Geoderma 324 (2018) 138–147

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Accounting for non-stationary variance in geostatistical mapping of soil properties

Alexandre M.J-C. Wadoux[a,*], Dick J. Brus[b], Gerard B.M. Heuvelink[a]

[a] Soil Geography and Landscape group, Wageningen University, The Netherlands
[b] Biometris, Wageningen University & Research, The Netherlands

ARTICLE INFO

Keywords:
Geostatistics
Pedometrics
Kriging
Non-stationarity
Uncertainty assessment
REML

ABSTRACT

Simple and ordinary kriging assume a constant mean and variance of the soil variable of interest. This assumption is often implausible because the mean and/or variance are linked to terrain attributes, parent material or other soil forming factors. In kriging with external drift (KED) non-stationarity in the mean is accounted for by modelling it as a linear combination of covariates. In this study, we applied an extension of KED that also accounts for non-stationary variance. Similar to the mean, the variance is modelled as a linear combination of covariates. The set of covariates for the mean may differ from the set for the variance. The best combinations of covariates for the mean and variance are selected using Akaike's information criterion. Model parameters of the selected model are then estimated by differential evolution using the Restricted Maximum Likelihood (REML) in the objective function. The methodology was tested in a small area of the Hunter Valley, NSW Australia, where samples from a fine grid with gamma K measurements were treated as measurements of the variable of interest. Terrain attributes were used as covariates. Both a non-stationary variance and a stationary variance model were calibrated. The mean squared prediction errors of the two models were somewhat comparable. However, the uncertainty about the predictions was much better quantified by the non-stationary variance model, as indicated by the mean and median of the standardized squared prediction error and by accuracy plots. We conclude that the non-stationary variance model is more flexible and better suited for uncertainty quantification of a mapped soil property. However, parameter estimation of the non-stationary variance model requires more attention due to possible singularity of the covariance matrix.

1. Introduction

Standard geostatistical mapping approaches predict a soil variable of interest at the unsampled nodes of a fine grid using measurements of this variable at sampling locations. In many cases predictions can be improved by exploiting a relation between the soil variable and one or more environmental covariates of which maps are available, such as terrain attributes derived from a digital elevation model and remote sensing images. This is usually done by modelling the soil variable as the sum of a linear combination of covariates and a spatially autocorrelated residual. This leads to Kriging with External Drift (KED) (Goovaerts, 1997). In situations where the covariates explain a considerable part of the variation of the soil variable, KED is superior to simple or ordinary kriging that both assume that the mean of the soil variable is constant within a global or local neighbourhood and not dependent on covariates.

In KED we allow for a non-stationary mean, but the variance is assumed stationary (i.e., constant). More specifically, it is assumed that the covariance between the soil variable Z at two locations $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ only depends on the separation vector $\mathbf{h}$: $Cov(Z(\mathbf{s}),Z(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$. Taking $\mathbf{h} = \mathbf{0}$ shows that the variance is assumed constant: $Var(Z(\mathbf{s})) = C(\mathbf{0})$ for all $\mathbf{s}$. However, in many cases the assumption of a stationary variance may be implausible, i.e. when the residual spatial variation is substantially different in different parts of the study area. For instance, McBratney and Webster (1981) identified several discontinuities in the variograms of soil colour and pH along a transect in north-east Scotland. The authors attributed the changes to boundaries between soil types. Similarly, Voltz and Webster (1990) found important differences between topsoil clay content variograms of contrasting Jurassic sediments.

In some cases, non-stationarity in the variance can be solved by transforming the data prior to geostatistical modelling, e.g. by a square-root or log-transformation (e.g., Jacques et al., 1999). Several solutions have been proposed in case a transformation does not solve the problem. Pintore and Holmes (2004) and later Haskard and Lark (2009) proposed to account for a non-stationary variance by spectral

* Corresponding author at: Soil Geography and Landscape group, Wageningen University, Droevendaalsesteeg 3, Wageningen 6708 BP, The Netherlands.
E-mail address: alexandre.wadoux@wur.nl (A. M.J-C. Wadoux).

https://doi.org/10.1016/j.geoderma.2018.03.010
Received 17 November 2017; Received in revised form 21 February 2018; Accepted 12 March 2018
Available online 30 March 2018

tempering. The method tempers a spectrum based on a stationary correlation matrix, but the modelled covariance structure can vary spatially while maintaining positive-definiteness. The authors showed that modifying the spectrum of the data according to a covariate on a transect gave a more realistic variance model for a case study on rates of emission of nitrous oxide from soils. Alternatively, McBratney and Minasny (2013) proposed to equalize variogram parameters by deformation of the geographic space. This method renders a stationary covariance function in the transformed space. Spatial predictions made in the transformed space are then back-transformed to the original geographic space. However, while this approach addresses differences in spatial correlation, it does not solve the non-stationary variance problem.

The work presented here builds on the work of Lark (2009) and Marchant et al. (2009). They demonstrated how a model in which the variance is a function of the spatial coordinate or covariates can be fitted by REML, and how such model can be used in geostatistical prediction of soil properties. The same approach is applied by Brus et al. (2016) in three-dimensional soil property mapping. They assumed that the residual variance is a stepwise or continuous function of depth, while in the horizontal plane, at a given depth, the residual variance was assumed constant.

The objective of this study is to test the approach proposed by Lark (2009) in a case study where several covariates are available for modelling the non-stationarity of the mean and variance. The best stationary variance model is compared with the best non-stationary variance model, using evaluation criteria that measure both the quality of the predictions as well as the quality of the estimated prediction uncertainty.

2. Statistical methodology

2.1. Model definition

A soil variable of interest Z at any location $\mathbf{s}$ in the study area $\mathscr{A}$ is modelled by:

$$Z(\mathbf{s}) = m(\mathbf{s}) + \sigma(\mathbf{s})\varepsilon(\mathbf{s}) \tag{1}$$

where $m(\mathbf{s})$ is the mean at location $\mathbf{s}$, $\sigma(\mathbf{s})$ the standard deviation at location $\mathbf{s}$ and ε a stationary, spatially correlated Gaussian random field with zero mean and unit variance. The mean m and standard deviation σ are deterministic functions that are modelled as linear combinations of covariates, unconditional on the observations:

$$Z(\mathbf{s}) = \sum_{k=0}^{K} \beta_k w_k(\mathbf{s}) + \sum_{l=0}^{L} \kappa_l g_l(\mathbf{s})\varepsilon(\mathbf{s}) \tag{2}$$

where the β_k and κ_l are regression coefficients (the latter are used for modelling the standard deviation), and the w_k and g_l spatially distributed covariates. We take $w_0(\mathbf{s}) = g_0(\mathbf{s}) = 1$ for all $\mathbf{s}$, so that β_0 and κ_0 are space-invariant constant contributions to the mean and standard deviation, respectively.

Let Z be measured at n locations $\mathbf{s}_i (i = 1, ..., n; \mathbf{s}_i \in \mathscr{A})$. The measurements $z(\mathbf{s}_i)$ are treated as realizations of the Gaussian field Z and prediction is done for Z at a new, unobserved location $\mathbf{s}_0$. Stacking the $z(\mathbf{s}_i)$ in a (column) vector $\mathbf{z}$ and changing to matrix notation yields:

$$\mathbf{z} = \mathbf{W}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{W}$ is the $n \times (K + 1)$ design matrix of covariates for the mean at the observation locations, $\boldsymbol{\beta}$ is the $(K + 1)$ vector of regression coefficients for the mean and $\boldsymbol{\varepsilon}$ is the n-vector of (standardized) residuals, which has variance-covariance matrix $\mathbf{R}$. $\mathbf{H}$ is an $n \times n$ diagonal matrix defined by:

$$\mathbf{H} = diag\{\mathbf{G}\boldsymbol{\kappa}\} \tag{4}$$

where $\mathbf{G}$ is the $n \times (L + 1)$ matrix of standard deviation covariates at observation locations and $\boldsymbol{\kappa}$ is an $(L + 1)$ vector of standard deviation regression coefficients. Note that while ε has variance-covariance matrix $\mathbf{R}$, the stochastic component $\mathbf{H}\boldsymbol{\varepsilon}$ of Eq. (3) has variance-covariance matrix $\mathbf{C} = \mathbf{HRH}'$. The parameters of the model defined by Eq. (3) are $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$ and the parameters of a model for the spatial autocorrelation of the standardized residual. In this work we will parametrize the spatial autocorrelation by an isotropic exponential correlogram $r(h) = r_0\left\{\exp\left(-\frac{h}{a}\right)\right\}$ (where $h > 0$ is the Euclidean distance between two locations, by definition $r(0) = 1$), thus introducing two more parameters, namely r_0 and a. Parameter r_0 equals one minus the nugget-to-sill ratio, while a refers to the spatial correlation length (or range, $3a$ being the effective range). Note that the stationary variance model is a special case of the non-stationary variance model. It is obtained by setting parameters κ_l, $l = 1...L$ to zero, so that $\sigma(s) = \kappa_0$ for all $\mathbf{s}$.

2.2. Parameter estimation and model selection

2.2.1. Parameter estimation

In estimation the parameters are subdivided in two subsets, the regression coefficients $\boldsymbol{\beta}$ for the mean, and all parameters of the stochastic part of the model, $\Phi = [\boldsymbol{\kappa}, r_0, a]$. For a stationary variance model the second subset reduces to $\Phi = [\kappa_0, r_0, a]$. The standard maximum likelihood estimates of Φ depend non-linearly on the regression coefficients for the mean $\boldsymbol{\beta}$, which introduces a bias in the estimates of Φ if both parameter subsets are estimated jointly (Lark and Webster, 2006). This problem can be avoided by restricted (or residual) maximum likelihood (REML) parameter estimation. REML first estimates Φ and next $\boldsymbol{\beta}$. Similar to standard maximum likelihood estimation, REML aims to find the vector Φ for which the observed data yield the highest probability density (i.e., likelihood, if treated as a function of the parameters instead of the data). The problem is that the likelihood of Φ depends on the regression coefficients for the mean, which are unknown and must also be estimated. Patterson and Thompson (1971) solved this problem by detrending the data by multiplying the data vector by a projection matrix. The new variable is a function of the original variable but independent of the regression coefficients for the mean. The associated restricted log-likelihood function is given by (Webster and Oliver, 2007):

$$L_r(\Phi|\mathbf{z}) = constant - \frac{1}{2}\log|\mathbf{C}| - \frac{1}{2}\log|\mathbf{W}'\mathbf{C}^{-1}\mathbf{W}| - \frac{1}{2}\mathbf{z}'\mathbf{P}'\mathbf{C}^{-1}(\mathbf{I} - \mathbf{Q})\mathbf{z} \tag{5}$$

where $\mathbf{I}$ is an identity matrix and $\mathbf{P}$ and $\mathbf{Q}$ are defined as:

$$\mathbf{P} = \mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' \tag{6}$$

$$\mathbf{Q} = \mathbf{W}(\mathbf{W}'\mathbf{C}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{C}^{-1} \tag{7}$$

After estimating Φ by maximising the restricted log-likelihood given in Eq. (5) above, the regression coefficients $\boldsymbol{\beta}$ for the mean can be estimated by Generalized Least Squares (GLS):

$$\boldsymbol{\beta} = (\mathbf{W}'\mathbf{C}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{C}^{-1}\mathbf{z} \tag{8}$$

Here, matrix $\mathbf{C}$ is computed from the optimized values for Φ. Note that the regression coefficients κ_l in Eq. (2) can be positive or negative, as long as the covariance matrix $\mathbf{C}$ is not singular.

2.2.2. Model selection

Two subsets of covariates must be chosen, one for the mean and one for the standard deviation. Suppose we have in total K candidate covariates for modelling the mean. For a subset of covariates of size k, there are $\binom{K}{k}$ possible combinations. Since the size is not fixed, we have $\sum_{k=0}^{K}\binom{K}{k}$ possible models in total for the stationary variance model.

Using the same set of candidate covariates for the standard deviation, for the non-stationary variance model the total number of models equals $\left(\sum_{k=0}^{K}\binom{K}{k}\right)^2$. Ideally, all model combinations are fitted and compared. Models with different sets of covariates for the mean cannot be compared on the basis of the restricted log-likelihood, because this is a function of just the covariance parameters Φ. Models can better be compared on the basis of the standard log-likelihood, using a likelihood ratio test or by comparing quality measures that are functions of the log-likelihood and the number of model parameters. Common quality measures are the Akaike Information Criterion (AIC, Akaike, 1998) and Bayesian Information Criterion (BIC, Kass and Wasserman, 1995). AIC, used in the case study hereafter, is defined as:

$$AIC = 2p - 2\log L \tag{9}$$

where p is the number of estimated parameters and $L = p(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\Phi})$ is the ordinary log-likelihood, given by (Diggle and Ribeiro, 2007, Eq. 5.13):

$$L(\boldsymbol{\beta},\ \boldsymbol{\Phi}|\mathbf{z}) = -\tfrac{1}{2}n\log(2\pi) - \tfrac{1}{2}\log|\mathbf{C}| - \tfrac{1}{2}(\mathbf{z}-\mathbf{W}\boldsymbol{\beta})'\mathbf{C}^{-1}(\mathbf{z}-\mathbf{W}\boldsymbol{\beta}) \tag{10}$$

The number of covariate combinations for the non-stationary variance model will be very large, unless the number of covariates K is small. Often, one will therefore resort to numerical search algorithms (e.g., greedy algorithms) to overcome large computation times. In the case study presented in Section 3, the small number of covariates K makes exhaustive search possible. Note that models are compared based on the AIC using the ordinary log-likelihood, while prediction is made with parameters estimated by restricted log-likelihood (Hoeting et al., 2006).

2.3. Kriging

Ignoring estimation errors in $\boldsymbol{\Phi}$ allows us to use a standard result from universal kriging (Webster and Oliver, 2007) to predict the soil variable of interest at a new, unobserved locations $\mathbf{s}_0$:

$$\hat{z}(\mathbf{s}_0) = (\boldsymbol{c}_0 + \mathbf{W}(\mathbf{W}'\mathbf{C}^{-1}\mathbf{W})^{-1}(\boldsymbol{w}_0 - \mathbf{W}'\mathbf{C}^{-1}\boldsymbol{c}_0))'\mathbf{C}^{-1}\mathbf{z} \tag{11}$$

where $\boldsymbol{w}_0$ is a vector of covariates for the mean at the prediction location and $\boldsymbol{c}_0$ is an n vector of covariances between the residuals at the observation and prediction location. Note that these are covariances of the (unstandardized) residuals $\sigma\cdot\varepsilon$ and thus depend on the standard deviation covariates $g_l(\mathbf{s}_0)$, their associated (estimated) regression coefficients κ_l and the correlogram of ε.

The associated prediction error is given by (Webster and Oliver, 2007):

$$Var(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)) = (\mathbf{g}(\mathbf{s}_0)'\boldsymbol{\kappa})^2 - \boldsymbol{c}_0'\mathbf{C}^{-1}\boldsymbol{c}_0 + (\boldsymbol{w}_0 - \mathbf{W}'\mathbf{C}^{-1}\boldsymbol{c}_0)'(\mathbf{W}'\mathbf{C}^{-1}\mathbf{W})^{-1}(\boldsymbol{w}_0 - \mathbf{W}'\mathbf{C}^{-1}\boldsymbol{c}_0) \tag{12}$$

where $(\mathbf{g}(\mathbf{s}_0)'\boldsymbol{\kappa})^2$ is the variance of $Z(\mathbf{s}_0)$. The first two terms on the right-hand side of Eq. (12) quantify the prediction error variance of the residuals, while the last term is the variance of the estimation error of the mean. Note that here we take uncertainty about the β_k into account, whereas uncertainty about the κ_l and variogram parameters r_0 and a is ignored. Taking the latter uncertainties into account is beyond the scope of this work.

2.4. Quality of prediction and estimated uncertainty

Let there be $(N - n)$ validation locations $\mathbf{s}_i$, $i = (n + 1)\ldots N$. In the case study discussed in Section 3, N is the number of nodes of a fine grid covering the study area, of which n are used for model calibration and $(N - n)$ for model evaluation. To quantify the quality of the predictions we computed the mean prediction error (ME), root mean squared prediction error (RMSE) and modelling efficiency coefficient (MEC). The latter is derived as follows (Janssen and Heuberger, 1995):

$$MEC = 1 - \frac{\sum_{i=n+1}^{N}(z(\mathbf{s}_i) - \hat{z}(\mathbf{s}_i))^2}{\sum_{i=n+1}^{N}(z(\mathbf{s}_i) - \bar{z})^2} \tag{13}$$

where $\bar{z}$ denotes the mean of the observations. MEC quantifies the improvement of the model made over using the mean of the observations as a predictor. MEC can become negative, while its optimal value is one.

To evaluate the quality of the prediction error variance, we used the standardized squared prediction error θ (Lark, 2000; Marchant et al., 2009):

$$\theta(\mathbf{s}_i) = \frac{(z(\mathbf{s}_i) - \hat{z}(\mathbf{s}_i))^2}{Var(Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i))}, \quad i = (n+1)\ldots N \tag{14}$$

For a model giving unbiased predictions and correct estimates of the prediction error variance, θ has a χ^2-distribution with one degree of freedom, so that the average value of θ over all validation locations should be close to 1, while its median should be close to 0.455.

2.4.1. Accuracy plots

Deutsch (1997) and Goovaerts (2001) proposed a visual assessment of the quality of the estimated prediction uncertainty through a so-called accuracy plot. Since the prediction error at each validation location s_i is normally distributed with zero mean and known variance (i.e., the kriging variance given by Eq. (12)), its cumulative distribution F_i is known too. From this one can easily compute for a given probability p a symmetric interval around the predicted value through computing the (1-p)/2 and (1 +p)/2 quantiles, and use these as the lower and upper bounds of a prediction interval. This is done for a series of values for p. The proportion of validation locations at which the p prediction interval includes the observed value is then obtained by:

$$\bar{\xi}(p) = \frac{1}{(N-n)}\sum_{i=n+1}^{N}\xi(\mathbf{s}_i; p) \quad \forall\ p \in [0, 1] \tag{15}$$

where $\xi(\mathbf{s}_i; p)$ is given by:

$$\xi(\mathbf{s}_i; p) = \begin{cases} 1 & \text{if } F_i^{-1}(\mathbf{s}_i; (1-p)/2) < z(\mathbf{s}_i) \le F_i^{-1}(\mathbf{s}_i; (1+p)/2) \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

A correct modelling of the uncertainty would entail that the proportion of validation locations where the p prediction interval covers the observed value approximately equals the nominal value p, for all values of p. These proportions are plotted in a scattergram against p; such scattergram is referred to as an accuracy plot. Ideally, the points in the accuracy plot are on the 1:1 line. The absolute deviation from this line can be summarized by an integral, calculated as:

$$A = \int_0^1 |\bar{\xi}(p) - p| dp \tag{17}$$

Ideally, $A = 0$ for a model that perfectly describes the uncertainty. Note that the measure A does not separate over- and underestimation of the uncertainty. Therefore we also derive P_O, the proportion of A that is above the 1:1 line (overestimation of uncertainty), given by:

$$P_O = \frac{1}{A}\int_0^1 \max\{0, \bar{\xi}(p) - p\} dp \tag{18}$$

and P_U, the proportion of A below the 1:1 line (underestimation of uncertainty):

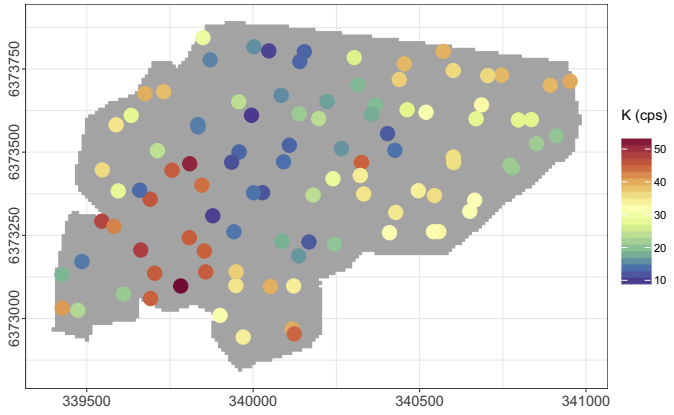

Fig. 1. Scarborough study area in the Hunter Valley, Australia, with location and values of gamma-radiomatric potassium at 100 locations.

$$P_U = \frac{1}{A}\int_0^1 \max\{0, p - \bar{\xi}(p)\}dp \tag{19}$$

Note that P_O and P_U sum to 1.

3. Case study

3.1. Study area and data

We tested the methodology in the 140 ha Scarborough area (Fig. 1), located in the Hunter Valley, Australia. Elevation ranges from 86 to 144 m above sea level with an average of 113 m. Radiometric gamma was surveyed using a vehicle-born passive gamma spectrometer. This yielded a raster file of 10 m × 10 m resolution of gamma-radiometric potassium (K) expressed in cps. The data collection procedure is detailed in Stockmann et al. (2012).

We used the R package spcosa (Walvoort et al., 2010) to divide the area into 50 compact geostrata of equal size, from which we selected two locations per stratum: one in the centre and one randomly. This resulted in a total of 100 sampling locations for the whole area (out of 10,473 grid locations), shown in Fig. 1.

In addition to gamma K observations at the 100 sampling locations, three covariates were derived from the 10 m resolution Digital Elevation Model (DEM):

- Topographic Wetness Index (TWI) (Fig. 2 (b)), which is the steady state wetness index, based on Moore et al. (1993).
- Slope (Fig. 2 (c)), which is the angle of inclination of the soil surface from the horizontal, derived using a 3 × 3 window and the method of Zevenbergen and Thorne (1987).
- Combined curvature (Fig. 2 (d)), which is a combination of profile and planform curvature, based on Moore et al. (1993).

3.2. Practical implementation

3.2.1. Parameter estimation and model selection

All covariates were centred on 0 and scaled to a standard deviation of 1, to enable direct comparison of the associated regression coefficients. The four candidate covariates chosen in Section 2.2.2 were used to compute the best stationary variance model and the best non-stationary variance model using the AIC criterion. Since there were only four candidate covariates the total number of models to be compared

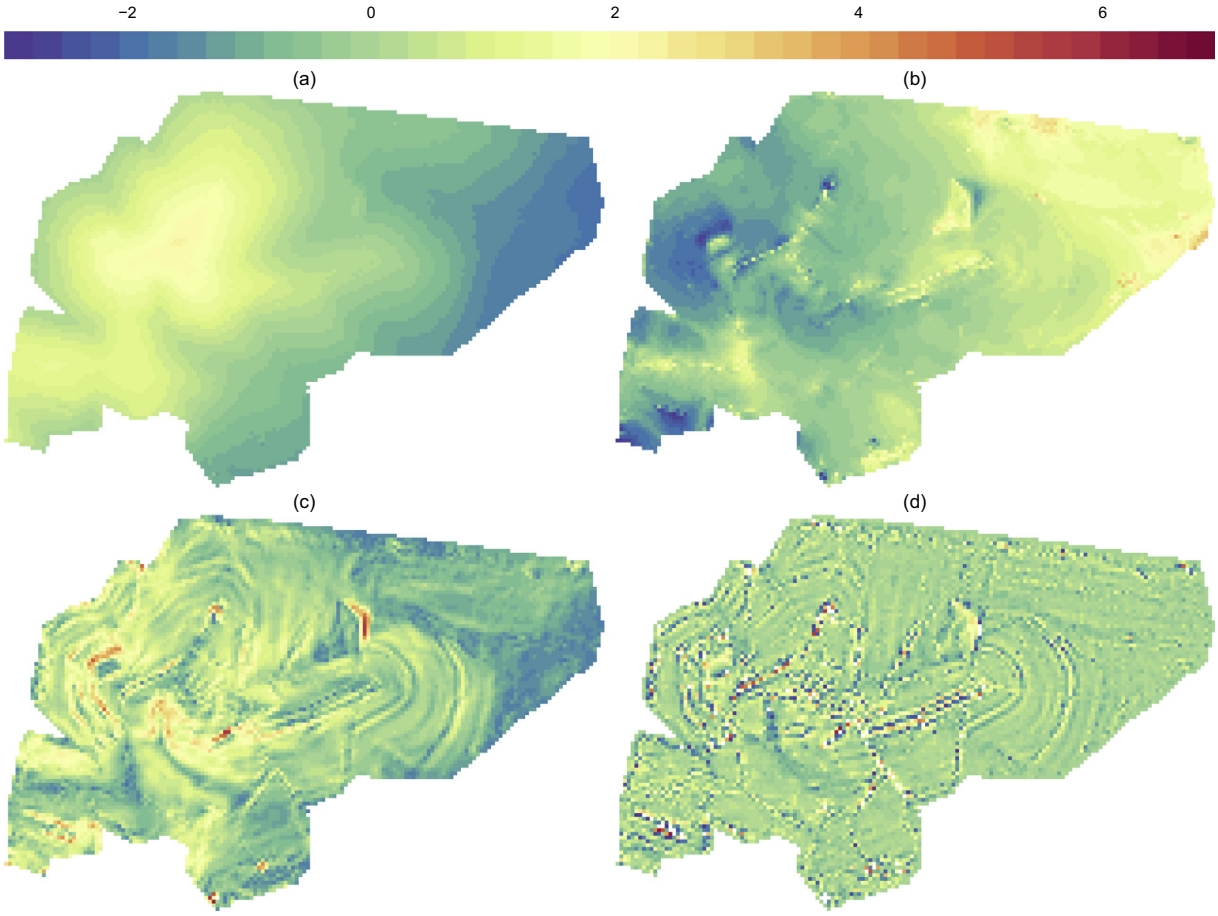

Fig. 2. Standardized covariates used in model selection: (a) DEM, (b) Topographic Wetness Index (TWI), (c) slope and (d) combined curvature.

Table 1
Estimated coefficients for the mean and standard deviation and variogram parameters for the stationary variance and non-stationary variance models.

Parameter	Associated with	Estimated value
Stationary variance model		
β_0 (cps)	Intercept for the mean	29.948
β_1 (m)	Elevation	−1.688
κ_0 (−)	Standard deviation	11.98
r_0 (−)	Short-distance correlation parameter	1.000
a (m)	Range parameter	149.2
Non-stationary variance model		
β_0 (cps)	Intercept for the mean	27.274
β_1 (%)	TWI	4.583
κ_0 (−)	Intercept for the standard deviation	18.34
κ_1 (m)	Elevation	7.115
κ_2 (%)	Slope	−2.161
r_0 (−)	Short-distance correlation parameter	0.999
a (m)	Range parameter	544.0

was only 16 for the stationary variance model and 256 for the non-stationary variance model, allowing exhaustive search. The global optimum of the restricted log-likelihood function was found using differential evolution (Storn and Price, 1997), implemented in the R package DEoptim (Ardia et al., 2015). The convergence threshold was fixed at 10^{-10}. Calculations were done in parallel on a standard desktop with eight cores. REML estimation of the model parameters for all combinations of covariates (272 models in total) took approximately 35 h. The standard deviation parameters were bounded to large positive and negative values to speed up computation (intercept between −50 and 50 and coefficients between −100 and 100). Likewise, variogram parameters were constrained within plausible ranges (the short distance correlation parameter r_0 was forced between 0 and 1 and the correlation length parameter a between 0 and one-third of the extent of the study area). In addition, any proposal combination of parameters in DEoptim that resulted in a near-singular or singular C matrix was rejected. This problem is discussed more extensively in the discussion.

3.2.2. Kriging

Predictions were made with global point kriging on a 10 m × 10 m resolution grid, excluding the 100 observation locations. As far as we know there are no existing R packages for kriging with non-stationary variance, so we implemented this in our own R script. The R script and a test case are available in Sawicka et al. (2017). The prediction to the 10,373 remaining grid cell centres ($N - n$) took less than a minute on a standard desktop computer.

3.3. Results

Based on the procedure detailed in Section 2.2.2, elevation was chosen as a covariate for the mean in the stationary variance model. For the non-stationary variance model, TWI was chosen for the mean while elevation and slope were chosen as covariates for the standard deviation. Table 1 presents the estimated coefficients for the mean, standard deviation and variogram parameters for the two models. Recall that the covariates are standardized to allow comparison between regression coefficients. Elevation has a negative effect on the mean gamma K for the stationary variance model. The stationary variance model also indicates strong residual spatial correlation at short distances (r_0 equals one), but the correlation decreases rapidly with distance (range parameter of 149.2 m for an area with an extent of about 1500 m in the East-West direction). The coefficient for the mean of the non-stationary variance model exhibits a strong positive relationship of gamma K with the wetness index (β_1 of about 4.6). The sign of the standard deviation coefficients κ_1 and κ_2 show that their associated covariates are positively and negatively correlated with the unconditional standard deviation, respectively. Elevation has a larger impact on the standard deviation ($\kappa_1 = 7.115$) but slope also has an important (negative) contribution ($\kappa_2 = -2.161$). In contrast with the stationary variance model, the range parameter of the non-stationary variance model is much larger ($a = 544$ m), while having a similar short-distance correlation parameter.

The maps of the (unconditional) means show large differences between the two models, both in terms of the magnitude of spatial variation and the spatial pattern (Fig. 3). There is more spatial variation on the map obtained with the non-stationary variance model. The difference in spatial pattern is due to the use of different covariates: elevation was used for the stationary variance model and wetness index for the non-stationary variance model. Despite the large differences in unconditional mean, the kriging prediction maps obtained with the two models are quite similar (Fig. 3). They have comparable ranges of predicted values and a similar spatial pattern. The non-stationary variance model exhibits greater fine-scale detail than the stationary variance model.

The spatial patterns of the unconditional standard deviation maps of the two models are clearly different (Fig. 4). The unconditional standard deviation of the stationary variance model is constant (equal to 11.98), while for the non-stationary variance model it varies smoothly through the area with high values in the East and low values in the West. The higher values are about five times greater than the low values, which clearly indicates that gamma K has non-stationary variance. The standard deviation shows a spatial trend due to the prominent effect of elevation, see Fig. 2 (a). This is refined at the local scale by the slope, where a large slope value leads to a smaller standard deviation (such as the two patches in the south of the area in Figs. 2 (c) and 4). The kriging standard deviation map of the stationary variance model has a familiar pattern with circular areas around observation locations with relatively low values. Uncertainty quickly increases with distance from the sampling locations because of the small range parameter (Table 1). The kriging standard deviation map of the non-stationary variance model shows the combined effect of the East-West standard deviation trend and the circular areas with low values near observation locations. On average, the kriging standard deviation of the non-stationary variance model is considerably smaller than that of the stationary variance model (see the mean kriging variance (MKV) in Table 2). This is confirmed by the θ statistics, which suggest that the stationary variance model severely overestimates the true uncertainty (see Section 3.4 below). For comparison, Fig. 5 shows a map of the local variance of the observation residuals of the non-stationary variance model.

3.4. Quality of predictions and prediction error variance

Table 2 shows that the log-likelihood (Eq. (10)) of the non-stationary variance model is larger than that of the stationary variance model, while the AIC (Eq. (9)) is smaller despite the larger number of model parameters. The non-stationary variance model provides a lower model efficiency measure (MEC of 0.692 against 0.714 for the stationary variance model) and a slightly larger accuracy measure (RMSE of 6.18 against 5.96 for the stationary variance model) but with a ME almost 20% smaller. This implies that the variance of the non-stationary variance model error is larger, because the mean squared error is the sum of the variance and the squared mean error.

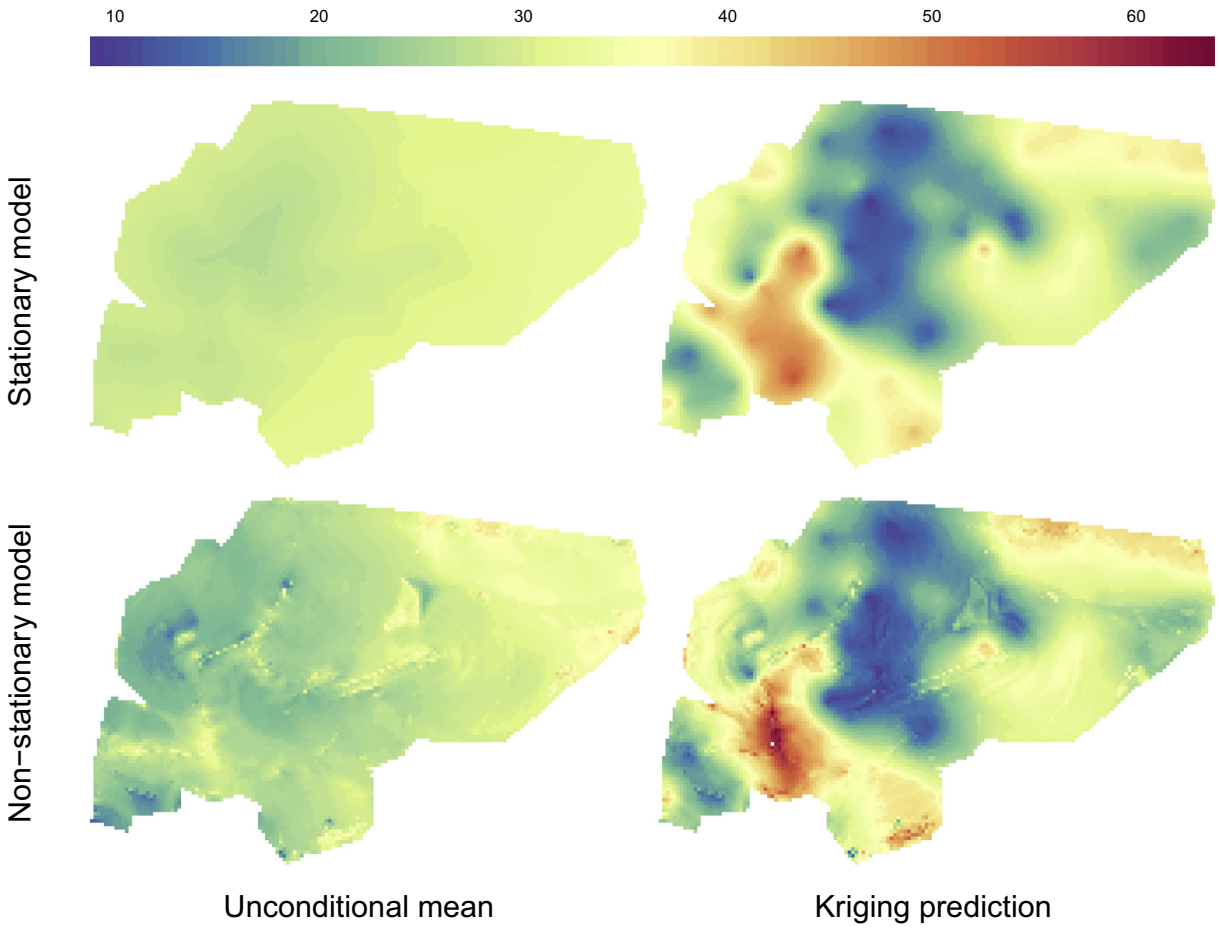

Fig. 3. Maps of the unconditional mean and kriging prediction.

The mean of θ (standardized squared prediction error) is closer to 1 for the non-stationary variance model. Fig. 6 shows the spatial pattern of the mean of θ for both models, computed using a local window. The mean θ of the stationary variance model has a clear spatial pattern, with an strong underestimation of the variance in the eastern part and an overestimation in the western part of the area. The pattern is much more smoothed for the non-stationary variance model, even though there remains a slight underestimation of the variance in the mid-eastern part of the area. The median of θ indicates a large improvement of the non-stationary variance model over the stationary variance model. The 0.2184 value of the median θ for the stationary variance model indicates that this model seriously overestimates the prediction error variance. This is confirmed by the accuracy plots (Fig. 7): prediction intervals as obtained with the stationary variance model are too wide, leading to larger absolute deviation from the nominal p values (the absolute deviation parameter A of the stationary variance model is six times larger than that of the non-stationary variance model). The over- and underestimation measures P_O and P_U show that for both models the main problem is overestimation of the uncertainty, although the mean θ and Fig. 7 suggest that for the non-stationary variance model, this is likely a chance effect and not significant. This problem is more severe for the stationary variance model, where absolute deviation from the nominal p is almost entirely due to overestimation of the uncertainty.

4. Discussion

For gamma K prediction, stationary and non-stationary variance models have very different unconditional mean maps. This is because different covariates were selected. Gamma K is linked to elevation for the stationary variance model and to topographic wetness for the non-stationary variance model. This difference may be explained by the large number of soil forming factors influencing spatial variation of gamma K. For example, Viscarra Rossel et al. (2007) showed that gamma K is mostly determined by soil minerals and soil particle size, with some effects of soil moisture and bulk density. In our case study, the area is covered by a wide range of parent materials, such as lithic sandstone, siltstone, mudstone, shale, limestone and volcanic rocks; as well as a large number of soil types such as Red Dermosols, followed by Brown, Black and Grey Dermosols (Kovac and Lawrie, 1991; Stockmann et al., 2012). These contrasting soil types and parent materials lead to large spatial variation of soil mineralogy and soil particle size, as well as soil moisture characteristics. None of these factors were directly included as a covariate, although each may have a (complex) relationship with the four covariates included in this study. It is therefore difficult to draw conclusions regarding a possible causal effect of the selected mean covariates on the gamma K distribution.

In spite of the apparent differences in the unconditional means between the two models, the final prediction maps were nearly the same and quite different from the unconditional means. This shows that in

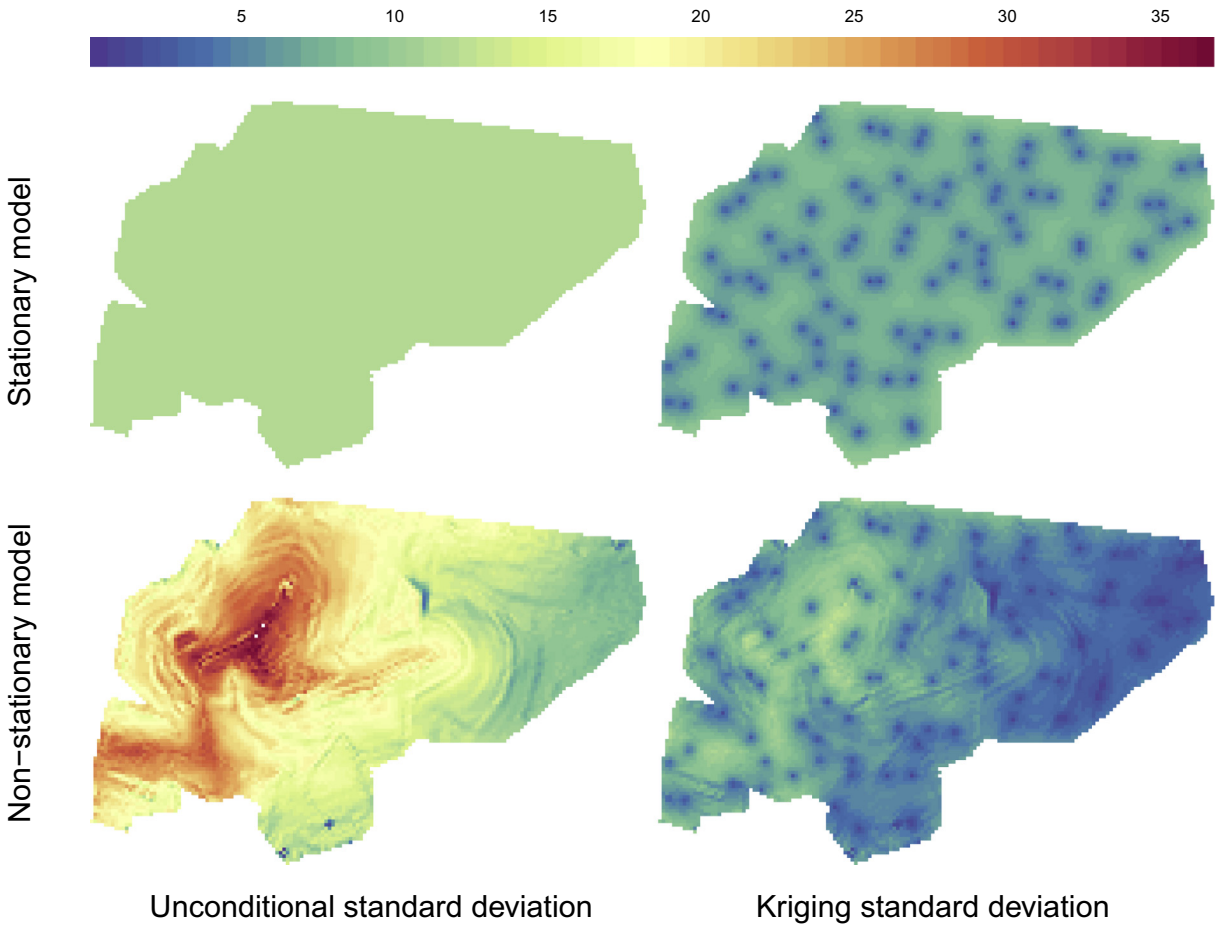

Fig. 4. Maps of the unconditional standard deviation and kriging standard deviations.

Table 2
Performance indicators for the stationary and non-stationary variance models.

	Stationary variance model	Non-stationary variance model
AIC	1667	756.2
Log-likelihood	−828.4	−371.2
MKV	49.4	37.4
ME	−0.84	−0.68
RMSE	5.96	6.18
MEC	0.714	0.692
Mean(θ)	0.718	1.126
Median(θ)	0.218	0.395
A	0.094	0.016
P_O	1.000	0.914
P_U	0.000	0.086

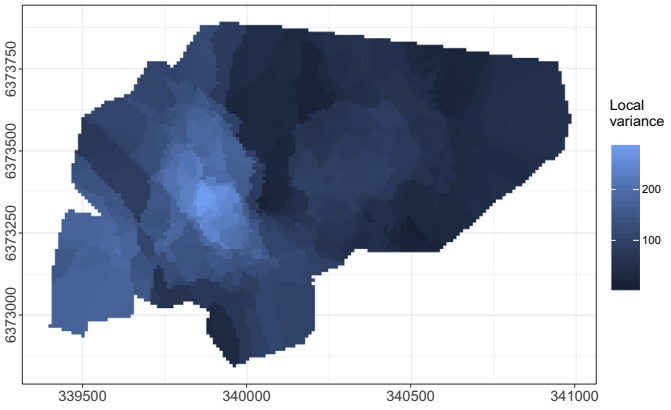

Fig. 5. Local variance of the observation residuals of the non-stationary variance model, calculated using twelve nearest neighbour observations as implemented in the R package RANN (Arya et al., 2017).

this study the kriging step is important. It not only improves prediction accuracy but also compensates for the differences made in modelling the unconditional mean, thus producing more robust predictions that are less sensitive to choices made during the model selection process.

Elevation and slope were selected to model the standard deviation of the non-stationary variance model. Elevation had a positive and slope a negative effect. Apparently, residual variation is greater at high elevation and shallow slopes. Stockmann et al. (2012) note that the top-of-hill vineyards were irrigated during the survey. This might have led to an increase of the local variance with elevation, since TWI does not account for this unexpected artificial process (reflected in Fig. 5). The role of slope on the standard deviation is more difficult to explain, and we do not wish to speculate. In fact, many authors have noted that interpretation of empirical digital soil mapping models is difficult (e.g., Bishop and McBratney, 2001). Recent work, such as by Angelini et al. (2017), have made a step towards 'conscious' digital soil mapping,

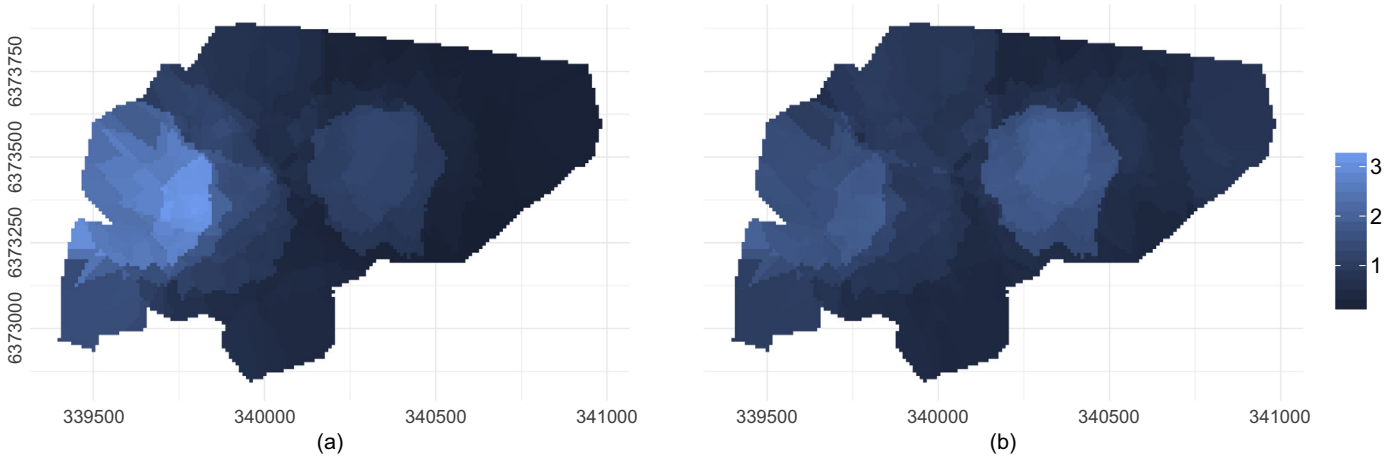

Fig. 6. Maps of the local mean θ for the (a) stationary variance model and (b) non-stationary variance model, calculated using twelve nearest neighbour observations and implemented with the R package RANN (Arya et al., 2017).

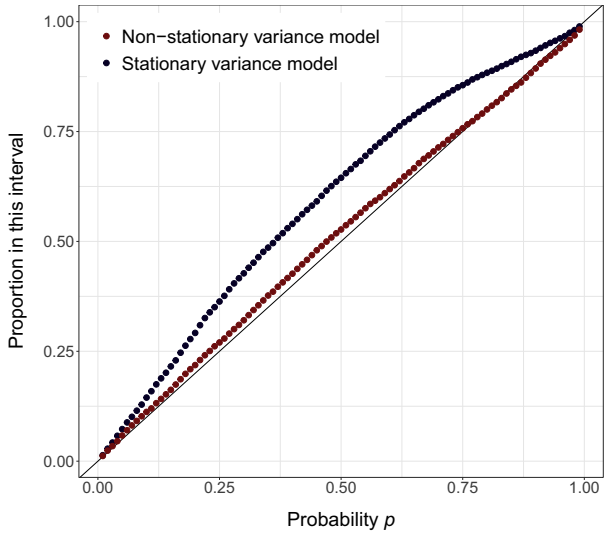

Fig. 7. Accuracy plot for the stationary variance and non-stationary variance models.

where the selection of covariates and their role within the model are primarily based on a soil-landscape conceptual model. While this is challenging for the mean, it is even more difficult to explain how local soil spatial variation (i.e., the standard deviation) is influenced by covariates. If non-stationary variance models gain popularity in future digital soil mapping research, then pedological interpretation of the selected models, including the structure of the standard deviation, requires attention.

For the case study, both models provide good predictive ability as shown by the ME, RMSE and MEC. The spatial patterns of the prediction maps also closely resembles those produced by others (e.g., McBratney and Minasny, 2013). In spite of the good performance, the stationary variance model did not provide satisfactory results regarding uncertainty quantification. The low median of θ and the accuracy plots in Fig. 7 show that the stationary variance model systematically overestimates the local standard deviation for most prediction intervals, except at the tails (0.1 and 0.9 predictive intervals). This is not surprising as the study area reveals strong non-stationarity (Fig. 5), which the stationary variance model cannot capture. The non-stationary variance model is more flexible and can address local differences in standard deviation. The median θ and accuracy plot statistics show that uncertainty quantification is significantly improved using this model. Given the substantial differences between the two kriging standard deviation maps shown in Fig. 4, this has important implications, such as for uncertainty propagation (Heuvelink, 1998) and sampling design optimization studies (Wadoux et al., 2017).

The non-stationary variance model did not improve the accuracy of the predictions. The RMSE of the non-stationary variance model is slightly greater than that of the stationary variance model and the MEC is slightly smaller. The mean kriging variance is about 25 % smaller, suggesting that the non-stationary variance model is more accurate than the stationary variance model, but this merely reflects that the stationary variance model systematically overestimated the uncertainty. The smaller RMSE of the stationary variance model suggests that having a more flexible variance leads to slightly worse predictions. We investigated this by comparing results of the non-stationary model with those of its sub-models (including the stationary version). The results (not shown) confirm that the non-stationary model provides slightly worse predictions than its stationary sub-model. There is no obvious explanation and this effect may be investigated more closely in future work.

While it was shown above that the assumption of a stationary variance was too restrictive for the case study and produced an unrealistic model of the true spatial variation of gamma K, the extension to a non-stationary variance model poses additional problems. We used a model in which the standard deviation is a linear combination of covariates. Parameter estimation and kriging require the inverse of the covariance matrix C, which depends on the covariates through matrix H defined in Eq. (4). Thus, C may become near-singular or even singular for specific combinations of the standard deviation covariates, which leads to numerical instability. Different approaches may be used to avoid this problem. Inspired by Marchant et al. (2009), Wadoux et al. (2017) propose to reject combinations of parameters suggested by differential evolution if these lead to a reciprocal condition number smaller than a given threshold. This seems to work fine but it affects the search for optimal parameters in parameter space, which might lead to sub-optimal parameter combinations. Alternatively, singularity might be tackled by making use of the generalized inverse (Sen and Srivastava, 2012), since kriging is about solving a set of linear equations which can also be accomplished using a generalized inverse. We have not investigated this and used the method proposed by Wadoux et al. (2017) instead. Another solution might be to model the log-transformed standard deviation as a linear combination of covariates (as in Pintore and Holmes, 2004). This would assure that the standard deviation is positive regardless of the parameter values. We explored this approach by taking $\sigma(s) = e^{(\sum_{l=0}^{L} x_l g_l(\mathbf{s}))}$ and re-estimating the parameters on the log-scale. We observed that a slight change in a parameter value may lead to a large change in the standard deviation. The estimated standard

deviation therefore became very unstable and near-singularity still occurred. Thus, we did not pursue this any further.

In this work, uncertainty about the mean regression coefficients was accounted for in the second term of Eq. (12), while uncertainty about the standard deviation coefficients and correlogram parameters was ignored. This may lead to underestimation of the actual 'true' uncertainty. Taking uncertainty about the correlogram parameters and standard deviation regression coefficients into account is possible, although it complicates the analysis. One possible approach is described in Lark (2002) and Marchant and Lark (2007). The authors use the derivative of the kriging prediction error variance and kriging weights with respect to the variogram parameters to infer a map of the covariance uncertainty, which can then be added to the prediction error variance map. We anticipate that this can be easily extended with respect to the standard deviation regression coefficients. Another technique to account for uncertainty in all parameters is to take a Bayesian approach, such as in Diggle and Ribeiro (2007), Chapt. 8. We judged uncertainty in the standard deviation coefficients and correlogram parameters to play a minor role, given that the total number of model parameters was much smaller than the number of observations, but taking these additional sources of uncertainty into account would certainly make a valuable extension.

There is also a need to further investigate the data requirements (number of observations and their spatial locations) for adequate fitting of non-stationary variance models. In this work, we fitted a total of seven parameters (including two additional standard deviation parameters) for the non-stationary variance model from 100 observation locations, which we considered adequate. However, the number of standard deviation parameters when using a non-stationary variance model might grow manifold, such as when using a more complex standard deviation function (e.g. splines). It has been demonstrated that covariance uncertainty is minimal when observations are clustered (Marchant and Lark, 2007), while interpolation error is reduced by spreading the observations in geographic and feature (i.e. covariate) space (Brus and Heuvelink, 2007). How large the sample size should be and which sampling design is best for estimation of the standard deviation coefficients has not been thoroughly explored, although Wadoux et al. (2017) indicate that the non-stationary variance model benefits from spreading the observations in the standard deviation covariate space, while keeping the sampling density fairly constant over the area.

Finally, we emphasize the value of improved uncertainty quantification, as obtained through the use of a non-stationary variance model. Map users often take the prediction map as their first interest, but visualization of the prediction alone can give a wrong impression about the quality of the map (Hengl and Toomanian, 2006) and bias the subsequent decision-making process (Goovaerts, 2001). Uncertainty quantification of the prediction is as important as the prediction itself to obtain a full impression about the quality of the maps. If the uncertainty is too large, users may decide to invest in obtaining a more accurate map (Heuvelink, 2014), but they can only make such decision if they have reliable information about the map quality. We also emphasize the importance of validating the uncertainty estimate, with measures such as θ statistics and accuracy plots. Our study showed that assuming stationarity in the variance can lead to erroneous quantification of uncertainty. In such cases we advocate the use of the non-stationary variance model because it is more flexible and leads to improved estimation of the prediction uncertainty.

5. Conclusion

We tested the non-stationary variance model developed in Lark (2009) for spatial interpolation of a soil property. We used multiple covariates to model the spatial standard deviation. Covariates were chosen based on the Akaike information criterion and model parameters fitted using a maximum likelihood approach. We compared the non-stationary variance model to the stationary variance model in a case study. The main conclusions are:

- When modelling a soil property that exhibits local differences in spatial variation, using a non-stationary variance model is recommended over a stationary variance model because it yields a more realistic quantification of prediction uncertainty. Using a constant standard deviation is often not realistic and may lead to local over- or underestimation of the uncertainty.
- Estimation of the parameters of a non-stationary variance model is hampered by near-singularity of the covariance matrix, for which several solutions are proposed but that need further investigation.
- In a case study mapping gamma K in the Hunter Valley, Australia, the kriging standard deviation maps of the stationary and non-stationary variance models were very different. Evaluation using independent validation data showed that the non-stationary model captured the uncertainty much better.
- In the case study different covariates were chosen to model the unconditional mean of the stationary and non-stationary variance models. However, the kriging prediction maps were nearly the same. This suggests that these are insensitive to choices made in the model selection process. Future research may show whether this is a consistent finding or case-dependent.

Acknowledgements

We thank Budiman Minasny of the University of Sydney, Australia, for providing the data used in the case study. This project received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000. We thank the anonymous reviewers for their helpful and constructive comments that helped to improve this manuscript.

References

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike. Springer, pp. 199–213.
Angelini, M.E., Heuvelink, G.B.M., Kempen, B., 2017. Multivariate mapping of soil with structural equation modelling. Eur. J. Soil Sci. 68, 575–591.
Ardia, D., Mullen, K.M., Peterson, B.G., Ulrich, J., 2015. DEoptim: Differential Evolution in R. http://CRAN.R-project.org/package=DEoptim Version 2.2-3.
Arya, S., Mount, D., Kemp, S.E., Jefferis, G., 2017. RANN: Fast Nearest Neighbour Search (Wraps ANN Library) Using L2 Metric. https://CRAN.R-project.org/package=RANN R package version 2.5.1.
Bishop, T., McBratney, A., 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103, 149–160.
Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138, 86–95.
Brus, D.J., Yang, R.-M., Zhang, G.-L., 2016. Three-dimensional geostatistical modeling of soil organic carbon: a case study in the Qilian mountains, China. Catena 141, 46–55.
Deutsch, C., 1997. Direct assessment of local accuracy and precision. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong'96, pp. 115–125.
Diggle, P.J., Ribeiro, P.J., 2007. Model based Geostatistics. Springer Series in Statistics.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press on Demand.
Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103, 3–26.
Haskard, K.A., Lark, R.M., 2009. Modelling non-stationary variance of soil properties by tempering an empirical spectrum. Geoderma 153, 18–28.
Hengl, T., Toomanian, N., 2006. Maps are not what they seem: representing uncertainty in soil-property maps. In: Proc. Accuracy, pp. 805–813.
Heuvelink, G.B.M., 1998. Error Propagation in Environmental Modelling with GIS. CRC Press.
Heuvelink, G.B.M., 2014. Uncertainty Quantification of GlobalSoilMap Products. Taylor & Francis Group London.
Hoeting, J.A., Davis, R.A., Merton, A.A., Thompson, S.E., 2006. Model selection for geostatistical models. Ecol. Appl. 16, 87–98.
Jacques, D., Mouvet, C., Mohanty, B., Vereecken, H., Feyen, J., 1999. Spatial variability of atrazine sorption parameters and other soil properties in a podzoluvisol. J. Contam. Hydrol. 36, 31–52.
Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecol. Model. 83, 55–66.
Kass, R.E., Wasserman, L., 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. J. Am. Stat. Assoc. 90, 928–934.
Kovac, M., Lawrie, J.W., 1991. Soil Landscapes of the Singleton 1:250.000 sheet. Soil Conservation Service of NSW, Sydney.

Lark, R.M., 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. Eur. J. Soil Sci. 51, 717–728.
Lark, R.M., 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. Geoderma 105, 49–80.
Lark, R.M., 2009. Kriging a soil variable with a simple nonstationary variance model. J. Agric. Biol. Environ. Stat. 14, 301–321.
Lark, R.M., Webster, R., 2006. Geostatistical mapping of geomorphic variables in the presence of trend. Earth Surf. Process. Landf. 31, 862–874.
Marchant, B.P., Lark, R.M., 2007. Optimized sample schemes for geostatistical surveys. Math. Geol. 39, 113–134.
Marchant, B.P., Newman, S., Corstanje, R., Reddy, K.R., Osborne, T.Z., Lark, R.M., 2009. Spatial monitoring of a non-stationary soil property: phosphorus in a Florida water conservation area. Eur. J. Soil Sci. 60, 757–769.
McBratney, A.B., Minasny, B., 2013. Spacebender. Spat. Stat. 4, 57–67.
McBratney, A.B., Webster, R., 1981. Spatial dependence and classification of the soil along a transect in Northeast Scotland. Geoderma 26, 63–82.
Moore, I.D., Gessler, P., Nielsen, G., Peterson, G., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57, 443–452.
Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.
Pintore, A., Holmes, C., 2004. Spatially adaptive non-stationary covariance functions via spatially adaptive spectra. In: Technical Report. University of Oxford, Oxford, UK.
Sawicka, K., Wadoux, A.M.J.-C., Heuvelink, G.B.M., 2017. Sample Design Optimisation Techniques and Associated Software. http://dx.doi.org/10.5281/zenodo.1065954. QUICS project deliverable 3.1.
Sen, A., Srivastava, M., 2012. Regression Analysis: Theory, Methods, and Applications. Springer Science & Business Media.
Stockmann, U., Minasny, B., McBratney, A.B., Hancock, G., Willgoose, G., 2012. Exploring short-term soil landscape formation in the Hunter Valley, NSW, using gamma ray spectrometry. In: Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012, Sydney, Australia. CRC Press, pp. 77.
Storn, R., Price, K., 1997. Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. 11, 341–359.
Viscarra Rossel, R.A., Taylor, H.J., McBratney, A.B., 2007. Multivariate calibration of hyperspectral γ-ray energy spectra for proximal soil sensing. Eur. J. Soil Sci. 58, 343–353.
Voltz, M., Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. Eur. J. Soil Sci. 41, 473–490.
Wadoux, A.M.J.C., Brus, D.J., Rico-Ramirez, M.A., Heuvelink, G.B.M., 2017. Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model. Adv. Water Resour. 107, 126–138.
Walvoort, D.J.J., Brus, D.J., De Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Comput. Geosci. 36, 1261–1267.
Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons.
Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landf. 12, 47–56.