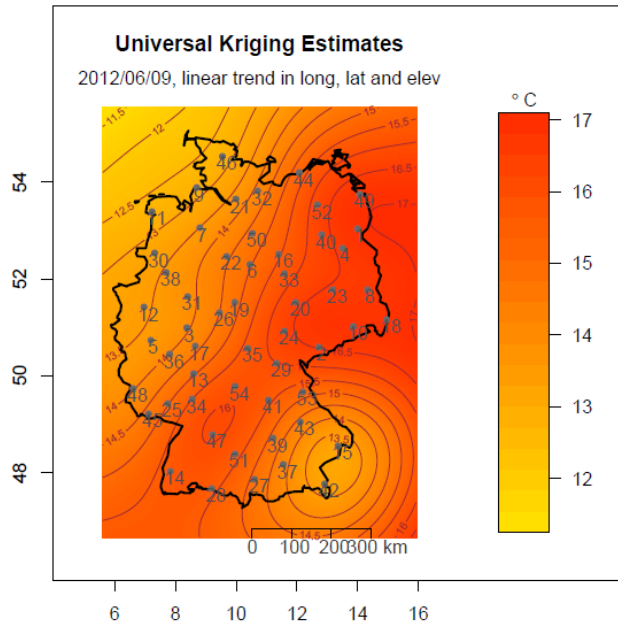
Example from the Bachelor thesis "Kriging methods in spatial statistics", Andreas Lichtenstern, TU München

Example Kriging



Example from the Bachelor thesis "Kriging methods in spatial statistics", Andreas Lichtenstern, TU München

Example Kriging



Example from the Bachelor thesis "Kriging methods in spatial statistics", Andreas Lichtenstern, TU München

Ausgangslage:

Gegeben sind bekannte Werte y_i (z.B. Messungen) für einen Parameter an gegebenen Positionen x_i . An beliebigen Positionen x sollen die ggf. unbekanntes Werte y vorhergesagt werden.

Interpolation:

Interpolation umfasst alle mathematischen Verfahren, welche eine Funktion $f(x) = y$ für eine solche Vorhersage verwenden. Klassischerweise hat bei einer linearen Interpolation $f(x)$ die Form eines gewichteten Mittels der gegebenen Werte y_i :

$$f(x) = \sum_i \lambda_i(x) \cdot y_i$$

mit den Gewichten $\lambda_i(x)$. Die Bestimmung der Gewichte hängt vom gewählten Verfahren ab.

Approximation und Interpolation:

Für eine klassische (exakte) Interpolation gilt, dass das Verfahren die Daten exakt reproduziert mit

$$f(x_i) = y_i$$

Bei einer Approximation (oder auch nicht-exakte Interpolation) ist dies nicht der Fall, die Daten werden hier nur angenähert mit

$$f(x_i) = y_i + \epsilon_i \rightarrow f(x_i) \approx y_i$$

ϵ_i ist dabei der unbekannte Approximationsfehler.

Interpolation und Extrapolation:

Findet die Vorhersage innerhalb der Daten statt, handelt es sich um eine Interpolation. Die Ergebnisse können zumeist als plausibel angesehen werden, da genügend Daten als Randbedingungen für die Vorhersage zu Verfügung stehen.

Findet die Vorhersage außerhalb der Daten oder sehr weit entfernt statt, handelt es sich um eine Extrapolation. Die Ergebnisse können zumeist nicht als plausibel angesehen werden und sollte sehr kritisch interpretiert werden, da die Daten nur sehr eingeschränkt als Randbedingungen verwendet werden können.

Lokale und globale Interpolation:

Bei einer lokalen Interpolationen werden nicht alle Datenwerte für die Interpolation verwendet. Dies bedeutet, dass es Datenwerte y_i gibt, für das Interpolationsgewicht gleich Null ist:

$$\exists i: \lambda_i(x) = 0$$

Bei einer globalen Interpolationen werden immer alle gegebenen Werte berücksichtigt, d.h. dass es keine Interpolationsgewichte gibt, welche Null sind:

$$\forall i: \lambda_i(x) \neq 0$$

Gitterfreie und gitterbasierte Interpolation:

Benötigt ein Interpolationsverfahren eine Vermaschung der Datenpunkte wird es als *gitterbasiert* bezeichnet. Es ist dann zumeist ein lokales Verfahren, welches nur die Datenwerte verwendet, welche sich auf die Zelle beziehen, in der der Interpolationspunkt x liegt.

Gitterfreie (grid-free) Verfahren benötigen keine Vermaschung, sind aber zumeist globale Verfahren.

Deterministische Interpolation:

Die Bestimmung der Gewichte basiert auf einem festgelegten mathematischen Modell. Es ist zumeist nicht initial möglich, die „Unsicherheit“ der Vorhersage abzuschätzen.

(Geo-)statistische Interpolation (z.B. kriging):

Die Bestimmung der Gewichte basiert zusätzlich auf der bekannten oder abgeschätzten räumlichen Korrelation der Daten. Dies ist zumeist aufwändiger als deterministische Vorhersage, erlaubt aber die zusätzliche Abschätzung der „Unsicherheit“ der Vorhersage.

Wahl des Interpolationsverfahrens:

Bei der Wahl des Vorhersageverfahrens sollten folgende Fragen berücksichtigt werden:

1. Entspricht das mathematische Modell des Verfahrens dem vorherzusagenden Parameter bzgl.
 - Skala (diskret/kontinuierlich/...)
 - Glattheit / Differenzierbarkeit ...
 - ... ?
2. Ist Extrapolation für Anwendung notwendig?
3. Ist für die Anwendung exakte Interpolation notwendig oder reicht auch Approximation?
4. Welche Komplexität des Verfahrens ist für die Daten/Anwendung angemessen?
5. ...

1. Definitionen, Funktionen, Anwendungen (Vorlesung 1)
2. Koordinatensysteme und -transformationen (Vorlesung 2+3)
3. Räumliche Datenmodellierung (Vorlesung 4 - 6)
4. Vermaschungen (Vorlesung 7 - 8)
5. Räumliche Interpolation (8 - 10)
- 6. Transformationen, Filtermethoden**

Filter Methods

Instead of the true image, we typically only know a **noisy image**

$$f = f_0 + n,$$

where the noise n can be of different type (typically Gaussian). **Denoising** describes the problem of obtaining a $g \approx f_0$ from knowledge of f .



Original Image

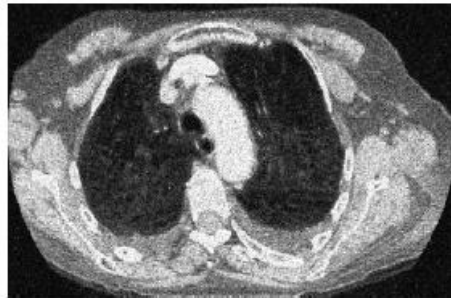


Image with Gaussian noise



Image with Salt and Pepper noise

Filter Methods

Instead of the true image, we typically only know a **noisy image**

$$f = f_0 + n,$$

where the noise n can be of different type (typically Gaussian). **Denoising** describes the problem of obtaining a $g \approx f_0$ from knowledge of f .

Denoising is often an initial step for further image processing.



Original Image

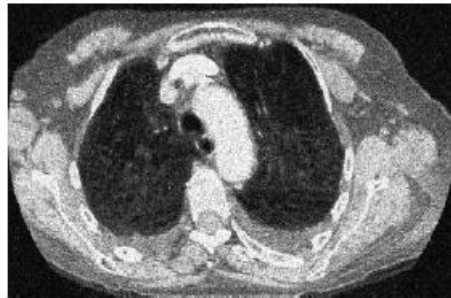


Image with Gaussian noise

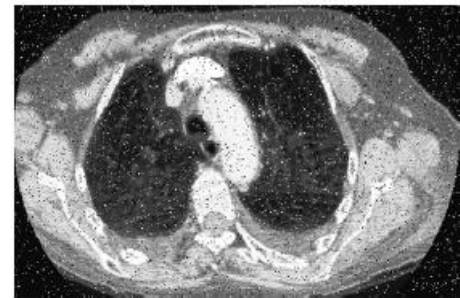


Image with Salt and Pepper noise

Filter Methods

Convolution: Given a kernel $w = [0, 1]^2 \rightarrow \mathbb{R}$, the convolution of f against w is defined via

$$g(x, y) = (w * f)(x, y) = \int_{[0,1]^2} w(x - u, y - v) f(u, v) du dv.$$

Fourier Transform: The Fourier transform of f defined via

$$F(u, v) = \hat{f}(u, v) = \int_{[0,1]^2} e^{-2\pi i(x,y) \cdot (u,v)} f(x, y) dx dy.$$

Convolution and Fourier Transform: The Fourier transform of a convolution has the property

$$G(u, v) = W(u, v)F(u, v).$$

Filter Methods

Convolution: Given a kernel $w = [0, 1]^2 \rightarrow \mathbb{R}$, the convolution of f against w is defined via

$$g(x, y) = (w * f)(x, y) = \int_{[0,1]^2} w(x - u, y - v) f(u, v) du dv.$$

Fourier Transform: The Fourier transform of f defined via

$$F(u, v) = \hat{f}(u, v) = \int_{[0,1]^2} e^{-2\pi i(x,y) \cdot (u,v)} f(x, y) dx dy.$$

Convolution and Fourier Transform: The Fourier transform of a convolution has the property

$$G(u, v) = W(u, v)F(u, v).$$

The choice of the kernel w or W in spatial or spectral domain, respectively, determines the properties of the convolution (filter).

Filter Methods

Representation in a discrete (pixel) setup:

Convolution: Given a kernel $w \in \mathbb{R}^{(2K+1) \times (2L+1)}$, the convolution of f against w is defined via

$$g(m, n) = (w * f)(m, n) = \sum_{k=-K}^K \sum_{l=-L}^L w(k, l) f(m - k, n - l)$$

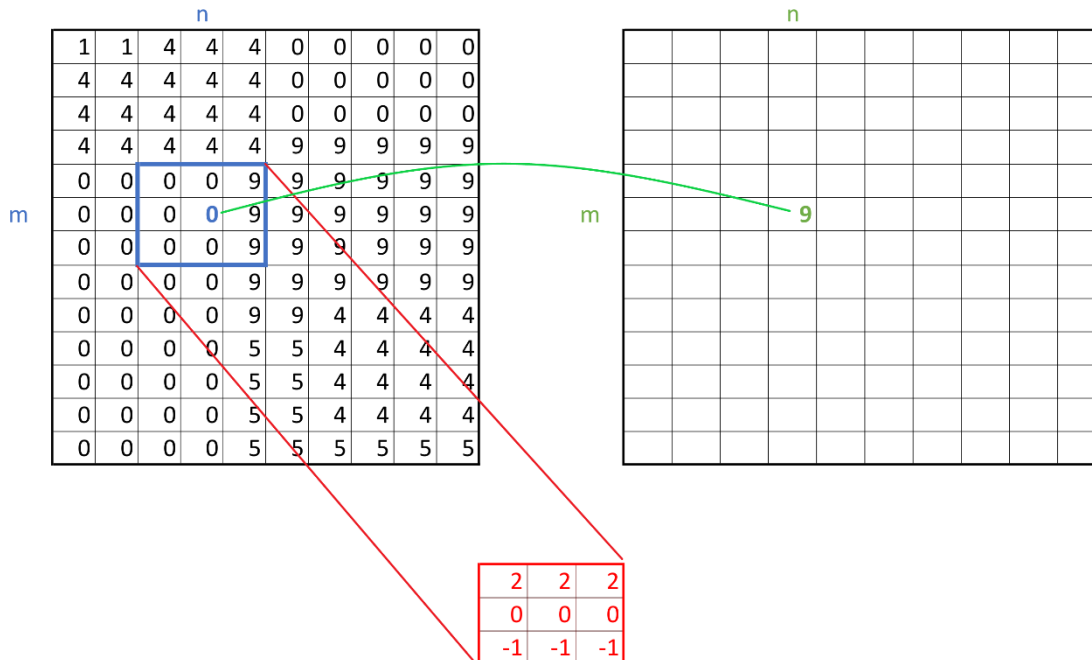
1	1	4	4	4	0	0	0	0	0
4	4	4	4	4	0	0	0	0	0
4	4	4	4	4	0	0	0	0	0
4	4	4	4	4	9	9	9	9	9
0	0	0	0	9	9	9	9	9	9
0	0	0	0	9	9	9	9	9	9
0	0	0	0	9	9	9	9	9	9
0	0	0	0	9	9	9	9	9	9
0	0	0	0	9	9	4	4	4	4
0	0	0	0	5	5	4	4	4	4
0	0	0	0	5	5	4	4	4	4
0	0	0	0	5	5	4	4	4	4
0	0	0	0	5	5	5	5	5	5

Filter Methods

Representation in a discrete (pixel) setup:

Convolution: Given a kernel $w \in \mathbb{R}^{(2K+1) \times (2L+1)}$, the convolution of f against w is defined via

$$g(m, n) = (w * f)(m, n) = \sum_{k=-K}^K \sum_{l=-L}^L w(k, l) f(m - k, n - l)$$



Filter Methods

Examples for kernels w :

Averaging Kernel in a $(2K + 1) \times (2L + 1)$ -neighbourhood. For $L = K = 1$:

$$w = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Gaussian Kernel with weight function $w(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}}$. For $L = K = 2$ and $\sigma = 1$:

$$w = \frac{1}{271} \begin{pmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{pmatrix}$$

Gaussian Filtering can be regarded as a special case of diffusion filtering.

Filter Methods

Original Image



Image with Gaussian noise



Gauss Filtered Image



Average Filtered Image



Examples for nonlinear methods:

Median Filter: For each pixel (m, n) we define a local environment $LA(m, n) = \{m - K, \dots, m + K\} \times \{n - L, \dots, n + L\}$ and choose the median intensity in that environment:

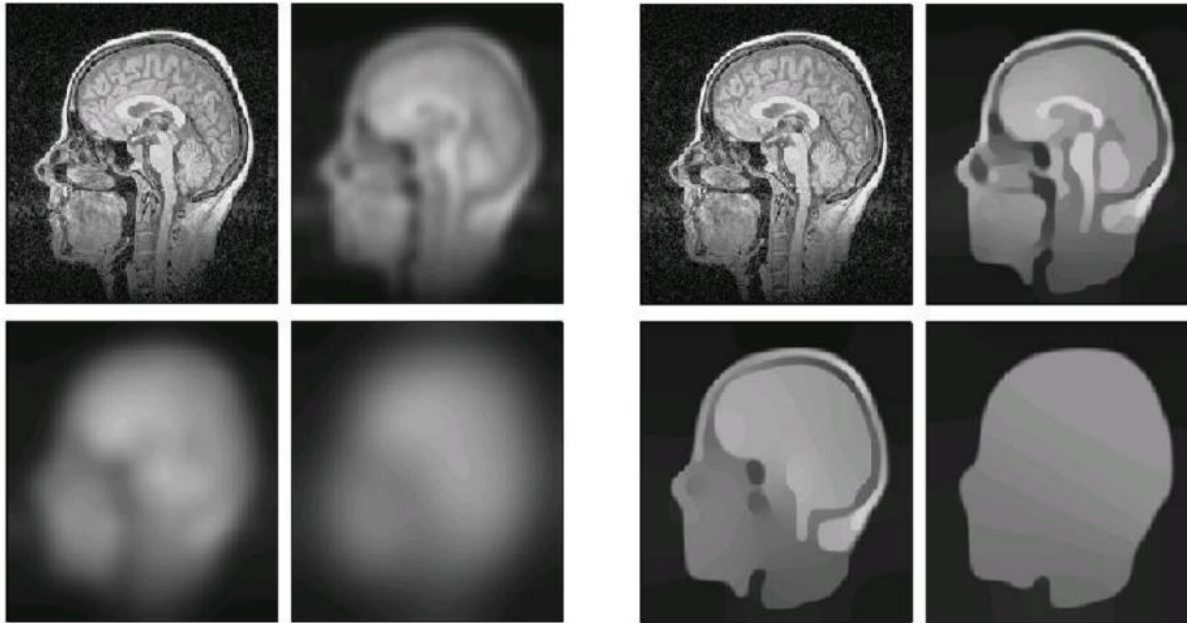
$$g(m, n) = \text{median}\{f(k, l) : (k, l) \in LA(m, n)\}$$

Nonlinear Diffusion: Let D denote the diffusion coefficient, which may depend on g and ∇g (latter accounts for 'edges'), then the diffusion equation reads

$$\begin{aligned} \frac{d}{dt}g(t, x, y) &= \nabla \cdot (D(g, \nabla g))\nabla g(t, x, y) \\ g(0, x, y) &= f(x, y). \end{aligned}$$

A popular choice for D is $D(|\nabla g|) = e^{-\frac{|\nabla g|^2}{2\sigma^2}}$ (Perona-Malik diffusion)

Filter Methods



Scale space behaviour of **linear** Diffusion filtering

Scale space behaviour of **Nonlinear** Diffusion filtering

'Anisotropic Diffusion in Image Processing', J. Weickert, Teubner, 1998

Where else could nonlinear diffusion filtering be used?

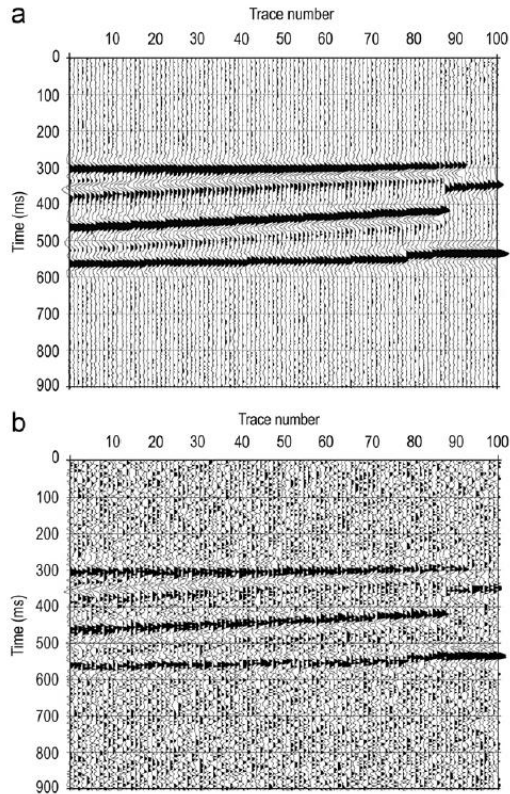


Fig. 3. Synthetic section contaminated with 10% (a) and 50% (b) random noise.

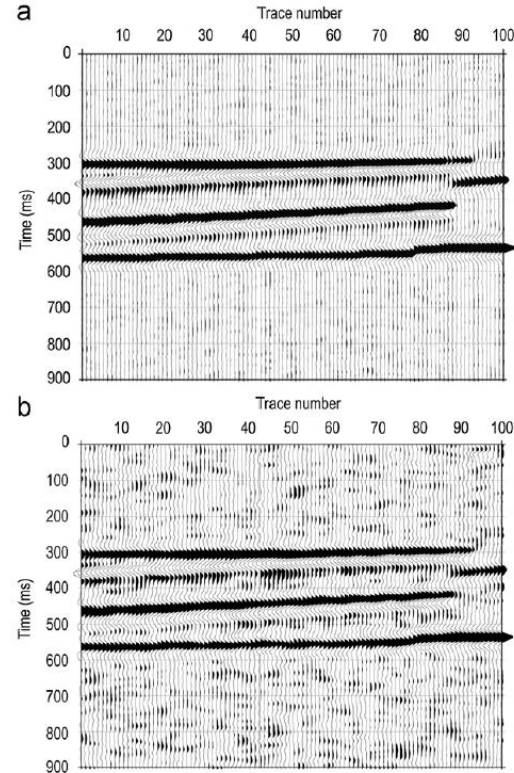
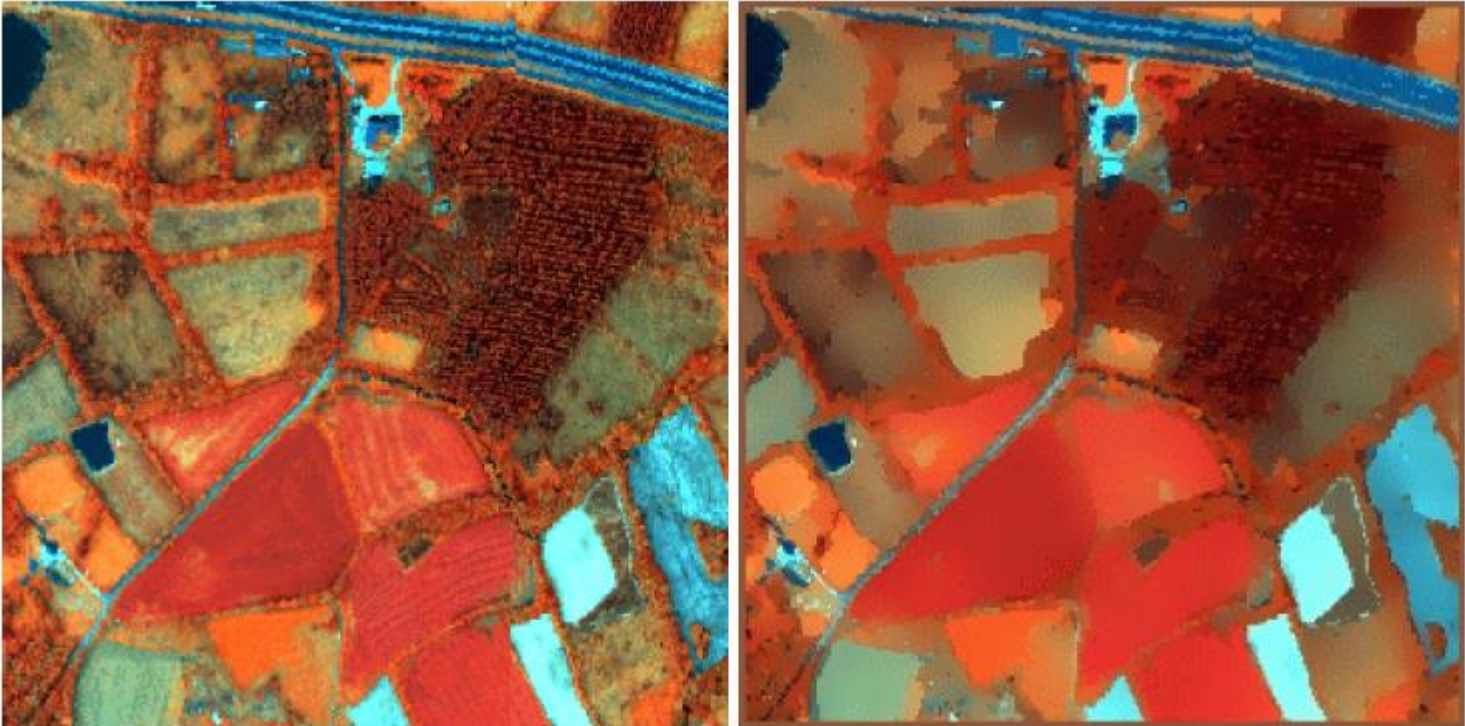


Fig. 4. Anisotropic non-linear diffusion filtering results: (a) section contaminated with 10% random noise and (b) section contaminated with 50% noise.

'Seismic noise attenuation by means of an anisotropic non-linear diffusion filter', Baddari et al., Computers & Geoscience, 2011

Filter Methods

Where else could nonlinear diffusion filtering be used?



'Nonlinear filtering of hyperspectral images with anisotropic diffusion', Lennon et al., IEEE Int. Geosc. Remote Sensing Symp., 2002