Kriging and Mixed Effects Models

Some Connections and a Case Study on Soil Degradation

A. Pollice

Dipartimento di Scienze Statistiche

Università degli Studi di Bari

TIES2002
Genova, 21 giugno 2002

PRIN’99: Metodi e tecniche statistiche per l’analisi di dati a struttura spaziale e spazio-temporale

1 Motivation

- Spatial interpolation and prediction for continuous processes are rarely performed with explicit reference to distributional assumptions. Geostatistical practice is seldom based on modern likelihood-based statistical methods.

- Second order assumptions prove to be insufficient to guarantee the optimality of Geostatistical methodologies as minimum variance predictors are only optimal under Gaussianity and when mean and variance parameters are known or considered so.

- Explicit distributional assumptions and the use of likelihood-based inferential procedures lead to the adoption of widely spread model-based statistical techniques in the Geostatistical context.

- Linear mixed effects models provide a convenient representation of spatial processes and efficient estimation and prediction criteria in the classical and Bayesian inferential frameworks. Famous Geostatistical predictors are obtained as special cases.

- Two illustrative case studies on agricultural data are proposed.

2 Stochastic model

- The spatial process $\{Y(\mathbf{t}); \mathbf{t} \in D\}$ is split in a sistematic term or mean effect, a spatially correlated component and a random noise
$$Y(\mathbf{t}) = X(\mathbf{t})\beta + S(\mathbf{t}) + \varepsilon(\mathbf{t})$$
  * $\mathbf{t}$ vector of n sampled spatial locations
  * $X(\mathbf{t}) = X$ matrix of spatially referenced non random variables (coordinates and/or covariates)
  * β *spatial trend* parameter vector
  * $S(\mathbf{t}) = S \sim N_n(\mathbf{0}, \sigma^2 H_{11}(\phi))$ spatial random effect
    σ^2 *partial sill*
    ϕ *range*
    $H_{11}(\phi)$ correlation matrix
      $h(\nu; \phi)$ valid correlation function
      ν distance between locations
  * $\varepsilon(\mathbf{t}) = \varepsilon \sim N_n(\mathbf{0}, \tau^2 I)$ measurement error process
    $\varepsilon \perp S$
    τ^2 *nugget*
- Marginal model
$$Y(\mathbf{t}) = Y \sim N_n(X\beta, \Sigma_{11})$$
  * $\Sigma_{11} = \tau^2 I + \sigma^2 H_{11}(\phi)$

3 ML estimation

- The n-dimensional data vector $y(\mathbf{t}) = \mathbf{y}$ is a single realization of the spatial process $\{Y(\mathbf{t}); \mathbf{t} \in D\}$.

- ML or GLS estimator for the *mean parameter* β (BLUE conditionally on Σ_{11})

$$\widehat{\beta} = \left(X^T \Sigma_{11}^{-1} X\right)^{-1} X^T \Sigma_{11}^{-1} \mathbf{y}$$

  * $\text{cov}(\widehat{\beta}|\Sigma_{11}) = (X^T \Sigma_{11}^{-1} X)^{-1}$ negatively biased estimate of $\text{cov}(\widehat{\beta})$

- ML estimates of spatial *covariance parameters* $\varpi = (\sigma^2, \phi, \tau^2)$ are obtained conditionally on $\widehat{\beta}$ by numeric iterative methods. They tend to be negatively biased as a consequence of considering $\widehat{\beta}$ as fixed.

- Unbiased estimates of spatial *covariance parameters* ϖ are obtained by iterative maximization of the *REML* likelihood function

$$L(\varpi; \mathbf{y} - X\widehat{\beta}) = \frac{L(\varpi, \beta; \mathbf{y})}{L(\varpi, \beta; \widehat{\beta})}$$

4 Prediction model

- Model for m unsampled locations $\mathbf{t}_0$

$$Y_0 = X_0\beta + S_0 + \varepsilon_0$$

  * $Y(\mathbf{t}_0) = Y_0$ unobserved realizations of the spatial process

  * X_0 matrix of coordinates/covariates for unsampled locations

  * $S_0 \sim N_m(\mathbf{0}, \sigma^2 H_{00}(\phi))$

$$\text{cov}(S, S_0) = \sigma^2 H_{10}(\phi)$$

  * $\varepsilon_0 \sim N_m(0, \tau^2 I) \qquad \varepsilon_0 \perp S_0 \qquad \varepsilon_0 \perp \varepsilon$

- Marginal model for observed and unobserved locations

$$\begin{pmatrix} Y \\ Y_0 \end{pmatrix} \sim N_{n+m}\left(\begin{pmatrix} X\beta \\ X_0\beta \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{10} \\ \Sigma_{10}^T & \Sigma_{00} \end{pmatrix}\right)$$

  * $\Sigma_{00} = \tau^2 I + \sigma^2 H_{00}(\phi)$
  * $\Sigma_{10} = \sigma^2 H_{10}(\phi)$

5 Predictions

- Minimum variance predictor conditional on β and ϖ (*simple kriging*)

$$E(Y_0|Y = \mathbf{y}) = X_0\beta + \Sigma_{10}^T\Sigma_{11}^{-1}(\mathbf{y} - X\beta)$$

- Best Linear Unbiased Predictor (BLUP) conditional on ϖ (*universal kriging*)

$$\widehat{Y}_0 = X_0\widehat{\beta} + \Sigma_{10}^T\Sigma_{11}^{-1}(\mathbf{y} - X\widehat{\beta})$$

  * Negative bias due to conditioning on ϖ.
  * MSE$(\widehat{Y}_0|\varpi)$ negatively biased estimate of MSE$(\widehat{Y}_0)$.
  * Exact predictor when $\mathbf{t}_0 = \mathbf{t} \Rightarrow \Sigma_{10} = \Sigma_{11}$

6 Measurement error

- Spatial covariance function $\sigma^2 h(\nu;\phi)$ is continuous at $\nu \to 0$.
- Nugget effect τ^2 is relegated to the additive random noise term ε.
- Covariance Σ_{10} between observations Y and unsampled realizations Y_0 is due to spatial random effects S and S_0 and not to ε
  * $\Sigma_{ii} = \tau^2 I + \sigma^2 H_{ii}(\phi) \qquad i = 0, 1$
  * $\Sigma_{10} = \sigma^2 H_{10}(\phi)$
- *Noiseless predictor* for sampled locations

$$\widehat{Y} = X\widehat{\beta} + \sigma^2 H_{11}\left(\tau^2 I + \sigma^2 H_{11}\right)^{-1}(\mathbf{y} - X\widehat{\beta})$$

  * James-Stein (shrinkage) estimator

$$\widehat{Y} = \mathbf{y} + \left(I - \sigma^2 H_{11}\Sigma_{11}^{-1}\right)(\mathbf{y} - X\widehat{\beta})$$

- *Residuals*

$$\mathbf{e} = \mathbf{y} - \widehat{Y} = -A\mathbf{y}$$

$$\mathbf{e}|\varpi \sim N_n\left(\mathbf{0}, A\Sigma_{11}A^T\right)$$

  * A known function of ϖ and X.

7 Bayesian inference

- Allows obtaining spatial predictors explicitly accounting for uncertainty in the unknown spatial correlation structure. Linear mixed models produce classical Geostatistical results as special cases.

- Response variable Y and parameters β and ϖ are random quantities. External information on model parameters is specified in terms of the *prior distribution* $\pi(\beta, \varpi)$, possibly depending on some *hyperparameters*.

- Prior distribution is updated after data collection: Bayesian inference on parameters β and ϖ is based on numerical summaries of the *posterior distribution* $\pi(\beta, \varpi|\mathbf{y})$ obtainable by Bayes' theorem.

- Bayesian predictions involve the consideration of the *predictive distribution*
$$p(\mathbf{y}_0|\mathbf{y}) = \iint p(\mathbf{y}_0|\mathbf{y}, \beta, \varpi)\pi(\beta, \varpi|\mathbf{y})d\beta d\varpi$$

- *Conjugate* priors induce posterior densities in the functional class of the prior.

- *Noninformative* or flat priors represent a lack of prior information and play a minimal role in the posterior.

8 Simplified Bayesian model

- Avoid prior specification for covariance parameters ϖ

  * $Y|\beta, S \sim N_n(X\beta + S, \tau^2 I)$

  * $\beta \sim N_p(\beta_0, B_0)$

  * $S \sim N_n(\mathbf{0}, \sigma^2 H_{11}(\phi)) \qquad S \perp \beta$

- Joint ditribution of the random vector $(Y, \beta, S)^T$

$$N_{2n+p}\left(\begin{pmatrix} X\beta_0 \\ \beta_0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} XB_0X' + \Sigma_{11} & XB_0 & \sigma^2 H_{11}(\phi) \\ B_0X' & B_0 & \mathbf{O} \\ \sigma^2 H_{11}(\phi) & \mathbf{O} & \sigma^2 H_{11}(\phi) \end{pmatrix}\right)$$

8.1 Simplified Bayesian model - Estimation

- The posterior distribution $p(\beta|\mathbf{y})$ is normal with mean (the Bayesian estimator of the *trend parameter*)

$$E(\beta|\mathbf{y}) = (X'\Sigma_{11}^{-1}X + B_0^{-1})^{-1}(X'\Sigma_{11}^{-1}\mathbf{y} + B_0^{-1}\beta_0)$$

  * When the non-informative prior $B_0^{-1} = \mathbf{O}$ is assumed, the expression of $E(\beta|\mathbf{y})$ reduces to that of the BLUE or GLM estimator.

- When $\beta = \beta_0$ is fixed, the posterior distribution for ϖ is proportional to the likelihood. Bayesian estimates of *co-variance parameters* correspond to ML estimates when the value of β equals a fixed unknown constant β_0.

- When the non-informative prior $\pi(\beta) = 1$ is assumed, the Bayesian estimates of *variance components* correspond to REML estimates.

8.2 Simplified Bayesian model - Prediction

- Extending the simplified model to include unobserved spatial locations, the random vector $(Y, Y_0, \beta, S, S_0)'$ results to be jointly distributed according to a $(2n + 2m + p)$-variate normal.
- The predictive distribution is also Gaussian with mean (the Bayesian *predictor*)

$$E(Y_0|\mathbf{y}) = X_0\beta_0+(XB_0X_0'+\Sigma_{10})'(XB_0X'+\Sigma_{11})^{-1}(\mathbf{y}-X\beta_0)$$

  * The Bayesian predictor reduces to the minimum variance predictor (simple kriging) when $B_0 = \mathbf{O}$, i.e. when a degenerate prior for β is assumed, assigning probability one to a specific value β_0.
  * If a non informative flat prior ($B_0^{-1} = \mathbf{O}$) is used, then it can be shown that the Bayesian predictor is equivalent to the BLUP (universal kriging).

9 Complete Bayesian models

- Taking into account uncertainty on covariance parameters is essential for a full Bayesian treatment of Geostatistical data. Formal Bayesian estimation of the spatial correlation would imply extending the simplified model to include prior specification for ϖ, deriving new expressions for the fixed effects estimator and the predictor, by averaging out these parameters according to their priors.

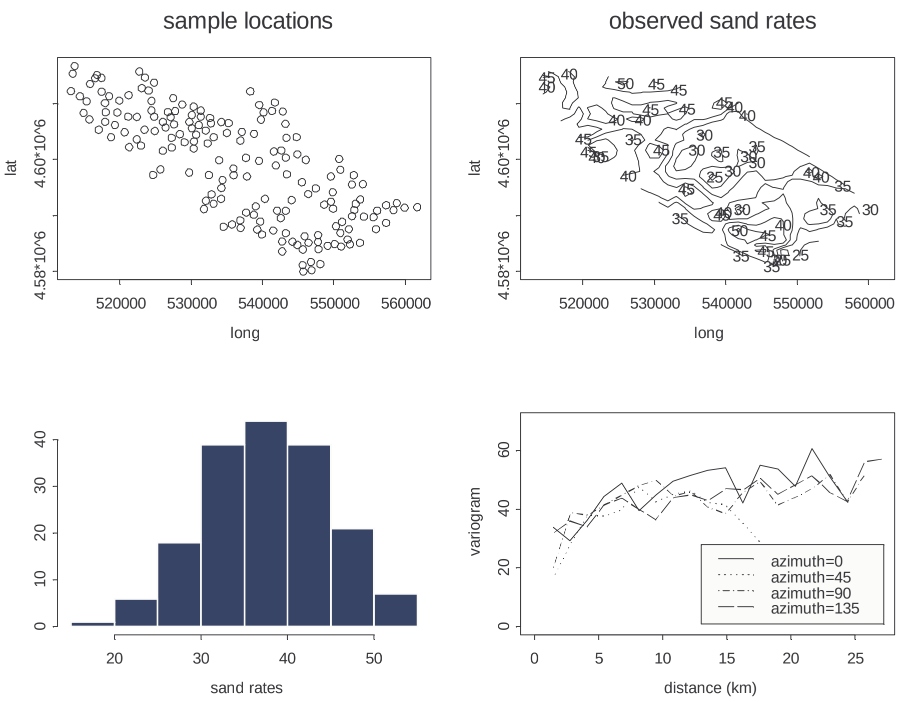
  * Empirical Bayes: estimates of ϖ can be obtained by the gaussian marginal likelihood for ϖ and substituted into former expressions of the Bayesian estimator and predictor (obtained conditionally on values of ϖ) [Searle et al., 1992].

  * Analytical expressions for the predictor and the estimates of parameters β and σ^2 are obtained conditioning on values of ϕ and τ^2, by the jointly conjugate Normal-Inverse Gamma prior for (β, σ^2) [Kitanidis, 1986; Ribeiro and Diggle, 1999].

  * When the spatial correlation structure is assumed unknown in all its parameters $\varpi = (\sigma^2, \phi, \tau^2)$, the posterior $\pi(\varpi|\mathbf{y})$ is not obtainable in closed form for currently used covariance functions and numerical simulation-based integration methods must be used instead [Ecker and Gelfand, 1999; Ribeiro and Diggle, 1999].

10 Case study: soil texture mapping

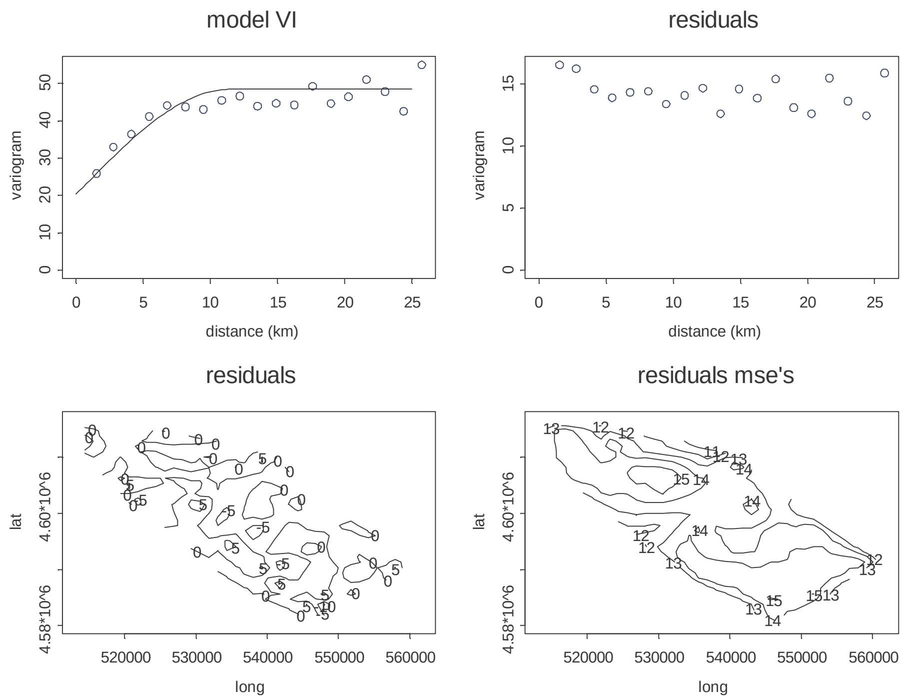
- Composition of the soil texture in sand, silt and clay measured at 175 sample locations over an area of about 75000ha [dataset provided by ”Istituto Sperimentale Agronomico - Bari”].
- Assumptions of Gaussianity, stationarity and isotropy



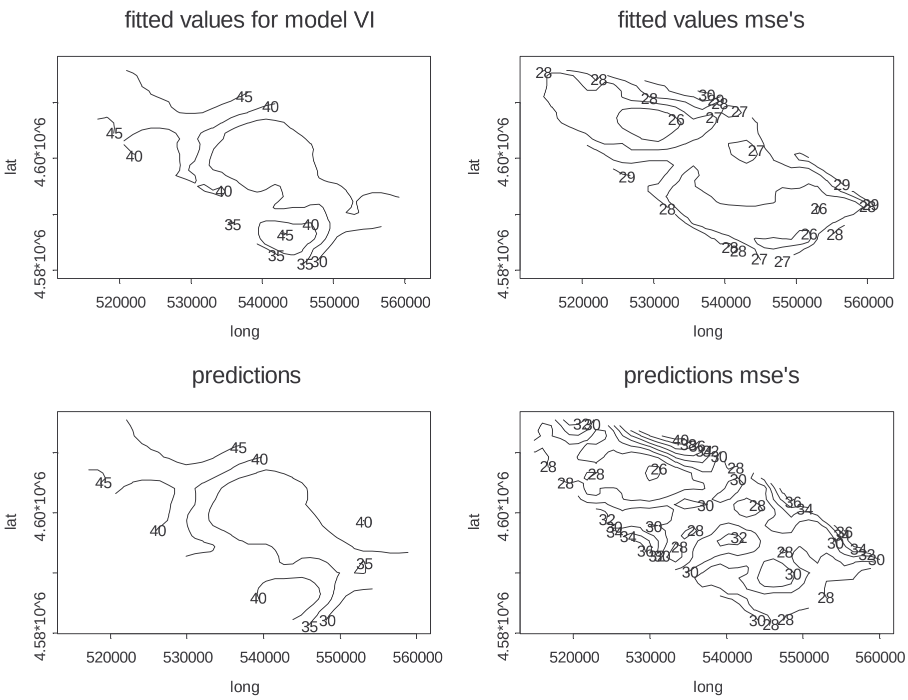
- REML estimates for combinations of three trend surfaces and three covariance structures (SAS PROC MIXED)

	COV	EFF	EST	SE	NUG	SIL	RAN	-2RLL	AIC	BIC
I	exp	long	-0.17	0.135	17.60	28.16	4.35	1112.9	1118.9	1128.4
		lat	0.23	0.172						
II	exp	long	-0.28	0.094	16.81	26.58	3.54	1112.9	1118.9	1128.4
III	exp	lat	0.37	0.149	17.99	30.42	5.08	1112.2	1118.2	1127.6
IV	sph	long	-0.18	0.115	20.46	22.81	9.25	1111.9	1117.9	1127.3
		lat	0.20	0.149						
V	sph	long	-0.28	0.082	22.05	20.58	8.84	1111.7	1117.7	1127.2
VI	sph	lat	0.40	0.134	20.36	28.07	11.42	1111.3	1117.3	1126.8
VII	gau	long	-0.18	0.120	16.81	26.58	3.54	1112.9	1118.9	1128.4
		lat	0.21	0.154						
VIII	gau	long	-0.28	0.080	24.34	17.56	4.24	1112.2	1118.2	1127.7
IX	gau	lat	0.38	0.127	24.96	21.62	5.38	1112.0	1118.0	1127.4

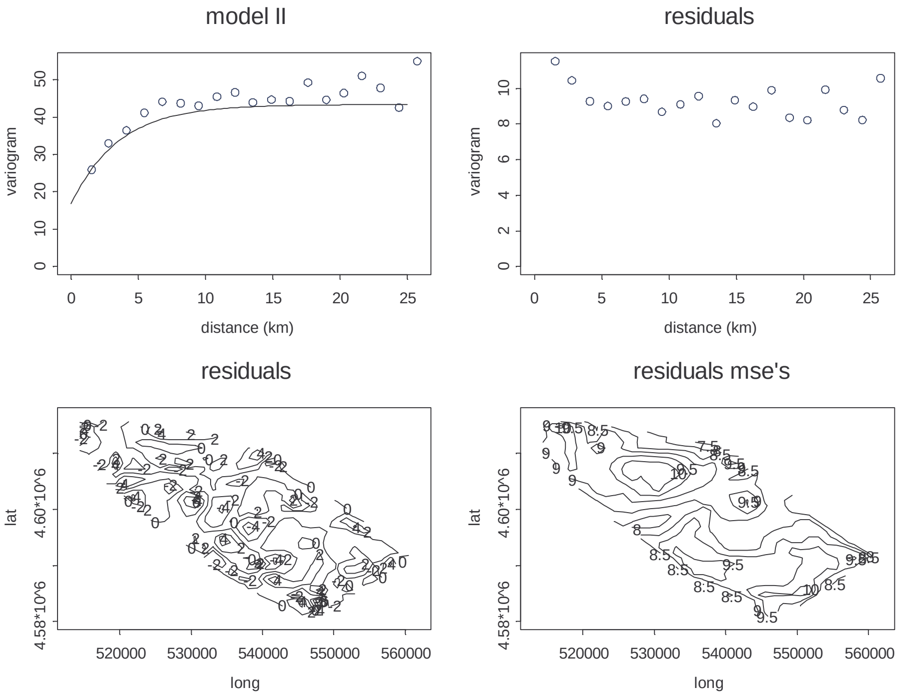
- Assessment of the chosen model VI



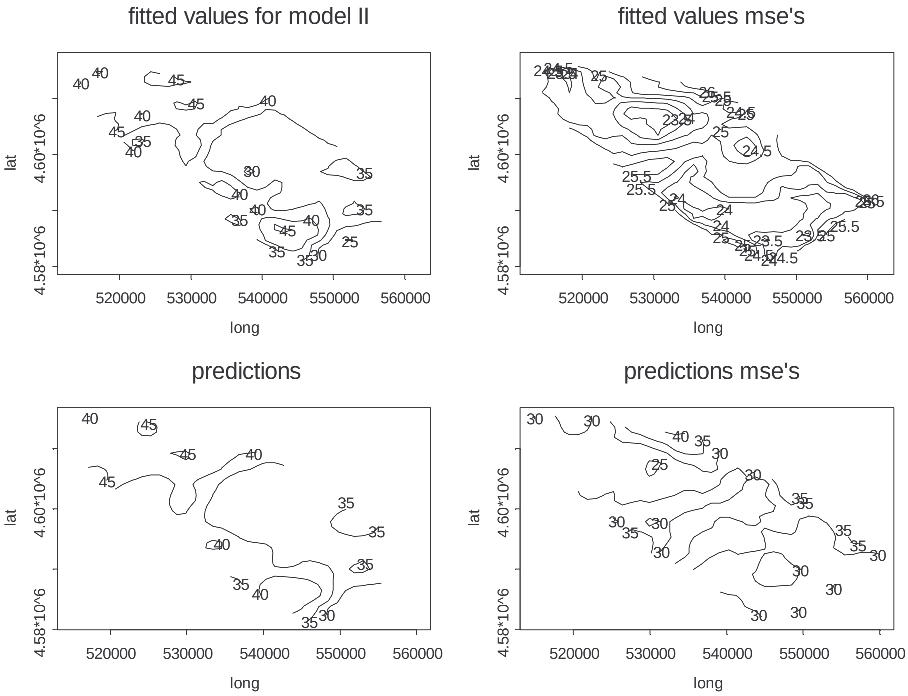
- Assessment of the chosen model VI



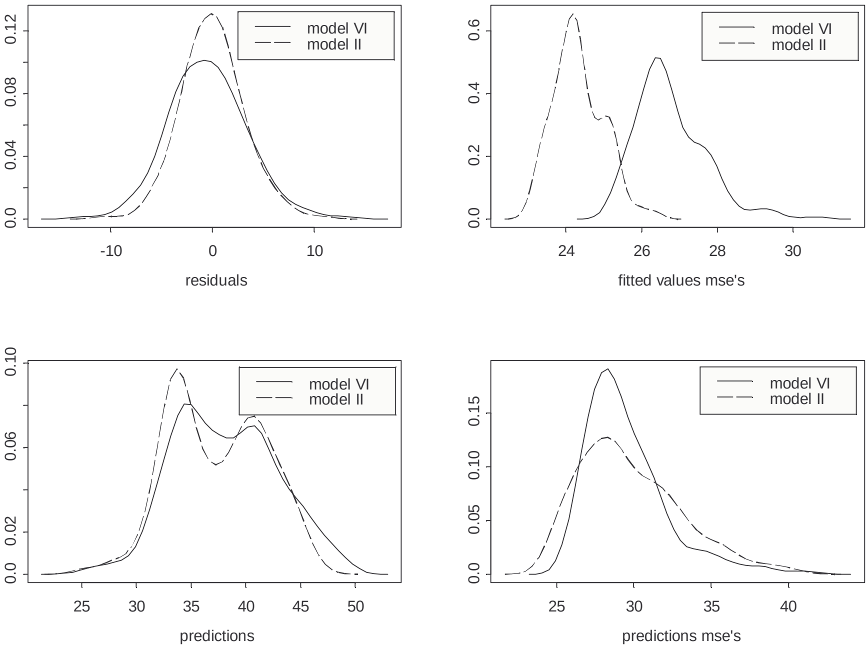
- Second best ”short range” model II



- Second best ”short range” model II



- Marginal comparisons between models VI and II



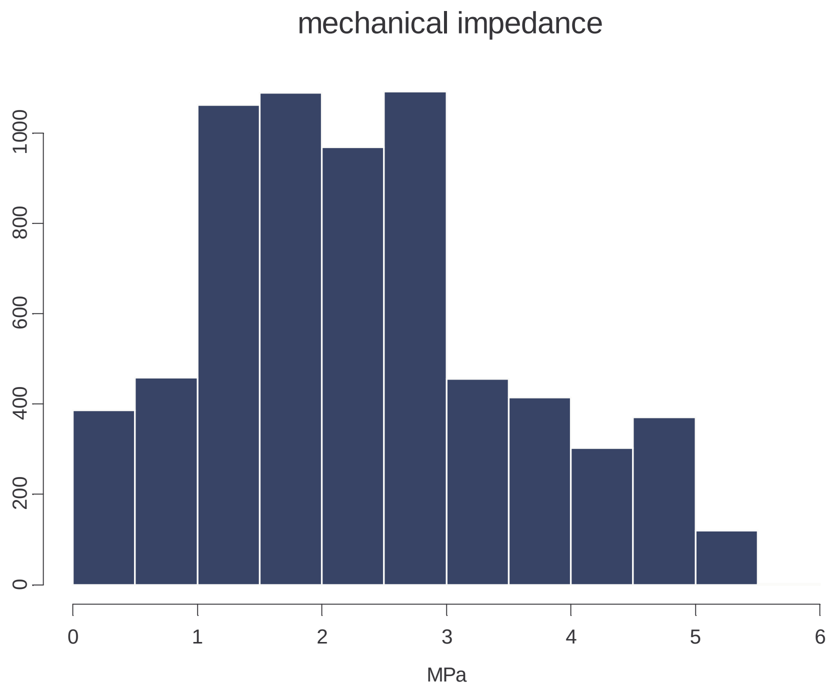
- Bayesian estimates for combinations of two covariance structures, four range prior specificarions and three trend surfaces. Priors for (β, σ) and τ were respectively assumed noninformative improper and discrete uniform (results of 10000 draws from the posterior distribution obtained by the R package geoR, Ribeiro and Diggle, 1999)

	COV	PRI	LON	LAT	NUG	SIL	RAN
I	exp	uni	-0.17	0.23	14.91	27.96	4.26
		rec	-0.18	0.20	14.23	26.69	3.16
		squ	-0.19	0.17	15.24	26.90	2.61
		exp	-0.19	0.17	15.38	27.14	2.44
II	exp	uni	-0.28		14.17	26.57	3.43
		rec	-0.28		15.00	26.47	2.72
		squ	-0.27		15.25	26.92	2.28
		exp	-0.27		15.54	27.42	2.22
III	exp	uni		0.37	15.08	30.16	5.03
		rec		0.36	15.50	29.05	3.87
		squ		0.35	15.24	28.57	3.27
		exp		0.34	15.24	28.57	2.88
IV	sph	uni	-0.18	0.20	19.72	22.76	11.46
		rec	-0.18	0.20	19.58	22.59	9.10
		squ	-0.18	0.19	19.40	22.39	8.82
		exp	-0.22	0.10	19.58	23.49	4.59
V	sph	uni	-0.28		19.83	22.03	8.93
		rec	-0.28		19.65	21.84	8.60
		squ	-0.28		18.96	21.07	8.22
		exp	-0.27		19.51	23.41	4.53
VI	sph	uni		0.40	21.35	27.84	11.52
		rec		0.40	21.29	27.77	11.40
		squ		0.36	19.00	24.79	9.21
		exp		0.20	18.20	23.74	7.45

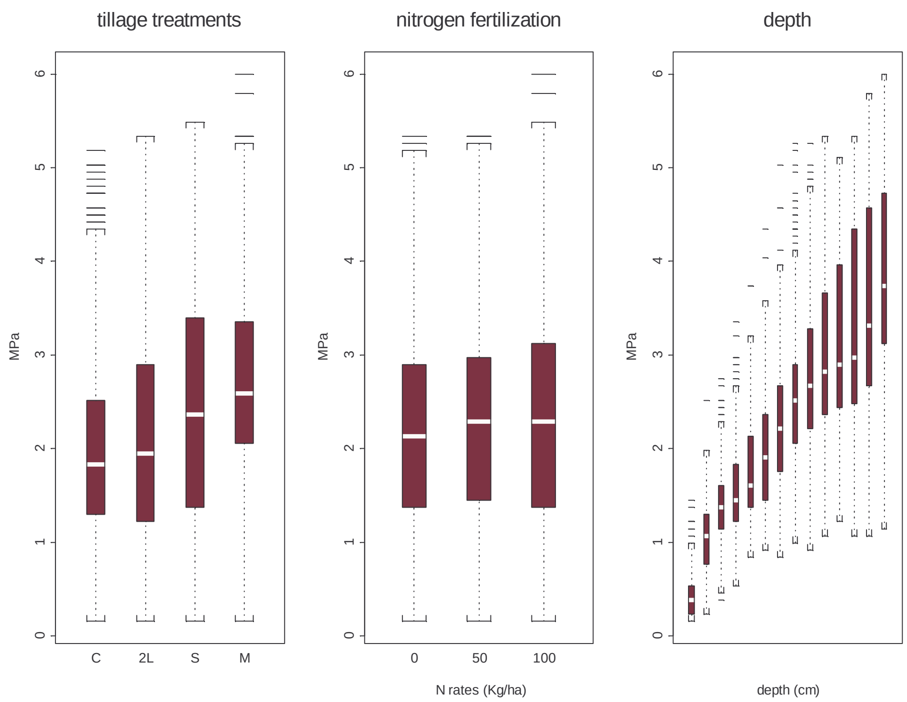
11 Case study: field trial modelling

- Conservation tillage as opposed to soil degradation.
- Mechanical impedance (resistance to penetration as measured by a cone penetrometer) is considered as a measure of the soil physical structure.
- Trial carried out at the farm of "Istituto Sperimentale Agronomico" within a durum wheat growing area in Apulia.
- Four different tillage treatments combined with three nitrogen rates.
- Recordings were taken twice during the 2001 crop season at $3.5cm$ vertical increments to a depth of $52cm$ over a regular grid with $3m \times 5m$ cells.
- Repeated split-plot design with two blocks: tillage treatments randomly assigned to whole plots; fertilization rates randomly assigned to split plots. Water content was measured at each subplot.
- Spatial heterogeneity is common in large field experiments, plot-to-plot variation within a block being affected by soil variation and fertility, and other site specific conditions.
- Mixed effects models provide a sensible method to account for within block spatial variation.

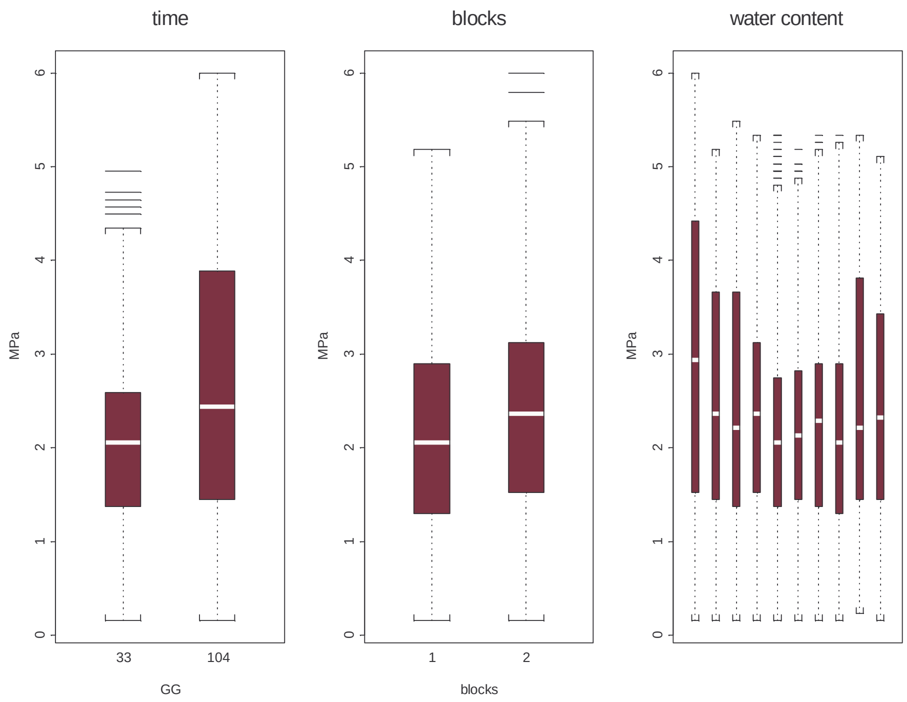
- Marginal distribution of the soil impedance



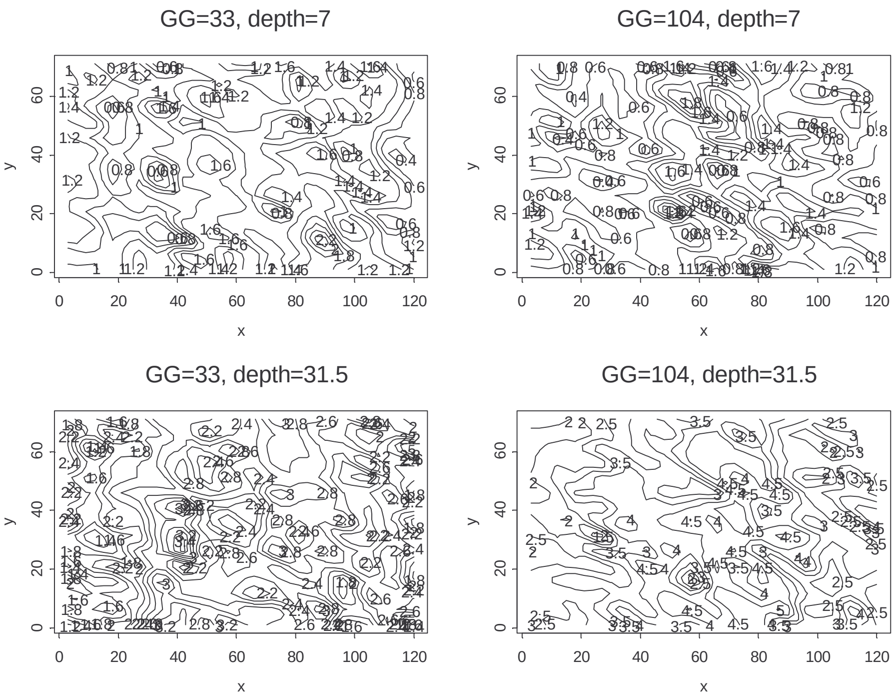
- Interactions of the soil impedance with tillage treatment nitrogen rate and depth



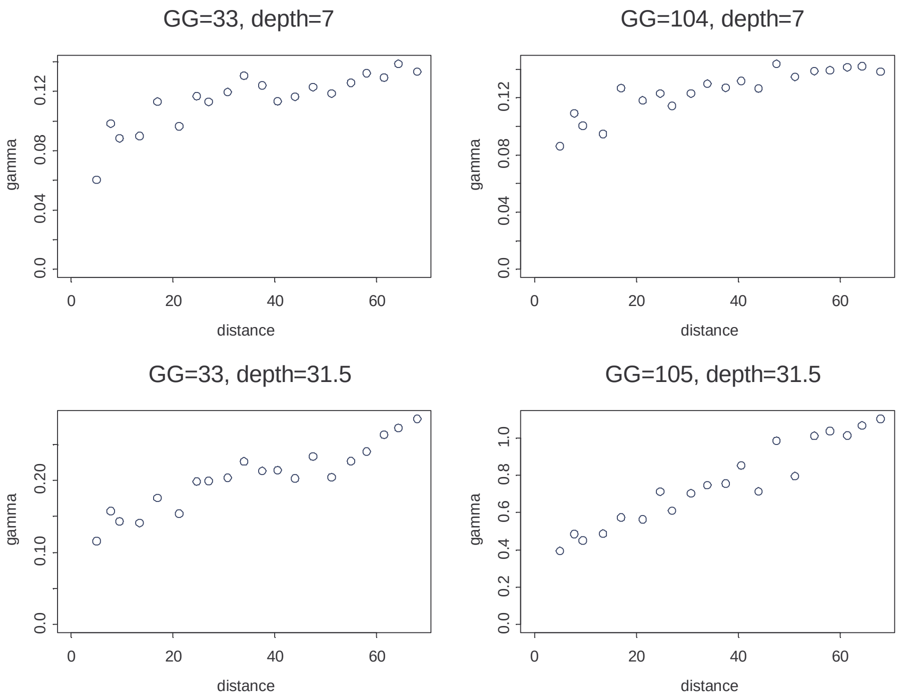
- Interactions with time, blocks and water content



- Contourplots of the soil impedance for two depths and the two time points



- Variograms of the soil impedance for two depths and the two time points



- Fitted model

$$Y = X\beta + S + \varepsilon$$

  * Y soil mechanical impedance
  * X covariates
    - tillage treatment
    - nitrogen fertilization rate
    - tillage-nitrogen interaction
    - water content
    - depth
    - day
    - block
  * S spatially correlated signal
  * ε measurement error term

- Correlation structure comparisons

	exp		sph		gau	
	topsoil	subsoil	topsoil	subsoil	topsoil	subsoil
-2RLL	2768.2	5371.7	2793.5	5369.3	3308.2	5982.2
AIC	2774.2	5377.7	2799.5	5375.3	3314.2	5988.2
BIC	2776.2	5379.7	2801.4	5377.2	3316.1	5990.1

- Exponential correlation structure: covariance parameter estimates

	topsoil		subsoil	
	EST	SE	EST	SE
σ^2	0.065	0.007	0.195	0.038
ϕ	17.360	3.135	40.006	11.167
τ^2	0.093	0.005	0.218	0.009

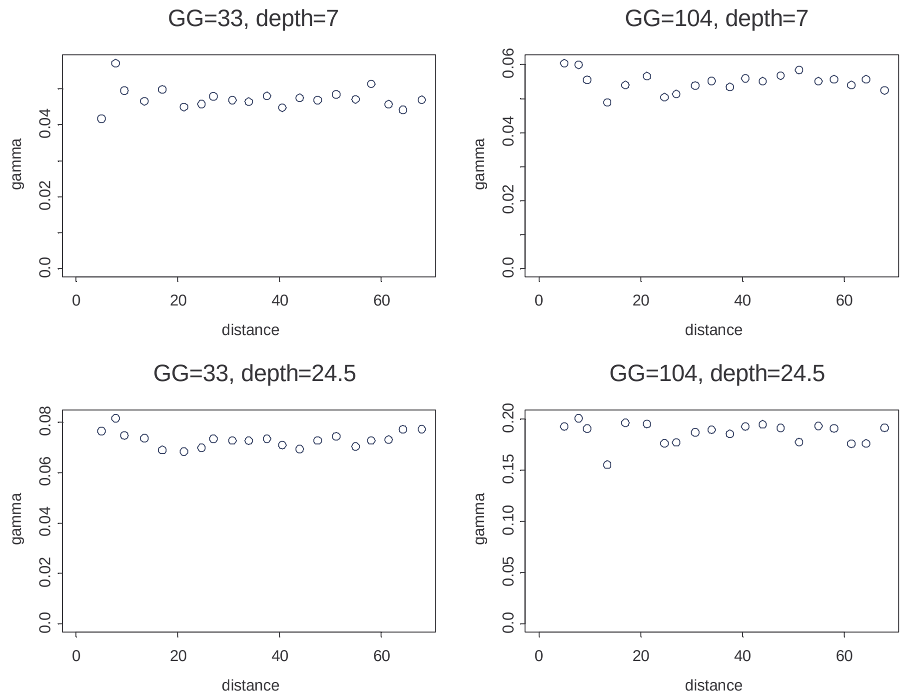
- Significance of tests on fixed effects

	topsoil	subsoil
tillage	$<.0001$	$<.0001$
nitrogen	0.0016	0.0002
tillage*nitrogen	$<.0001$	0.0001
water content	0.0003	0.0663
depth	$<.0001$	0.0128
time	0.2348	0.0001
block	0.0015	0.1805

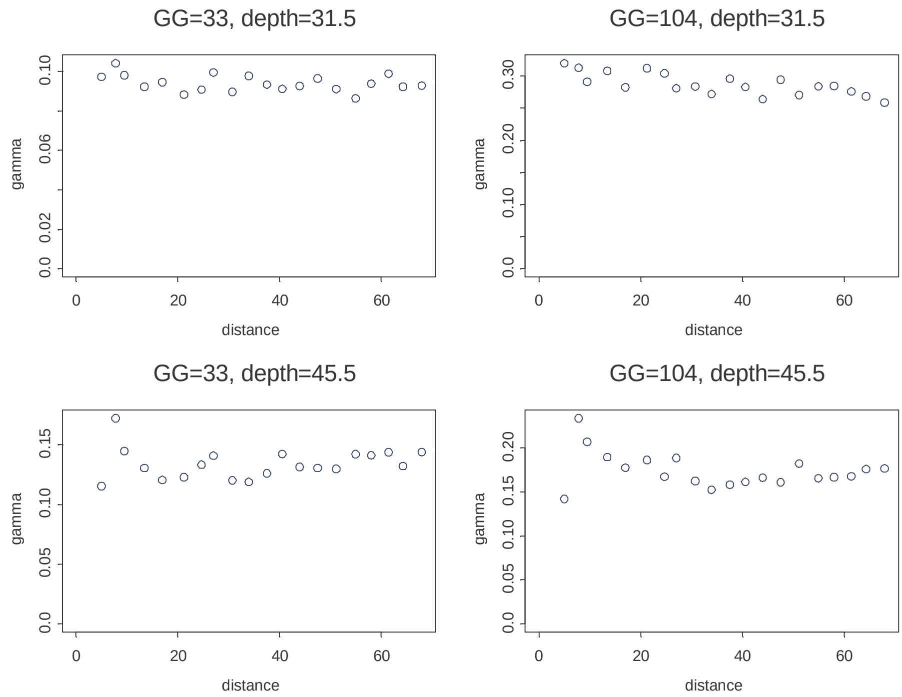
- Least squares means

		topsoil		subsoil	
		EST	SE	EST	SE
tillage	conventional	1.366	0.028	2.616	0.065
	two-layer	1.286	0.028	3.121	0.065
	surface	1.396	0.030	3.390	0.070
	minimum	1.903	0.030	3.239	0.070
nitrogen	0	1.449	0.025	3.026	0.064
	50	1.516	0.025	3.098	0.064
	100	1.499	0.025	3.150	0.064
depth	3.5 $\vee$ 14	0.445	0.060	2.446	0.164
	7 $\vee$ 31.5	1.090	0.060	2.687	0.164
	10.5 $\vee$ 35	1.378	0.060	2.925	0.164
	14 $\vee$ 38.5	1.555	0.060	3.088	0.164
	17.5 $\vee$ 42	1.745	0.060	3.261	0.164
	21 $\vee$ 45.5	1.960	0.060	3.467	0.164
	24.5 $\vee$ 49	2.241	0.060	3.767	0.164
time	33	1.458	0.032	2.562	0.088
	104	1.518	0.032	3.621	0.088
block	1	1.419	0.029	3.048	0.069
	2	1.557	0.029	3.135	0.070

- Variograms of residuals for the topsoil model



- Variograms of residuals for the subsoil model



12 Future directions

- Semiparametric and nonparametric extensions of mixed effects models:
  * Finite mixtures [L. Magder, S. Zeger (1996) JASA];
  * Nonparametric density estimation [Tao et al. (1999)];
  * Dirichlet process prior on random effects [A. Gelfand, A. Kottas (2001)].
- Extensions of the spatial model:

$$Y(\mathbf{t}) = X(\mathbf{t})\beta + S(\mathbf{t}) + \varepsilon(\mathbf{t})$$

  * Generalized linear mixed effects models [P. Diggle, R. Mojeed, J. Tawn (1998), Appl. Stats.];
  * Bivariate linear mixed effets model [S. Banerjee, A. Gelfand, W. Polasek (2000), JSPI];
  * Spatially correlated random effects obtained as a convolution of a white noise process and a smoothing kernel function [D. Higdon (2001), Tech. Rep. Duke Univ.].