A disposition of interpolation techniques

Martin Knotters[1], Gerard Heuvelink[1]

[1]Soil Science Centre

Wageningen University and Research Centre - Alterra

Aim

- To help PBL in choosing from a large variety of interpolation techniques

Ca. 120 methods, 130 references. 5 reviews, 20 comparisons.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.

Main disposition

- Interpolation in space
  - Methods that quantify accuracy
    - Methods without use of ancillary information
    - Methods using ancillary information
    - Methods incorporating process knowledge
  - Methods that do not quantify accuracy
    - ...
- Interpolation in time
  - ...
    - ...
- Interpolation in space and time
  - ...
    - ...

- Linear interpolation, Triangular Irregular Network
- Inverse distance weighing

$$z(\mathbf{s}_0) = \frac{\sum_{i=1}^{n} \frac{z(\mathbf{s}_i)}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}}, \tag{1}$$

  in which $\mathbf{s}_0$ is the location to be interpolated to, $\mathbf{s}_i, i = 1 \ldots n$ is the n locations where z has been observed, d is the distance and p is the power.

- Nearest neighbour algorithm (Thiessen polygons). The power p in Eq. (1) goes to infinity.

Interpolation in space, no quantified accuracy, no ancillary information

- Linear interpolation, Triangular Irregular Network
- Inverse distance weighing

$$z(\mathbf{s}_0) = \frac{\sum_{i=1}^{n} \frac{z(\mathbf{s}_i)}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}}, \tag{1}$$

  in which s_0 is the location to be interpolated to, $\mathbf{s}_i, i = 1 \ldots n$ is the n locations where z has been observed, d is the distance and p is the power.
- Nearest neighbour algorithm (Thiessen polygons). The power p in Eq. (1) goes to infinity.

- Linear interpolation, Triangular Irregular Network
- Inverse distance weighing

$$z(\mathbf{s}_0) = \frac{\sum_{i=1}^{n} \frac{z(\mathbf{s}_i)}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}}, \tag{1}$$

  in which s_0 is the location to be interpolated to, $s_i, i = 1 \ldots n$ is the n locations where z has been observed, d is the distance and p is the power.

- Nearest neighbour algorithm (Thiessen polygons). The power p in Eq. (1) goes to infinity.

- Linear interpolation, Triangular Irregular Network
- Inverse distance weighing

$$z(\mathbf{s}_0) = \frac{\sum_{i=1}^{n} \frac{z(\mathbf{s}_i)}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}}, \tag{1}$$

  in which s_0 is the location to be interpolated to, $\mathrm{s}_i, i = 1 \ldots n$ is the n locations where z has been observed, d is the distance and p is the power.
- Nearest neighbour algorithm (Thiessen polygons). The power p in Eq. (1) goes to infinity.

Example of IDW of Swiss rainfall data

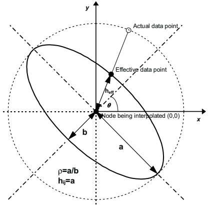


Figure 1. *Illustration of the concept of anisotropy-corrected effective distance.*

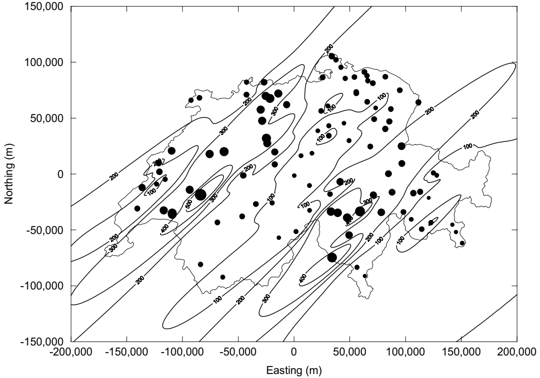
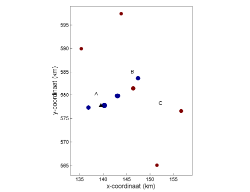


Figure 4. *Predicted rainfall contours based on 100 given training points, superimposed on Switzerland's boarder, shown (size of dots is proportional to the magnitude of the recorded rainfall).*

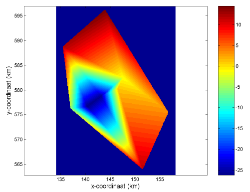
anisotropic IDW

RMSE = 6.3 mm

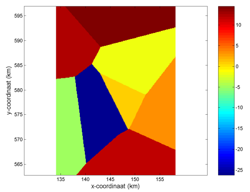
Bodemdaling Waddenzee



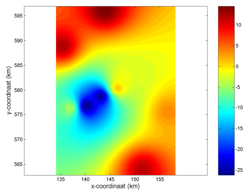
Observations



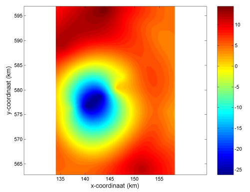
Triangular Irregular Network



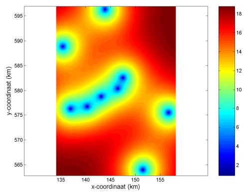
Thiessen polygons



Inverse distance weighing



Ordinary kriging



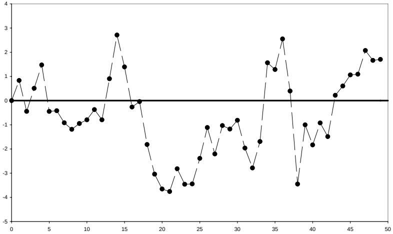
Kriging standard deviation

Interpolation in space, (no) quantified accuracy, no ancillary information

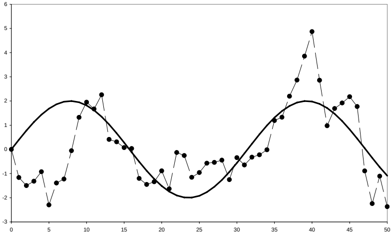
Kriging methods (no secondary variables)

RF-models:

$Z(\mathbf{s}) = m + R(\mathbf{s})$,
where $\mathbf{s} = x, y$.



$Z(\mathbf{s}) = m(\mathbf{s}) + R(\mathbf{s})$,
where $m(\mathbf{s})$ is a function of the spatial co-ordinates x, y.



Kriging methods (no secondary variables)

- numerical variables: simple kriging, ordinary kriging, etc.
- categorical variables: indicator kriging
- non-Gaussian distributed data: indicator kriging, lognormal kriging, disjunctive kriging, multi-Gaussian kriging, trans-Gaussian kriging
- *exact* interpolator
- short introduction: $E\{Z\}$-kriging (Dennis Walvoort, available at www.ai-geostats.org)

Kriging methods (no secondary variables)

- numerical variables: simple kriging, ordinary kriging, etc.
- categorical variables: indicator kriging
- non-Gaussian distributed data: indicator kriging, lognormal kriging, disjunctive kriging, multi-Gaussian kriging, trans-Gaussian kriging
- *exact* interpolator
- short introduction: $E\{Z\}$-kriging (Dennis Walvoort, available at www.ai-geostats.org)

Kriging methods (no secondary variables)

- numerical variables: simple kriging, ordinary kriging, etc.
- categorical variables: indicator kriging
- non-Gaussian distributed data: indicator kriging, lognormal kriging, disjunctive kriging, multi-Gaussian kriging, trans-Gaussian kriging
- *exact* interpolator
- short introduction: $E\{Z\}$-kriging (Dennis Walvoort, available at www.ai-geostats.org)

Kriging methods (no secondary variables)

- numerical variables: simple kriging, ordinary kriging, etc.
- categorical variables: indicator kriging
- non-Gaussian distributed data: indicator kriging, lognormal kriging, disjunctive kriging, multi-Gaussian kriging, trans-Gaussian kriging
- *exact* interpolator
- short introduction: $E\{Z\}$-kriging (Dennis Walvoort, available at www.ai-geostats.org)

Kriging methods (no secondary variables)

- numerical variables: simple kriging, ordinary kriging, etc.
- categorical variables: indicator kriging
- non-Gaussian distributed data: indicator kriging, lognormal kriging, disjunctive kriging, multi-Gaussian kriging, trans-Gaussian kriging
- *exact* interpolator
- short introduction: $E\{Z\}$-kriging (Dennis Walvoort, available at www.ai-geostats.org)

Kriging methods (no secondary variables)

- numerical variables: simple kriging, ordinary kriging, etc.
- categorical variables: indicator kriging
- non-Gaussian distributed data: indicator kriging, lognormal kriging, disjunctive kriging, multi-Gaussian kriging, trans-Gaussian kriging
- *exact* interpolator
- short introduction: $E\{Z\}$-kriging (Dennis Walvoort, available at www.ai-geostats.org)

Example Ordinary Kriging vs. Indicator Kriging

Swiss rainfall data

Anisotropic variograms

Ordinary kriging	Indicator kriging

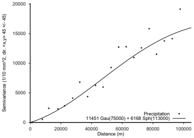


Figure 3. *Sample variogram for 45° (+ symbols) with fitted model (solid curve).*

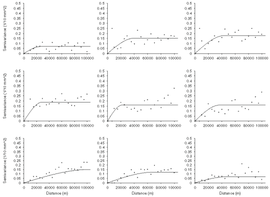


Figure 7. *Sample indicator variograms for 45° (+ symbols) with fitted models (solid curves).*

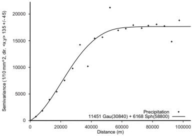


Figure 4. *Sample variogram for 135° (+ symbols) with fitted model (solid curve).*

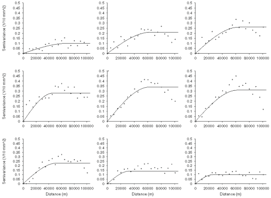


Figure 8. *Sample indicator variograms for 135° (+ symbols) with fitted models (solid curves).*

Example Ordinary Kriging vs. Indicator Kriging

Maps of Swiss rainfall

Ordinary kriging	Indicator kriging

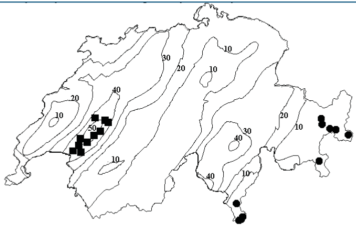


Figure 5. *Isolines of OK estimates, with a 10 mm interval. The squares and circles show the locations of the ten maximum and ten minimum estimates respectively.*

RMSE = 5.97 mm

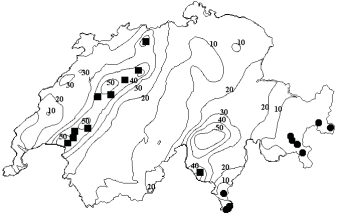


Figure 9. *Isolines of IK estimates, with a 10 mm interval. The squares and circles show the locations of the ten maximum and ten minimum estimates respectively.*

RMSE = 6 mm

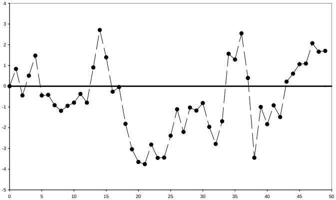
Kriging methods (secondary variables)

- Stratified kriging
- Cokriging methods

$$Z_i(\mathbf{s}) = m_i(\mathbf{s}) + R_i(\mathbf{s}), \text{ with } i = 1 \cdots n \text{ variables}$$

- Combinations of regression and kriging:
  - Kriging combined with linear regression/kriging with uncertain data

RF-model: $Z(\mathbf{s}) = m + R(\mathbf{s})$

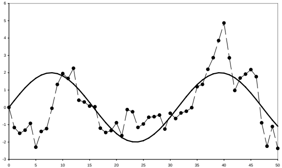


Accuracy of the data on z is known, and accounted for in the kriging system.

Kriging methods (secondary variables), continued

- Combinations of regression and kriging (continued):
  - Universal kriging, kriging with an external drift, kriging with a trend model, kriging with a guess field, regression kriging, residual kriging

    RF-model: $Z(\mathbf{s}) = m(\mathbf{s}) + R(\mathbf{s})$

    $m(\mathbf{s})$ is a function of secondary variables. Data on the secondary variable(s) need to be available at the prediction points!

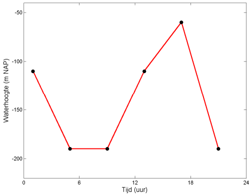
Other methods

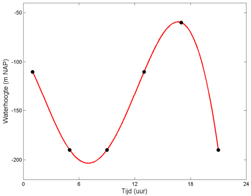
- Bayesian Maximum Entropy (BME)
- Markov Random Fields (MRF)

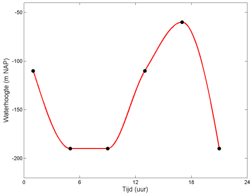
BME and MRF can be applied to both numerical and categorical variables.

Brus, D.J. and G.B.M. Heuvelink, 2007. *Towards a Soil Information System with quantified uncertainty. Three approaches for stochastic simulation of soil maps.* Wageningen, WOt-rapport 58.

Splines: quantified accuracy or not?







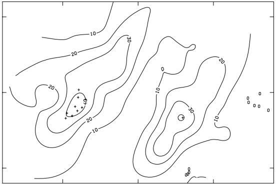


Figure 9. *Isolines of the untransformed interpolated values (mm), overlaid with the 10 smallest and 10 largest estimated rainfall data values.*

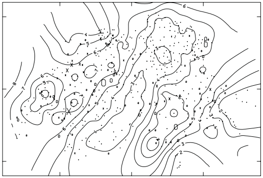


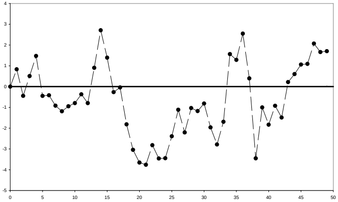
Figure 8. *Proportional symbol map of untransformed data residuals, overlaid with isolines of estimated standard errors (mm).*

RMSE = 5.6 mm

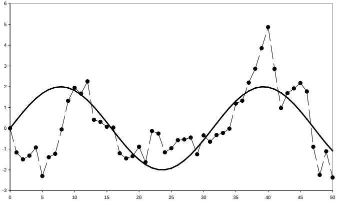
Smoothing spline: niet exact door de waarnemingen (meetfout)

Kriging in time

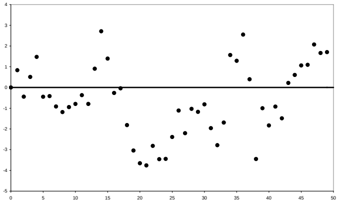
RF-model: $Z(t) = m + R(t)$



RF-model: $Z(t) = m(t) + R(t)$



Data-filling, ARIMA-modelling



AR(1) model:

$$z_t - \mu = \phi_1(z_{t-1} - \mu) + \epsilon_t$$

$\mu = 0,\ \phi_1 = 0.8,\ \sigma_\epsilon^2 = 1$

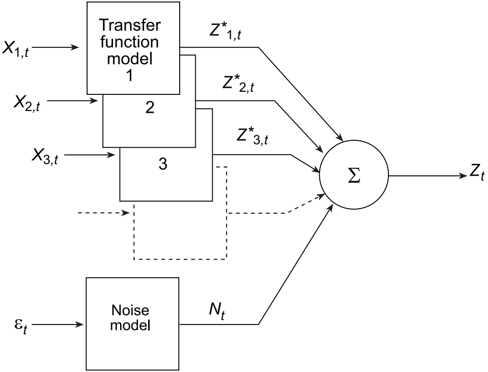
ARMA(1,1) model:

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \epsilon_t - \theta_1\epsilon_{t-1}$$

SARIMA(p,d,q)×(P,D,Q) model:

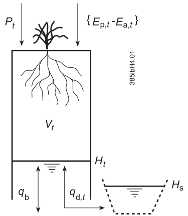
$$(\nabla^d \nabla_s^D Z_t - \mu) = \frac{\phi(B)\Phi(B^s)}{\theta(B)\Theta(B^s)}\epsilon_t$$

Data-filling, TFN-modelling



$$Z_t = Z^*_{1,t} + Z^*_{2,t} + \cdots + Z^*_{n,t} + N_t$$

Physically based time series models



$$H_t - \mu = a_1(H_{t-\triangle t} - \mu) + b_0 P_t + \epsilon_t,$$

with

$$\begin{aligned} a_1 &= \mathrm{e}^{-\triangle t/(\varphi\gamma)}, \\ b_0 &= \gamma(1 - a_1), \\ \mu &= \gamma q_{\mathrm{b}} + H_{\mathrm{s}}, \\ \epsilon &= \gamma\left\{[E_{\mathrm{p}}(t) - E_{\mathrm{a}}(t)] - \frac{\triangle V}{\triangle t}\right\}\left\{1 - \mathrm{e}^{-\triangle t/(\varphi\gamma)}\right\} \end{aligned}$$

State-space approach

State equation:

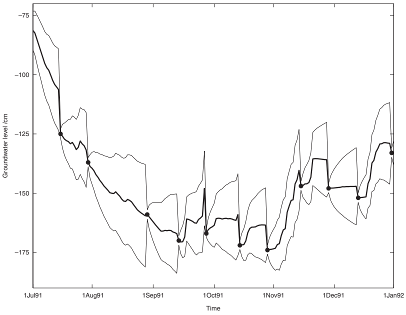
$$Z(t) = g\left\{Z(t-1)\right\} + \epsilon(t)\,,$$

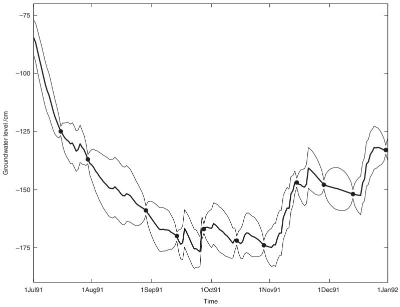
measurement equation:

$$Y(t) = h\left\{Z(t)\right\} + \eta(t),$$

where $Y(t)$ is the measurement, and $\eta(t)$ is the measurement error.
Kalman-filtering: update of prediction of $Z(t)$ by using all past and present observations $Y(t)$ with known accuracy.
Kalman-smoothing: predicting the present state from past, present and future observations.

State-space approach





Kalman filtering

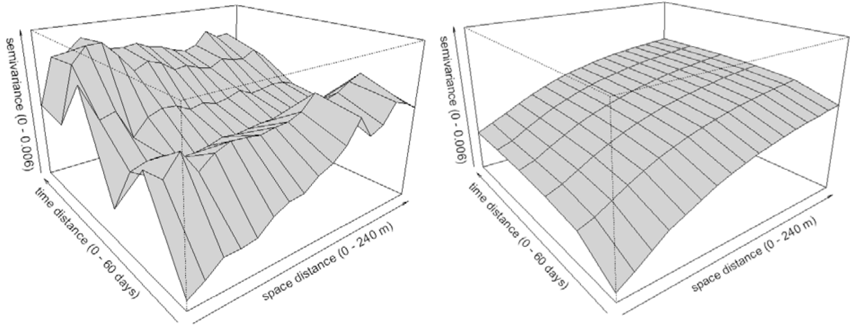
Kalman smoothing

Variance of measurement error = 1 cm^2

Space-time kriging

RF-model:

$$Z(\mathbf{s}, t) = m(\mathbf{s}, t) + R(\mathbf{s}, t)$$



Normalized Difference Vegetation Index

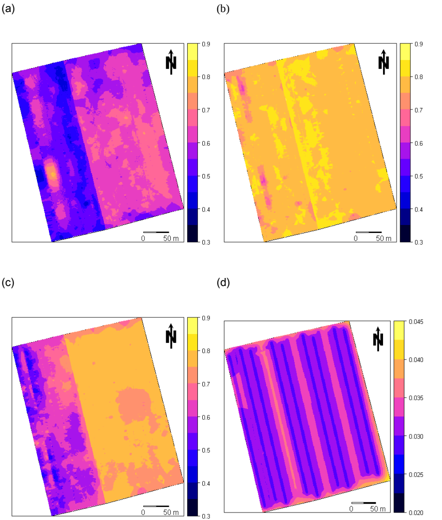


Figure 8. Space-time kriged predictions for three arbitrary days: (a) DOY 165, (b) DOY 200 and (c) DOY 235; (d) kriging standard deviations for DOY 235.

Regionalized time series models

$$H_t - \mu = a_1(H_{t-\triangle t} - \mu) + b_0 P_t + \epsilon_t$$

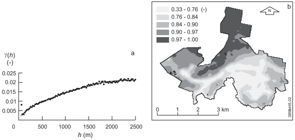


Figure 5.2: Spatial structure of $\tilde{a}_1(\mathbf{u})$, **a** *Sample variogram (dots) and fitted variogram model (line),* **b** *Map of* $\tilde{a}_1(\mathbf{u})$.

a_1

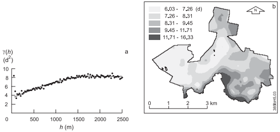


Figure 5.3: Spatial structure of $\tilde{b}_0(\mathbf{u})$, **a** *Sample variogram (dots) and fitted variogram model (line),* **b** *Map of* $\tilde{b}_0(\mathbf{u})$.

b_0

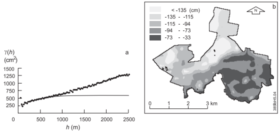


Figure 5.4: Spatial structure of $\tilde{\mu}(\mathbf{u})$, **a** *Sample variogram (dots) and fitted variogram model (line),* **b** *Map of* $\tilde{\mu}(\mathbf{u})$.

μ

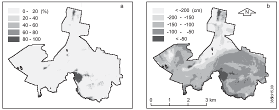


Figure 5.9: **a** *Map of the risk that at April 1st in a future year the water table depth will be shallower than 50 cm;* **b** *Map of the expected water table depth at April 1st in a future year. Values in cm.*

maps of water table depths

Other methods

- State-space approach
- Optimal interpolation, variational methods

Criteria:

- applicability in space, time and space-time;
- quantification of accuracy of interpolated values;
- applicability at numerical and/or categorical variables;
- incorporation of ancillary information;
- incorporation of process knowledge;
- applicability for up- and downscaling;
- complexity of application, and required computation time;
- constraints on the size and conditions of the dataset;
- availability of implementations by means of software tools.