A statistical model for spatial resolution enhancement

Alfred Stein
ITC

Spatial resolution enhancement - why?

- Some measurements and models act at points – statements for areas of land are required
- Other observations (remote sensing) concern average values – how to obtain more detailed information
- MAUP: the modifiable area unit problem

MAUP

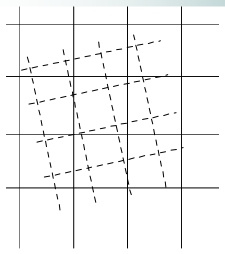
- Study areas can be divided in many ways in non-overlapping units
- Per class accuracies are considerably affected by changing scale and aggregation level
- No unique spatial resolution is appropriate for the detection and discrimination of all geographical entities
- MAUP has an effect on ML image classification
- In remote sensing: pixels correspond to units of information, but different sensors have different pixel sizes → different results

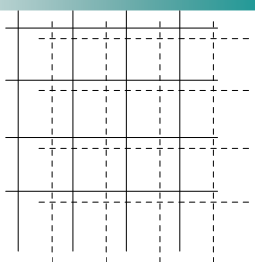
Discrepancies between data need and availability

- In spatial studies we consider the support size δ of the data, i.e. the spatial extent for which data are representative.
- We define the scale as the spatial resolution r.
- Resolution enhancement applies to data at a coarse scale (level I, r_1) towards a fine scale of interest (level II, r_2) where $r_1 \geq r_2$

A first approach: remote sensing

- Essential to overlay different grids
  - data from different images with the same sensor may lead to a different grid
  - data from different sensors may have different resolutions
- Coal fires
- Locusts
- Paddy field area
- Sub-pixel land cover resolution





Resampling methods

- Methods:
- Nearest neighbor: only the closest observation
- Bilinear interpolation: a plane through the four neighborhood points
- Bicubic interpolation: fit a pair of 3rd degree polynomials through 16 nb points
- Pixel unmixing

A statistical model

- We consider two scale levels
- The true spatial pattern x is modeled in a way that is independent of the two scale levels.
- In the simplest situation, we take x to be a Gaussian random field.
- This random field is determined by the parameters μ (trend), λ as variance parameter in the covariance function and K (discretization).
- The second characteristic is the spatial variable $y(s)$, governed by random function x, data support H and prior variance κ^{-1}. Further, x results into the scaled field Ax.
- Distinction into a continuous and a discrete model.

The continuous model

Data $y(s)$ are considered with s referring to a point or a contiguous area
A field $x=(x(s),s\in S)$ is defined over a region $S \subset R^2$.

$$y(s) = \int H(s)\, x(s)\, ds$$

where $H(s)$ determines the relation between $y(s)$ and the Gaussian random field.
The continuous model is then defined as

$$(Hx)_i = \int_{s\in S} H_i(s)\, x(s)\, ds$$

The observed data $y=y(s_i), i=1,2,\ldots, n$ are modeled as Gaussian and independent, given x, with constant *a priori* variance κ^{-1} and expectation linear in x:

$$y \approx N(H x,\ \kappa^{-1} I).$$

For $i = 1,\ldots,n$ the point spread functions $H_i(s)$ encode the *support* of the data.

- $H_i(s)$ is constant on $[s_i-\delta, s_i+\delta] \times [s_j-\delta, s_j+\delta]$ and 0 elsewhere.
- $H_i(s)$ will be close to a delta function at s_i if y_i is measured at s_i and represents an extremely localized quantity at that position,
- $H_i(s)$ will be non-zero over some neighborhood surrounding s_i if y_i represents a quantity that is naturally spatially aggregated

To make inferences about x on a possibly different resolution, we consider linear functionals A on X, so that Ax is the vector

$$Ax = \left(\int_{s\in S} A_j(s)\, x(s)\, ds \right)_{j=1}^{J}$$

Typically, interest concerns inference about (an aggregated version of) x over the whole region S
Therefore, the $A_j(s)$ will have supports forming a partition of S.
For example, in R^2, $A_{ij}(s) = m^2$ on $[(i-1)/m, i/m] \times [(j-1)/m, j/m]$ and 0 otherwise.

The discrete model

Lattice data are defined at the nodes of a fine-mazed grid, $T = \{1,\ldots,m_r\} \times \{1,\ldots,m_c\}$.

At n grid nodes or small contiguous sets of nodes a positive value occurs.

The discrete model has for x a finite-dimensional Gaussian random vector, with x_k corresponding to the value at location t_k, k=1,2,…,$K = m_r \cdot m_c$.

The linear functionals H and A become matrices H_{ik} and A_{jk}, of dimensions $n \times K$ and $J \times K$, where J is the number of data after scaling.

A model for scaling then becomes

$y_i = \Sigma_j H_{ij}\, x(s_j) + \varepsilon_i$

The matrix H thus transforms the data lattice T_1 towards lattice T_2.

The $x(s_j)$ represent the random function at the nodes of T_1, the index j takes these values in some predefined order, and the index i runs over T_2 also in some predefined order.

Correlation function models

Let the covariance function as a function of s and t be defined as $\text{cov}_Y(s,t) = E(Y(s)\cdot Y(t)) - \mu_Y(s)\cdot\mu_Y(t) = \text{cov}_Y(|s\text{-}t|)$. Then the correlation function equals

$$K(s,t) = \frac{K_Y(s,t)}{\sqrt{\sigma_Y(s)\sigma_Y(t)}}$$

where $\sigma_Y(s) = \sigma_Y(t) = K_Y(0)$ equals the variance at locations s and t, respectively.

Several functional models are permissible as a correlation function.

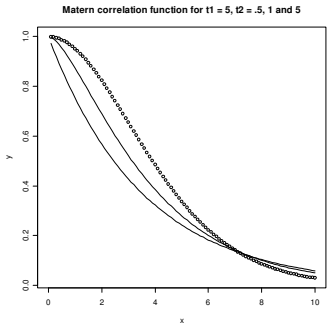
Matèrn correlation function

The Matèrn correlation function is defined as

$$K_\theta(s,t) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)}\cdot\left(\frac{|s-t|}{\theta_1'}\right)^{\theta_2} K_{\theta_2}\left(\frac{|s-t|}{\theta_1'}\right)$$

The Bessel K-function of order θ_2, depends on scale parameter θ_1 and smoothness parameter θ_2, whereas $\theta_1' = \theta_1/(2\sqrt{\theta_2})$.

- ✓ If $\theta_2 = 0.5$, then $K_\theta(s,t)$ equals the exponential correlation function $K_{1/2}(s,t) = \exp\{-|s\text{-}t|/\theta_1'\}$;
- ✓ For $\theta_2 \to \infty$, $K_\theta(s,t)$ equals the Gaussian correlation function $K_\infty(s,t) = \exp\{-(s\text{-}t)^2/\theta_1'^2\}$.



Maximum likelihood estimation

Maximum likelihood estimators of β for fixed θ and λ are given by

$$\hat{\beta} = (X'G^{-1}X)^{-1}X'G^{-1}Y$$

with the matrix G given by $\lambda\cdot K_\theta(s,t)$, X by possible covariates and Y containing the data.

The profile likelihood for θ and λ, apart from an additive constant equals

$$\hat{l}^* = -\tfrac{1}{2}\log|G| - \tfrac{1}{2}n\log(Y - X\hat{\beta})'G^{-1}(Y - X\hat{\beta})$$

In the discrete model, the probability distribution for the observations equals $y \sim N(0, (\kappa H^T H + \lambda G)^{-1})$.

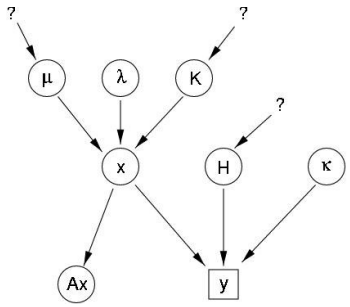
This allows us to write down the likelihood function as

$$p(y \mid \theta) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{\sqrt{|B_\theta|}} \cdot \exp(-\tfrac{1}{2}(y-\mu)^T B_\theta^{-1}(y-\mu))$$

where $\theta = (\theta_1, \theta_2, \kappa, \lambda)$, the vector of four parameters describing the covariances and $B_\theta = \kappa H^T H + \lambda G^{-1}$.

- Prior information for the various parameters may result from hierarchical assumptions, i.e.from any particular distribution assumed to be valid.
- As an alternative, such information may be obtained from source scale levels.
- In this case, prior data for μ, λ and K are derived at level I, and at level II for the matrix K and the function H.

Directed a-cyclical graph



Algorithm

- From the *n* data and the problem definition the resolution δ and the scale size and number of discretization steps *m* follow.
- The matrix *B* is filled, using the (Matèrn) correlation function.
- For the numerical integration we let the step size *h* depend upon *n* and the scale *m* as $m = 1 + \delta/h$ and $n = 1 + \varepsilon/h$, depending on the support size of the data.
- The correlation is computed for the step sizes provided by these values of *m* and *n*.

- The profile likelihood is solved for any set of covariables.
- Find the optimum - we used the ve08ad routine, as implemented, which to our experience is extremely sensitive to good prior estimates and to a similar range of changes in parameters.
- Apply the results in a scaling routine, allowing for any discretisation, either coarser or finer than the grid of the data.

- As an alternative to likelihood estimation, the variogram is estimated using squared pair differences of observations group in a range of succeeding distance classes with bin width x_i:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (y(s_i) - y(s_i + h))^2$$

Here, $y(s_i)$ and $y(s_i+h)$ is a pair of points separated by a distance between $h-\xi$ and $h+\xi$, the total number of pairs being equal to $N(h)$.

Distributions

In both continuous and discrete space versions, inference about A*x* given *y* follows from the standard results about conditional distributions in multivariate normal distributions:

$$E(Ax \mid y) = E(Ax) + \operatorname{cov}(Ax, y)\operatorname{var}(y)^{-1}(y - E(y))$$

$$\operatorname{var}(Ax \mid y) = \operatorname{var}(Ax) - \operatorname{cov}(Ax, y)\operatorname{var}(y)^{-1}\operatorname{cov}(y, Ax)$$

In the discrete case, suppose that $E(x)=\mu$ and $\operatorname{var}(x)=\Sigma$ are known; then $E(y)=H\mu$, $\operatorname{var}(y) = \kappa^{-1}I+H\Sigma H^T$ and $\operatorname{cov}(A x,y)= A\Sigma H^T$, so that

$E(A x \mid y) = A\mu + A\Sigma H^T(\kappa^{-1}I+H\Sigma H^T)^{-1}(y-H\mu)$
$\operatorname{var}(A x \mid y) = A\Sigma A^T - A\Sigma H^T(\kappa^{-1}I+H\Sigma H^T)^{-1}H\Sigma A^T$.

In the continuous case, suppose that $\mu(s)=E(x(s))$ and $\Sigma(s,s') = \operatorname{cov}(x(s),x(s')) = \lambda \cdot K(s,s')$ are known; then

$A\mu_j = \int_{s\in S} A_j(s)\,\mu(s)ds$

$A\Sigma H^T)_{ji} = \int_{s\in S}\int_{s'\in S} A_j(s)\,\Sigma(s,s')\,H_i(s')\,ds\,ds'$

$(H\Sigma H^T)_{ji} = \int_{s\in S}\int_{s'\in S} H_j(s)\,\Sigma(s,s')\,H_i(s')\,ds\,ds'$

and

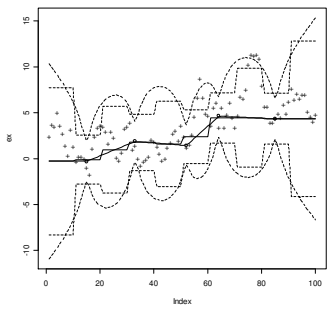
$(A\Sigma A^T)_{ji} = \int_{s\in S}\int_{s'\in S} A_j(s)\,\Sigma(s,s')\,A_i(s')\,ds\,ds'$

With $B = \operatorname{var} y \in M_{n,n}$, $D = \operatorname{cov}(AX,AX) \in M_{m,m}$ and $C = \operatorname{cov}(AX,y) \in M_{m,n}$ and assuming an isotropic covariance model, i.e. $K(s,t) = K(|s-t|)$ this yields the following matrix version for scaling:

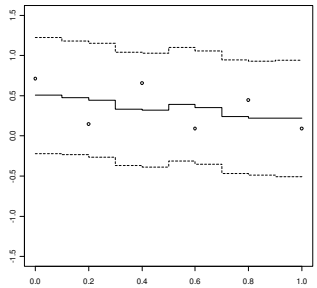
$E(A x|y) = A y + C \cdot B^{-1}(y-H\mu)$
$\operatorname{var}(A x \mid y) = D-C\,B^{-1}C^T$

To evaluate matrices B, C and D, the double integrals have to be evaluated. In most practical studies numerical integration procedures have to be applied.

Example with n = 100, m = 10

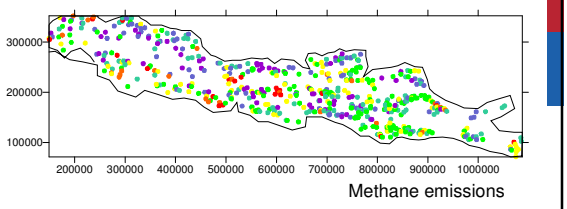


Example with n = 5, m = 10



First application

Prior estimates were obtained on the basis of collected data at the rice field level. Values for $\mu = 4.3 \cdot 10^4$ and $\kappa = 1.908$ equal the mean and the reciprocal of the variance at the data level, the value of λ is obtained from the estimated covariance function. Values for H and K are given by the target level requirements.

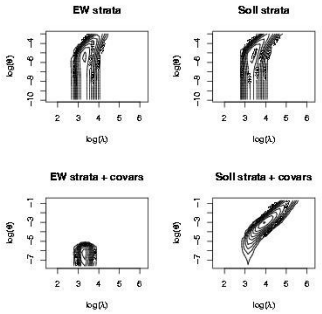


Methane emissions

Covariance function estimates

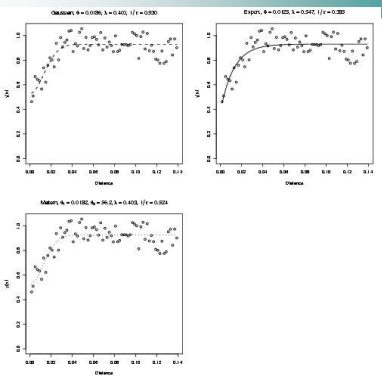
- Likelihood functions for the parameters λ and θ of the exponential covariance function using the profile likelihoods are obtained.
- For the east-west strata, or the soil strata, a low value of θ indicates that the data are almost independent, with a parameter $\lambda \approx 30$.
- Spatial structure occurs, though, after inclusion of [Fe] and [OC] into the profile.
- In a combination with the EW-strata, the optimum $(\theta, \lambda) = (0.0025,6)$, and in a combination with the soil strata $(\theta, \lambda) = (0.018,13.5)$.

Profile likelihoods for θ and λ with different profiles.



Estimated correlation function coefficients

	Gaussian	Exponential	Matèrn
θ_1	0.0186	0.0125	0.0182
θ_2	-	-	56.2
λ	0.400	0.547	0.403
$1/\kappa$	0.530	0.385	0.524



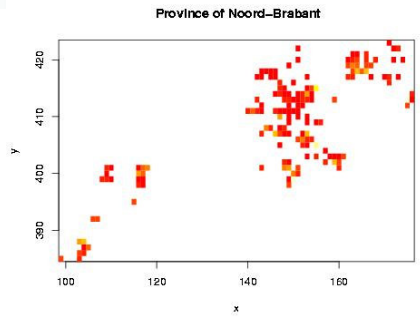
Application 2:

- Data on heavy metals (Zn, Cd) are collected at the provincial level at coordinate precision of 1 km
- Data are required at the city level (Oss), resolution of 1 m
- Test data are available

Data and aim

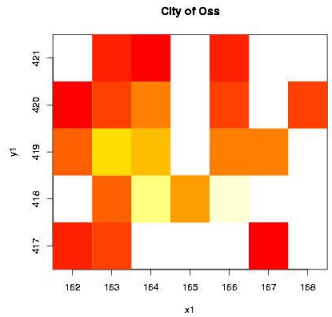
- In the province of Noord-Brabant (appr. 5000 km^2) in the southern Netherlands data on several metals have been collected from groundwater monitoring wells.
- The aim: to investigate the distribution of these metals under agricultural and urban use of the land at the provincial level. Coordinates were recorded at a 1 km precision (δ = 1 km); n = 161 locations with average concentrations in 1 km^2 grid cells.

Prior information was obtained from the data at the 1 km interval: μ = 250 ppm for Zn and 1 ppm for Cd, λ = 0.48 for Zn and 8736 for Cd equals sill - nugget parameter of the covariance function; K equals the number of 10 $\times$ 10 m cells in the discretized target space, κ equals the reciprocal of the variances; H equals the function with support at the 10 $\times$ 10 cells at level II.



Data and aim

- Within the city of Oss, sampling is relatively intense (62 monitoring wells), covering approximately 56 % of the city. Coordinates are available at 1 m precision scale (δ = 1 m).
- For these data, we consider downscaling of the 1 km^2 data towards the 1 m^2 resolution, and use the data available at that resolution for validation.



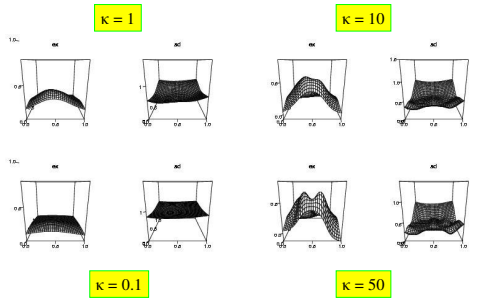
Parameters

Zinc and Cadmium data were downscaled in the Oss area using the Matèrn correlation function with parameter vector $\theta = (\kappa,\lambda,\theta_1,\theta_2)^T = (10, 10, 1, 0.5)^T$ for Zinc and $(10, 0.56, 1, 0.5)^T$ for Cadmium. Notice that the data points are confined to a limited part of that area only, as the extent in the *y*-direction is somewhat smaller than that in the *x*-direction. Data are downscaled towards raster cells of 0.05×0.05, corresponding to an actual resolution of 500 m $\times$ 500 m$ grid cells.

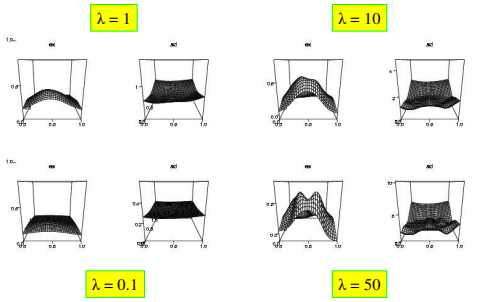
Sensitivity analysis

- We tested the sensitivity of the Matèrn correlation function with parameter vector $\theta = (\kappa,\lambda,\theta_1,\theta_2)^T = (10, 10, 1, 0.5)^T$.
- The upper two figures display the expected values and the standard deviations.
- The high peak in the center that we could observe as well on the original data is well represented, whereas lower values occur closer to the edges.
- The log-likelihood for the choices of θ_1 and θ_2 was equal to -20.5.

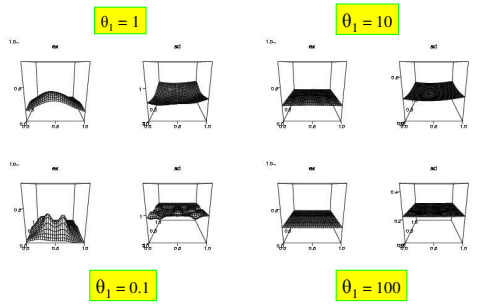
Effects of choices for values of $\kappa = 1$, $\kappa = 0.1$, $\kappa = 10$ and $\kappa = 100$ on downscaling of Zinc data.



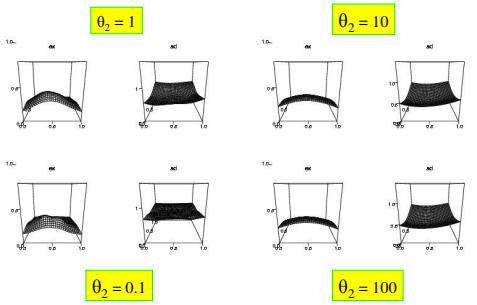
Effects of choices for values of $\lambda = 1$, $\lambda = 0.1$, $\lambda = 10$ and $\lambda = 50$ on downscaling of Zinc data.



Effects of choices for values of $\theta_1 = 1$, $\theta_1 = 0.1$, $\theta_1 = 10$ and $\theta_1 = 100$ on downscaling of Zinc data.



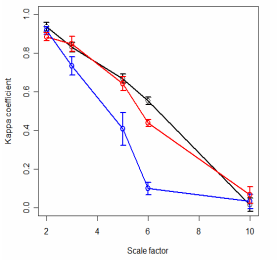
Effects of choices for values of $\theta_2 = 1$, $\theta_2 = 0.1$, $\theta_2 = 10$ and $\theta_2 = 100$ on downscaling of Zinc data.



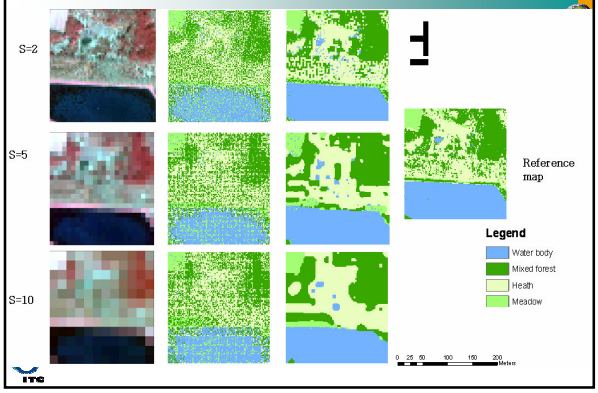
Application from remote sensing

- Classification of a RS image
- Additional field data have been collected
- Suitability of Markov Random Field-based method for Super-Resolution Land Cover Mapping
- A simulated annealing approach has been used
- How does the scale factor S affect the quality of the SRM?

Neighborhood size determination



SRM from remote sensing data



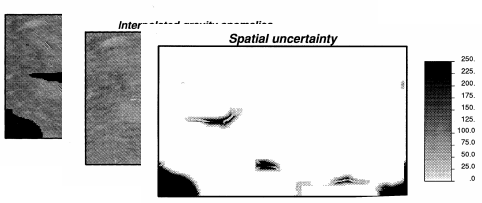
We found that for this study

- Neighborhood size growing with the scale factor performs better than 2nd order neighborhood size.
- Large objects are favored with MRF-based SRM due to the rich contextual information
- The quality of SRM is decreasing with increasing scale factor
- The MRF-based method is suitable for super-reolution land cover mapping however care should be taken on the parameters setting

Prior information

- Prior information and information on related variables may increase the quality of the scaled variables.
- For downscaling, information at level I collected in X_S yields a regression model $f(Xs,\hat{\beta})$
- With data at level II collected in X_D, predictions can be made. An interpolation routine gives estimates at any required level of detail.
- For upscaling, also data at the demand level on related variables can be included.

More examples



Conclusions

- A difference between continuous and discrete scaling models is relevant, the continuous one to be analyzed using random fields, the discrete one by a lattice type of an approach.
- Only minor differences exist between scaling *up* and scaling *down* - a similar statistical model applies.
- Selection of the correct correlation structure in terms of the use of co-variates is essential.
- A sensitivity analysis showed that correlation parameters that describe the shape and the range of dependence are crucial. Those that describe the variance or distinguish between spatial and non-spatial variation appear to be somehow less relevant for scaling.