Mapping soil properties using a non-stationary variance geostatistical model

Alexandre Wadoux*[1], Dick Brus[2], Gerard Heuvelink[1]

[1]Soil Geography and Landscape group, Wageningen University & Research
[2]Biometris, Wageningen University & Research

*alexandre.wadoux@wur.nl

Introduction

Conventional geostatistical models assume that the property being monitored is the realisation of a second order stationary random process

$$Z(s) = \mu + \sigma\varepsilon(s)$$

μ, σ are constant

$$Cov(\,\varepsilon(s), \varepsilon(s+h)\,) = C(h)$$

if $h = 0, Cov(\varepsilon(s),\ \varepsilon(s)) = Var\big(\varepsilon(s)\big) = C(0)$

But this is often an invalid assumption

- Can be checked with exploratory analysis of the observation

WAGENINGEN UR
For quality of life

Non-stationary variance model

Ordinary kriging

$$Z(s) = \underbrace{\mu + \sigma\varepsilon(s)}_{\text{mean + random error}}$$

- <u>Simple solutions exist for non-stationarity</u>

Universal kriging (non-stationary mean)

$$Z(s) = \underbrace{\mu(s) + \sigma\varepsilon(s)}_{\text{spatial trend + random error}}$$

Universal kriging with non-stationary variance

$$\underbrace{Z(s)}_{\text{target variable}} = \underbrace{\mu(s)}_{\text{spatial trend}} + \underbrace{\sigma(s)}_{\text{spatial standard deviation}} \cdot \underbrace{\varepsilon(s)}_{\text{error}}$$

Non-stationary variance model

Mean at location s

$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$

Standard deviation at location s

Standardised random error

Non-stationary variance model

In matrix notation

$$\boldsymbol{z} = \boldsymbol{F}\boldsymbol{\beta} + \underbrace{\boldsymbol{G}\boldsymbol{\kappa} \cdot \boldsymbol{\varepsilon}}$$

$\boldsymbol{C} = diag\{\boldsymbol{G}\boldsymbol{\kappa}\} \cdot \boldsymbol{R} \cdot diag\{\boldsymbol{G}\boldsymbol{\kappa}\}^T$ is the variance-covariance matrix

($\boldsymbol{R}$ is the correlation matrix)

Predictions at new location

$$\hat{\boldsymbol{z}}(\boldsymbol{s}_0) = \mathbf{f}(\boldsymbol{s}_0)^T\hat{\boldsymbol{\beta}} + \boldsymbol{g}(\boldsymbol{s}_0)^T\hat{\boldsymbol{\kappa}} \cdot \hat{\boldsymbol{\varepsilon}}(\boldsymbol{s}_0)$$

Prediction error variance at new location

$$\sigma^2(\boldsymbol{s}_0) = \underbrace{\boldsymbol{c}(0) - \boldsymbol{c}_0^T\boldsymbol{C}^{-1}\boldsymbol{c}_0}$$

Prediction error variance of the residuals

$$+\underbrace{(\boldsymbol{f}_0 - \boldsymbol{F}^T\boldsymbol{C}^{-1}\boldsymbol{c}_0)^T(\boldsymbol{F}^T\boldsymbol{C}^{-1}\boldsymbol{F})^{-1}\boldsymbol{f}_0 - \boldsymbol{F}^T\boldsymbol{C}^{-1}\boldsymbol{c}_0}$$

Error variance of the trend

Parameter estimation

With exponential correlogram

$$r(h) = c_0 + (1 - c_0)\left\{\exp\left(\frac{-3h}{a}\right)\right\}$$

We need to estimate $\Phi = [\kappa_i, c_0, a]$ and β_i

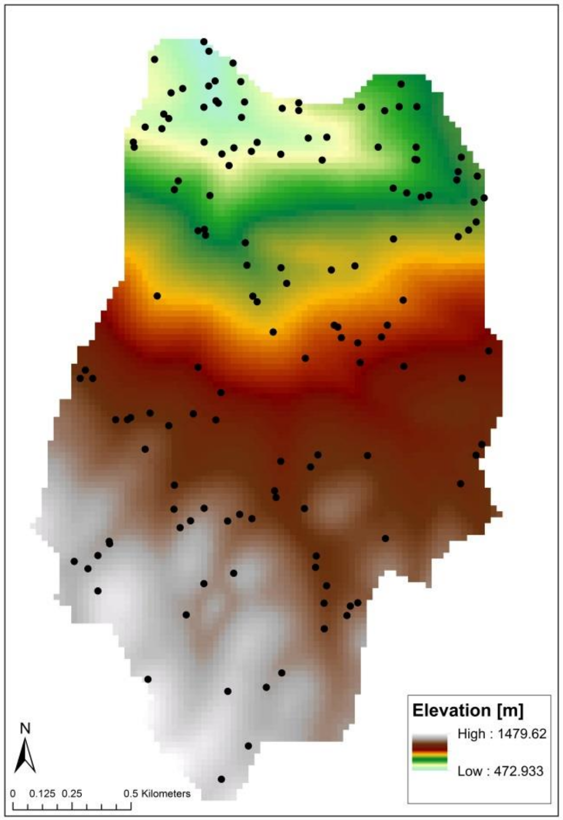
Independent of β_i, Restricted log-likelihood:

$$\ell(\Phi|\boldsymbol{z}) = Constant\ - \frac{1}{2}\ln|\boldsymbol{C}| - \frac{1}{2}\ln|\boldsymbol{X}^T\boldsymbol{C}^{-1}\boldsymbol{X}| - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{C}^{-1}(\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{z}$$

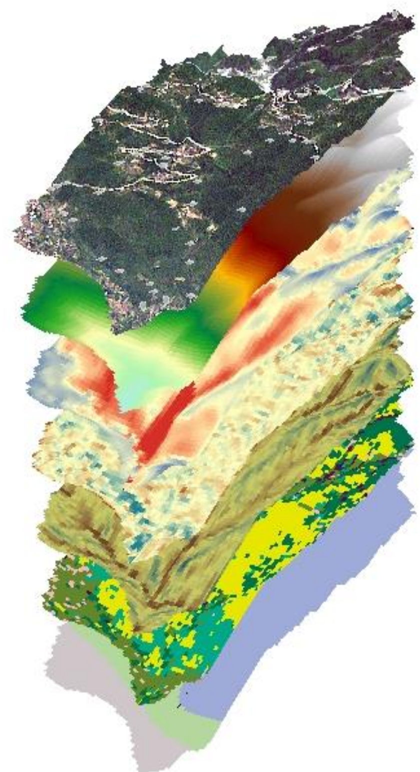
β_i are estimated with GLS using REML estimates of κ_i, c_0, a

Case study in central China

Soil clay content (0-20 cm)

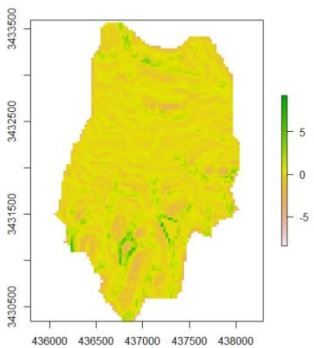


- 140 topsoil samples with clay content analysis (%)
- 13 covariates

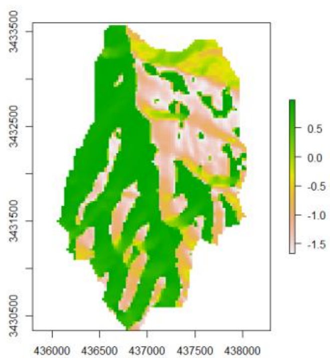


Covariates

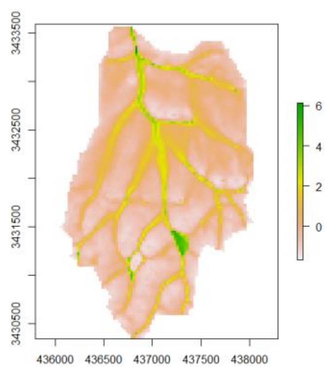
$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$



Profil curvature

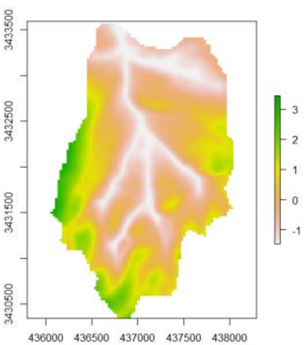


Northing

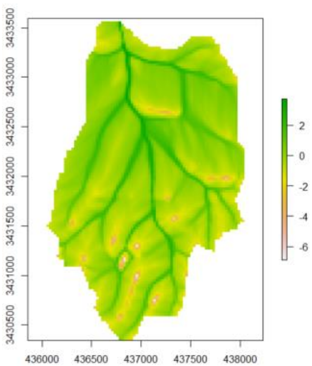


Wetness Index

+

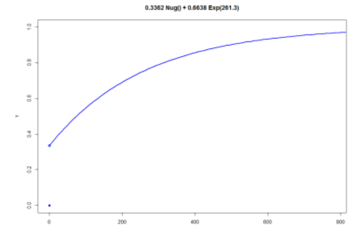


Altitude above channel network



Catchment area

·

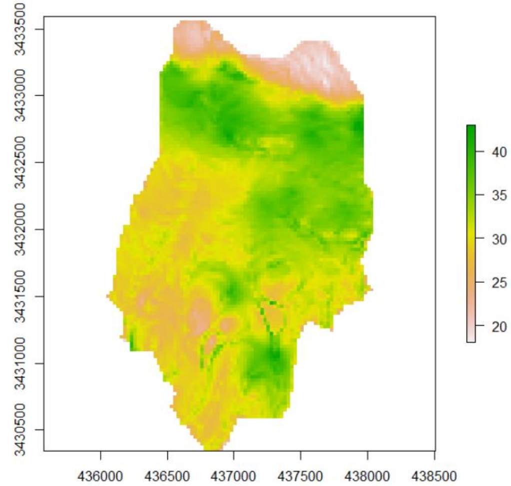


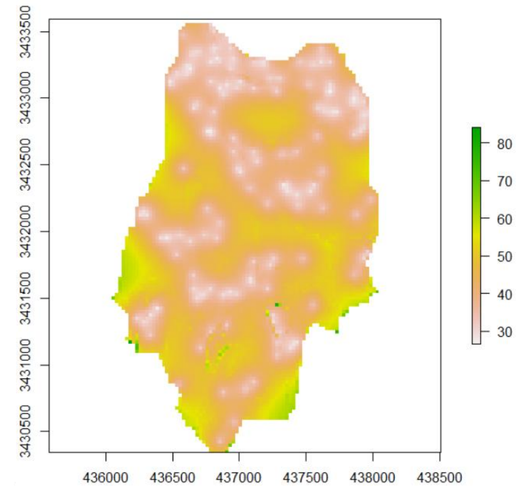
Results

Prediction | **Prediction error variance**

Stationary model

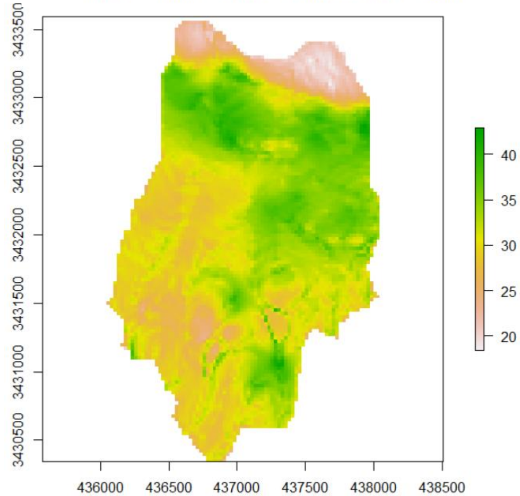
$$Z(s) = \mu(s) + \sigma\varepsilon(s)$$

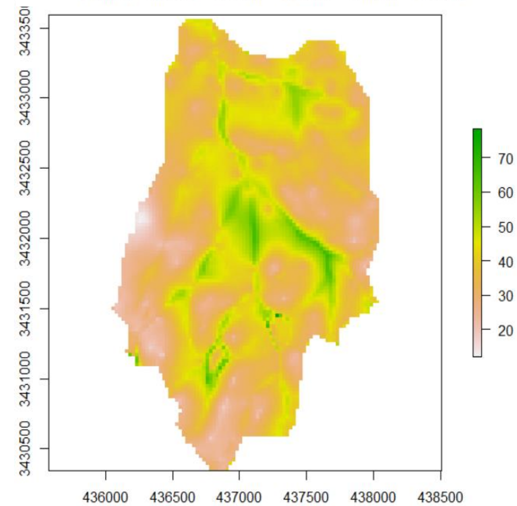




Non-stationary variance model

$$Z(s) = \mu(s) + \sigma(s) \cdot \varepsilon(s)$$





Summary statistics

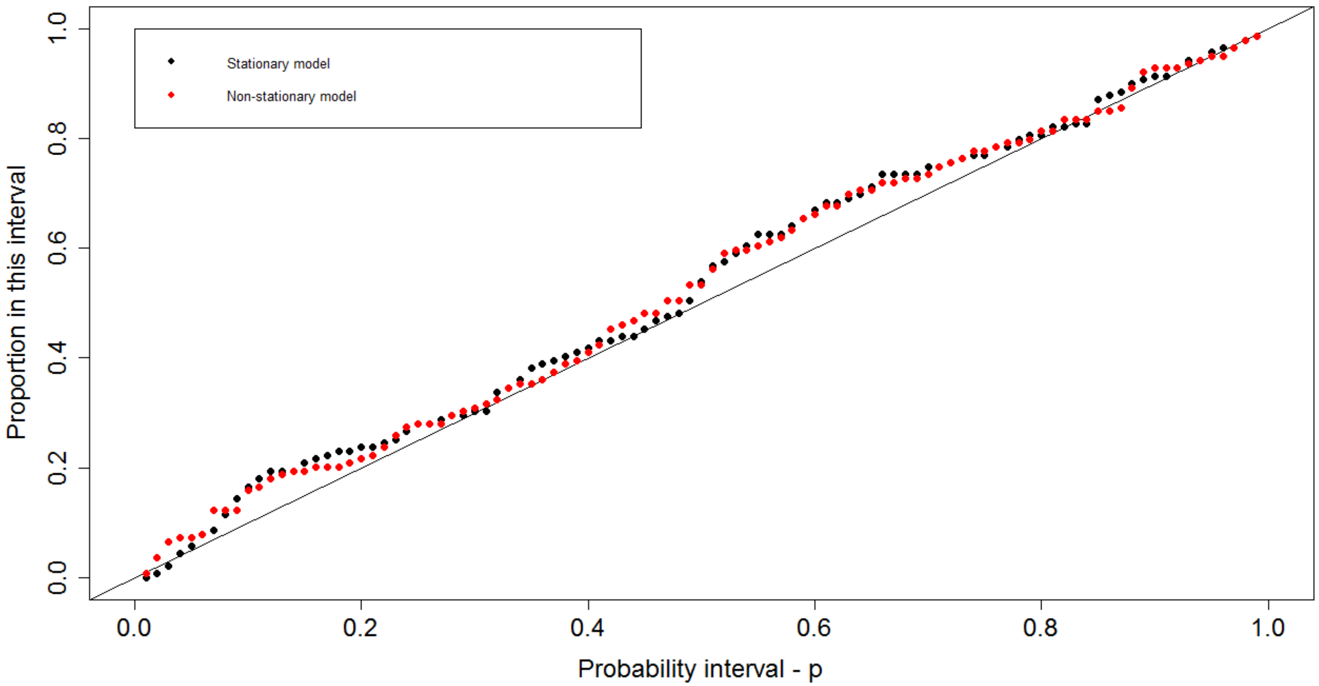
$$MUKV = \int_{s \in A} var\left(Z(s) - \hat{Z}(s)\right) ds$$

$$var(z\ score) = var\left(\frac{\hat{Z}(s) - Z(s)}{\sigma(s)}\right)$$

$$\theta(s) = \frac{\left(Z(s) - \hat{Z}(s)\right)^2}{\sigma^2(s)}$$

	Log-likelihood	AIC	RMSE	R^2	var(z score)	Standardised squared prediction error $\theta(s)$	MUKV
Stationary model	-459.9	933.7	6.045	0.4044	0.9484	0.9416	43.24
Non-stationary variance model	-457.0	932.0	6.035	0.4063	0.9516	0.9448	37.79

Accuracy plot



	Precision (Deutsch, 1997)	Goodness (Goovaert, 2001)
Stationary model	0.9430	0.9694
Non-stationary variance model	0.9435	0.9703

Interested ?

1. Wadoux, A, Brus, DJ, Rico-Ramirez, MA, and Heuvelink, GBM, Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model. 2017; Advances in Water Resources (in Press)

Questions ?

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.