CrossMark

A fully non-stationary linear coregionalization model for multivariate random fields

Francky Fouedjio[1] ●

Abstract This paper introduces an extension of the traditional stationary linear coregionalization model to handle the lack of stationarity. Under the proposed model, coregionalization matrices are spatially dependent, and basic univariate spatial dependence structures are non-stationary. A parameter estimation procedure of the proposed non-stationary linear coregionalization model is developed under the local stationarity framework. The proposed estimation procedure is based on the method of moments and involves a matrix-valued local stationary variogram kernel estimator, a weighted local least squares method in combination with a kernel smoothing technique. Local parameter estimates are knitted together for prediction and simulation purposes. The proposed non-stationary multivariate spatial modeling approach is illustrated using two real bivariate data examples. Prediction performance comparison is carried out with the classical stationary multivariate spatial modeling approach. According to several criteria, the prediction performance of the proposed non-stationary multivariate spatial modeling approach appears to be significantly better.

Keywords Linear coregionalization model · Non-stationarity · Local stationarity · Non-parametric · Cokriging · Cosimulation

✉ Francky Fouedjio
francky.fouedjiokameni@csiro.au

1   CSIRO Mineral Resources, 26 Dick Perry Ave, Kensington, WA 6151, Australia

1 Introduction

Continuously indexed datasets with multiple variables are often present in many disciplines in the geosciences. When the interest is in predicting or simulating, multivariate random fields are a natural modeling choice for multiple variables measured at several locations. The spatial dependence structure of the multivariate random field is commonly depicted by direct and cross covariance functions (or direct and cross variograms) describing the relationship between variables across locations. Their modeling and estimation are fundamental for prediction and simulation.

Simplifying modeling assumptions are often made on direct and cross covariance functions (or direct and cross variograms). They include the stationarity assumption which states that direct and cross covariance functions (or direct and cross variograms) remain constant over the whole study domain. This assumption is driven more by mathematical convenience than by reality. In practice, it is often violated due to some local influences or localized effects which can be reflected by computing local stationary direct and cross variograms whose characteristics may vary spatially, particularly when data come from large study domains. In such case, using a stationary multivariate spatial modeling approach would be liable to produce less accurate prediction, including an incorrect evaluation of the prediction error. So research has been directed toward developing non-stationary multivariate spatial modeling approaches.

Although many non-stationary univariate spatial modeling approaches have been developed over the last decade (see Fouedjio 2016 for a review), non-stationary multivariate spatial modeling approaches are very few in the current literature (see Genton and Kleiber 2015 for a

🙋 Springer

review). The main difficulty is in specifying valid non-stationary cross covariance functions (or cross variograms). Existing non-stationary multivariate spatial modeling approaches include works by Gelfand et al. (2004), Majumdar et al. (2010), Kleiber and Nychka (2012), and Kleiber and Porcu (2015). The most popular stationary multivariate spatial model commonly known as linear coregionalization model (Journel and Huijbregts 1978; Bourgault and Marcotte 1991; Goulard and Voltz 1992; Grzebyk and Wackernagel 1994; Vargas-Guzmán et al. 2002; Wackernagel 2003) was extended to the first plausible non-stationary multivariate spatial model by Gelfand et al. (2004). The idea consists in letting the coefficients of basic univariate spatial dependence structures of the model to vary across space. However, basic univariate spatial dependence structures of the model remain stationary. Majumdar et al. (2010) generalize the stationary multivariate convolution model proposed by Majumdar and Gelfand (2007) to the non-stationary setting. The modeling approach is based on convolutions of spatially varying covariance kernels. Kleiber and Nychka (2012) and Kleiber and Porcu (2015) respectively extend the stationary Matérn and Cauchy cross covariance functions to allow for spatially varying variance, scale, smoothness, and range-dependence parameters.

This paper introduces an extension of the conventional stationary linear coregionalization model to handle the lack of stationarity. Under the proposed model, coregionalization matrices are spatially dependent, and basic univariate spatial dependence structures are non-stationary. A three-step parameter estimation procedure of the proposed non-stationary linear coregionalization model is developed under the local stationarity setting. Firstly, a matrix-valued local stationary variogram kernel moment estimator is defined at any location of interest. Then, a weighted local least squares procedure is carried out to estimate parameters at a set of locations referred to as anchor locations covering the study region. Finally, estimated parameters at any location of interest are obtained through a kernel smoothing technique. The proposed estimation framework is distribution free and computationally attractive. It does not impose any distributional assumptions except the existence of the first and second order moment structures. It does not require any calculation of inverse or determinant of matrices. Local parameter estimates are knitted together for purposes of prediction and simulation. The proposed non-stationary multivariate spatial modeling approach is illustrated using two practical examples: geophysical and geochemical bivariate data.

The format of the remainder of the paper is as follows. Section 2 introduces the non-stationary linear coregionalization model. Section 3 laid out the parameter estimation details. Section 4 tackles cokriging and conditional cosimulations in this non-stationary framework. Section 5 contains illustrations of the proposed non-stationary multivariate spatial modeling approach using two practical examples: geophysical and geochemical bivariate data. Section 6 offers concluding discussion and few summary remarks.

2 Proposed model

Consider a p-dimensional vector-valued random field $\mathbf{Z}(\mathbf{x}) = \left[Z_1(\mathbf{x}), \ldots, Z_p(\mathbf{x})\right]^T$ defined over a spatial domain of interest G of $\mathbb{R}^d (d \geq 1)$, and having the following representation:

$$\mathbf{Z}(\mathbf{x}) = \boldsymbol{\mu}(\mathbf{x}) + \sum_{u=1}^{r} \mathbf{A}^{(u)}(\mathbf{x})\mathbf{W}^{(u)}(\mathbf{x}) + \mathbf{A}^{(0)}(\mathbf{x})\boldsymbol{\epsilon}(\mathbf{x}), \ \forall \mathbf{x} \in G, \tag{1}$$

where $\boldsymbol{\mu}(\cdot)$ is an unknown vector-valued fixed function; $\left\{\mathbf{A}^{(u)}(\cdot)\right\}_{u=0}^{r}$ are unknown $p \times p$ matrix-valued fixed functions; $\left\{\mathbf{W}^{(u)}(\cdot)\right\}_{u=1}^{r}$ are mutually independent p-dimensional vector-valued random fields whose their components are mutually independent, each with zero mean, unit variance, and basic non-stationary univariate correlation function $\rho_{NS}^{(u)}(\cdot, \cdot)$; $\boldsymbol{\epsilon}(\cdot)$ is a p-dimensional vector-valued random field independent of $\left\{\mathbf{W}^{(u)}(\cdot)\right\}_{u=1}^{r}$ and whose components are standard white noises.

Under the model defined in Eq. (1), the first and second order moment structures of the vector-valued random field $\mathbf{Z}(\cdot)$ are expressed as follows (the proof is given in Appendix):

$$\mathbb{E}(\mathbf{Z}(\mathbf{x})) = \boldsymbol{\mu}(\mathbf{x}); \tag{2}$$

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{y})) &= \sum_{u=1}^{r} \mathbf{A}^{(u)}(\mathbf{x})\mathbf{A}^{(u)}(\mathbf{y})^T \rho_{NS}^{(u)}(\mathbf{x}, \mathbf{y}) \\
&\quad + \mathbf{A}^{(0)}(\mathbf{x})\mathbf{A}^{(0)}(\mathbf{y})^T \rho_S^{(0)}(\|\mathbf{x} - \mathbf{y}\|) \\
&\equiv \mathbf{C}(\mathbf{x}, \mathbf{y}),
\end{aligned} \tag{3}$$

where $\rho_S^{(0)}(\cdot)$ is the nugget correlation function defined by $\rho_S^{(0)}(\|\mathbf{h}\|) = 1$ for $\|\mathbf{h}\| = 0$ and $\rho_S^{(0)}(\|\mathbf{h}\|) = 0$ $\|\mathbf{h}\| > 0$. The superscript T denotes matrix transposition.

The matrix-valued covariance function $\mathbf{C}(\cdot, \cdot)$ described in Eq. (3) is a valid covariance structure by construction (the proof is given in Appendix). The covariance matrix $\mathbf{C}(\mathbf{x}, \mathbf{y})$ is a $p \times p$ matrix that contains direct covariances $\{\mathbb{C}\mathrm{ov}(Z_i(\mathbf{x}), Z_i(\mathbf{y})\}_{i=1}^p$ along its major diagonal and cross covariances $\{\mathbb{C}\mathrm{ov}(Z_i(\mathbf{x}), Z_j(\mathbf{y})\}_{i,j=1, i \neq j}^p$ off that diagonal. The variance-covariance matrix $\mathbf{C}(\mathbf{x}, \mathbf{x})$ is expressed as the sum of the $(r + 1)$ coregionalization matrices

$\left\{ \mathbf{A}^{(u)}(\mathbf{x}) \mathbf{A}^{(u)}(\mathbf{x})^T \right\}_{u=0}^r$. The basic non-stationary univariate correlation functions family $\left\{ \rho_{NS}^{(u)}(\cdot, \cdot) \right\}_{u=1}^r$ is chosen in the class of closed-form non-stationary univariate correlation functions with locally varying geometric anisotropy introduced by Paciorek and Schervish (2006). Thus, the matrix-valued covariance function $\mathbf{C}(\cdot, \cdot)$ takes the following form:

$$
\begin{aligned}
\mathbf{C}(\mathbf{x}, \mathbf{y}) = & \sum_{u=1}^r \mathbf{A}^{(u)}(\mathbf{x}) \mathbf{A}^{(u)}(\mathbf{y})^T \phi_{\mathbf{xy}}^{(u)} \rho_S^{(u)}\left(\sqrt{Q_{\mathbf{xy}}^{(u)}(\mathbf{x} - \mathbf{y})} \right) \\
& + \mathbf{A}^{(0)}(\mathbf{x}) \mathbf{A}^{(0)}(\mathbf{y})^T \rho_S^{(0)}(\|\mathbf{x} - \mathbf{y}\|),
\end{aligned}
\tag{4}
$$

where $\quad \phi_{\mathbf{xy}}^{(u)} = \left| \boldsymbol{\Sigma}_{\mathbf{x}}^{(u)} \right|^{\frac{1}{4}} \left| \boldsymbol{\Sigma}_{\mathbf{y}}^{(u)} \right|^{\frac{1}{4}} \left| \dfrac{\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)} + \boldsymbol{\Sigma}_{\mathbf{y}}^{(u)}}{2} \right|^{-\frac{1}{2}}, \quad Q_{\mathbf{xy}}^{(u)}(\mathbf{h}) =$

$\mathbf{h}^T \left(\dfrac{\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)} + \boldsymbol{\Sigma}_{\mathbf{y}}^{(u)}}{2} \right)^{-1} \mathbf{h}, \quad \forall \mathbf{h} \in \mathbb{R}^d; \quad \{ \boldsymbol{\Sigma}^{(u)}(\cdot) : \ \mathbb{R}^d \to PD_d(\mathbb{R}),$ $\mathbf{x} \mapsto \boldsymbol{\Sigma}_{\mathbf{x}}^{(u)} \}_{u=1}^r$ are mappings from $\mathbb{R}^d$ to $PD_d(\mathbb{R})$ the set of real-valued positive definite d-dimensional square matrices; $\left\{ \rho_S^{(u)}(\cdot) \right\}_{u=1}^r$ are continuous isotropic stationary univariate correlation functions positive definite on $\mathbb{R}^d$, for all $d \in \mathbb{N}^\star$ (e.g., Matérn family, Cauchy family, max stable family). Matrix-valued functions $\left\{ \boldsymbol{\Sigma}^{(u)}(\cdot) \right\}_{u=1}^r$ are parametrized via the spectral decomposition, thus ensuring their positive definiteness: $\forall \mathbf{x}_0 \in G, \quad \boldsymbol{\Sigma}_{\mathbf{x}_0}^{(u)} = \boldsymbol{\Psi}_{\mathbf{x}_0}^{(u)} \boldsymbol{\Lambda}_{\mathbf{x}_0}^{(u)}$ $\boldsymbol{\Psi}_{\mathbf{x}_0}^{(u)T}$, where $\boldsymbol{\Lambda}_{\mathbf{x}_0}^{(u)}$ is the diagonal matrix of eigenvalues and $\boldsymbol{\Psi}_{\mathbf{x}_0}^{(u)}$ is the eigenvector matrix formulated in 2D as follows:

$$
\boldsymbol{\Lambda}_{\mathbf{x}_0}^{(u)} = \begin{pmatrix} \lambda_1^{(u)^2}(\mathbf{x}_0) & 0 \\ 0 & \lambda_2^{(u)^2}(\mathbf{x}_0) \end{pmatrix},
$$
$$
\boldsymbol{\Psi}_{\mathbf{x}_0}^{(u)} = \begin{pmatrix} \cos \psi^{(u)}(\mathbf{x}_0) & \sin \psi^{(u)}(\mathbf{x}_0) \\ -\sin \psi^{(u)}(\mathbf{x}_0) & \cos \psi^{(u)}(\mathbf{x}_0) \end{pmatrix},
$$

with $\lambda_1^{(u)}(\mathbf{x}_0), \lambda_2^{(u)}(\mathbf{x}_0) > 0$ controlling the local ranges and $\psi^{(u)}(\mathbf{x}_0) \in [0, \pi[$ specifying the local geometric anisotropy angle.

In Eq. (4), matrix-valued coregionalization functions $\left\{ \mathbf{A}^{(u)}(\cdot) \right\}_{u=0}^r$, and matrix-valued geometric anisotropy functions $\{ \boldsymbol{\Sigma}^{(u)}(\cdot) \}_{u=1}^r$ are enabled to vary spatially. Thus, the resulting matrix-valued covariance function $\mathbf{C}(\cdot, \cdot)$ is non-stationary, i.e., the simple and cross covariance functions are now dependent on the spatial location pair and not just the lag vector, $\mathbb{C}ov(Z_i(\mathbf{x}), Z_j(\mathbf{y})) = C_{ij}(\mathbf{x}, \mathbf{y})$. In the non-stationary model defined by Eq. (4), when $\mathbf{x} \neq \mathbf{y}$, we may have $\mathbf{A}^{(u)}(\mathbf{x}) \neq \mathbf{A}^{(u)}(\mathbf{y})$ and so $\mathbf{A}^{(u)}(\mathbf{x}) \mathbf{A}^{(u)}(\mathbf{y})^T \neq \mathbf{A}^{(u)}(\mathbf{y}) \mathbf{A}^{(u)}(\mathbf{x})^T$. Thus, this non-stationary model

may provide a non-symmetric cross-covariance matrix function, $C_{ij}(\mathbf{x}, \mathbf{y}) \neq C_{ji}(\mathbf{x}, \mathbf{y})$. It is also important to note that by definition $\mathbf{C}(\mathbf{x}, \mathbf{y}) = \mathbf{C}(\mathbf{y}, \mathbf{x})^T$. The non-stationary model described by Eq. (4) includes also the conventional stationary linear coregionalization model as a special case. Moreover, it is more flexible than the spatially varying linear coregionalization model of Gelfand et al. (2004) where basic univariate correlation functions are rather considered stationary.

3 Estimating model parameters

Suppose that the p-dimensional vector-valued random field $\mathbf{Z}(\cdot)$ is observed at known spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset G$. Let us assume that all components of $\mathbf{Z}(\cdot)$ has all data values at all sample locations (isotopic sampling). Without loss of generality, we consider that the study domain G is part of $\mathbb{R}^2 (d = 2)$. The goal is to estimate at any location of interest the following parameters: the vector-valued mean function $\boldsymbol{\mu}(\cdot)$, the matrix-valued coregionalization functions $\left\{ \mathbf{A}^{(u)}(\cdot) \right\}_{u=0}^r$, and the matrix-valued geometric anisotropy functions $\left\{ \boldsymbol{\Sigma}^{(u)}(\cdot) \right\}_{u=1}^r$ characterized by $\left\{ \lambda_1^{(u)}(\cdot), \lambda_2^{(u)}(\cdot), \psi^{(u)}(\cdot) \right\}_{u=1}^r$.

The estimation is carried out under the local stationarity or quasi-stationarity assumption (Matheron 1971; Wackernagel 2003). In the quasi-stationarity setup, parameter functions $\boldsymbol{\mu}(\cdot), \left\{ \mathbf{A}^{(u)}(\cdot) \right\}_{u=0}^r, \left\{ \lambda_1^{(u)}(\cdot) \right\}_{u=1}^r, \left\{ \lambda_2^{(u)}(\cdot) \right\}_{u=1}^r,$ and $\left\{ \psi^{(u)}(\cdot) \right\}_{u=1}^r$ are smooth functions varying slowly in space so that at any location of interest $\mathbf{x}_0 \in G$, one can define a neighborhood $\mathcal{V}_{\mathbf{x}_0} = \{\mathbf{x} \in G, \ \|\mathbf{x}_0 - \mathbf{x}\| \leq \eta\}$ wherein the first and second order moment structures of $\mathbf{Z}(\cdot)$ are approximately stationary. Thus, $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{V}_{\mathbf{x}_0} \times \mathcal{V}_{\mathbf{x}_0}$ Eqs. (2) and (4) are reduced as follows:

$$
\boldsymbol{\mu}(\mathbf{x}) \approx \boldsymbol{\mu}(\mathbf{y}) \approx \boldsymbol{\mu}(\mathbf{x}_0), \tag{5}
$$

$$
\begin{aligned}
\mathbf{C}(\mathbf{x}, \mathbf{y}) \approx & \sum_{u=1}^r \mathbf{B}^{(u)}(\mathbf{x}_0) \rho_S^{(u)}\left(\sqrt{(\mathbf{x} - \mathbf{y})^T \boldsymbol{\Sigma}_{\mathbf{x}_0}^{(u)-1}(\mathbf{x} - \mathbf{y})} \right) \\
& + \mathbf{B}^{(0)}(\mathbf{x}_0) \rho_S^{(0)}(\|\mathbf{x} - \mathbf{y}\|) \\
\equiv & \ \mathbf{C}_S(\mathbf{x} - \mathbf{y}; \mathbf{x}_0),
\end{aligned}
\tag{6}
$$

where $\left\{ \mathbf{B}^{(u)}(\mathbf{x}_0) = \mathbf{A}^{(u)}(\mathbf{x}_0) \mathbf{A}^{(u)}(\mathbf{x}_0)^T \right\}_{u=0}^r$ are local coregionalization matrices.

Thus, locally, the vector-valued non-stationary mean function is constant, and the matrix-valued non-stationary covariance function is reduced to a matrix-valued stationary covariance function. The local estimation is carried out

through a three-step estimation scheme. Firstly, a matrix-valued local stationary variogram kernel moment estimator is defined at any location of interest. Then, a weighted local least squares procedure is performed to estimate parameters at a set of locations referred to as anchor locations covering the study domain. Finally, parameter estimates at any location of interest are obtained through a kernel smoothing technique. All of these different elements are described in this section.

3.1 Matrix-valued local stationary variogram kernel estimator

The matrix-valued local stationary covariance function $\mathbf{C}_S(\cdot; \mathbf{x}_0)$ obtained in Eq. (6) being an even function, it is equivalent to the matrix-valued local stationary variogram $\mathbf{\Gamma}_S(\cdot; \mathbf{x}_0) = \mathbf{C}_S(\mathbf{0}; \mathbf{x}_0) - \mathbf{C}_S(\cdot; \mathbf{x}_0)$. Thus, the estimation can be formulated in a framework with variograms. A non-parametric kernel moment estimator of the matrix-valued local stationary variogram function at a target location $\mathbf{x}_0 \in G$ and for a spatial lag vector $\mathbf{h} \in \mathbb{R}^d$, $\|\mathbf{h}\| \leq \eta$ is defined as follows:

$\eta = \xi$, Eq. (7) leads to the classical moving window estimator.

In Eq. (7) the kernel function $K_\xi(\cdot)$ is chosen as the Gaussian kernel $\left(K_\xi(\|\mathbf{h}\|) \propto \exp(-\frac{1}{2\xi^2}\|\mathbf{h}\|^2) \right)$ whose support is non-compact and therefore includes all sample locations. By doing this, the matrix-valued local stationary variogram kernel estimator is not limited to the local information, remote sample locations are also considered albeit down-weighted. This choice helps to reduce the instability of the matrix-valued local stationary variogram kernel estimator at regions with low sampling density. It avoids artifacts caused by the only use of sample locations close to the target location. Moreover, it avoids the problem of non-smooth local parameter estimates which is incompatible with the local stationarity assumption. Regarding the size of the quasi-stationarity neighborhood η, it sets according to the bandwidth parameter ξ. $\eta = \sqrt{3}\xi$ so that the standard deviation of the Gaussian kernel corresponds to one of the uniform kernel whose support is compact. Other possible choices for η include a quantile of the Gaussian kernel (e.g., $\eta \approx 2\xi$) or the full width at half

$$\widehat{\mathbf{\Gamma}}_S(\mathbf{h}; \mathbf{x}_0) = \frac{\sum_{\mathcal{V}(\mathbf{h})} \tilde{K}_\xi(\|\mathbf{x}_0 - \mathbf{s}_k\|)\tilde{K}_\xi(\|\mathbf{x}_0 - \mathbf{s}_{k\prime}\|)[\mathbf{Z}(\mathbf{s}_k) - \mathbf{Z}(\mathbf{s}_{k\prime})][\mathbf{Z}(\mathbf{s}_k) - \mathbf{Z}(\mathbf{s}_{k\prime})]^T}{2\sum_{\mathcal{V}(\mathbf{h})} \tilde{K}_\xi(\|\mathbf{x}_0 - \mathbf{s}_k\|)\tilde{K}_\xi(\|\mathbf{x}_0 - \mathbf{s}_{k\prime}\|)}, \quad (7)$$

where $\tilde{K}_\xi(\|\mathbf{x}_0 - \mathbf{s}_k\|) = K_\xi(\|\mathbf{x}_0 - \mathbf{s}_k\|)/\sum_{t=1}^n K_\xi(\|\mathbf{x}_0 - \mathbf{s}_t\|)$ are standardized weights with $K_\xi(\cdot)$ a positive kernel function with bandwidth parameter $\xi > 0$. $\mathcal{V}(\mathbf{h}) = \{(\mathbf{s}_k, \mathbf{s}_{k\prime}) \in G \times G : \mathbf{s}_k - \mathbf{s}_{k\prime} = \mathbf{h}\}$ is the set of all pairs of sample locations separated by vector $\mathbf{h}$. In case of irregularly sampled data where there are usually not enough sample locations separated by exactly $\mathbf{h}$, $\mathcal{V}(\mathbf{h})$ is commonly modified by $\{(\mathbf{s}_k, \mathbf{s}_{k\prime}) \in G \times G : \mathbf{s}_k - \mathbf{s}_{k\prime} \in \mathcal{T}(\mathbf{h})\}$, where $\mathcal{T}(\mathbf{h})$ is a tolerance region surrounding $\mathbf{h}$.

In Eq. (7) each pair of sample locations receives a weight proportional to the product of the individual weights. Pairs of sample locations close to the target location have more influence on the matrix-valued local stationary variogram kernel estimator than those which are faraway. The role of the kernel function is to smoothly down-weight the influence of distant sample locations. Taking $K_\xi(\|\mathbf{h}\|) \propto 1$ and η as the radius of the study domain G reduces Eq. (7) to the conventional moment estimator for a global matrix-valued stationary variogram used in practice. For $K_\xi(\|\mathbf{h}\|) \propto \mathbb{1}_{\|\mathbf{h}\| < \xi}$ and

maximum ($\eta = \sqrt{2\log(2)}\xi$).

3.2 Parameter raw estimates

The matrix-valued local stationary variogram kernel estimator defined in Eq. (7) is used to estimate to the unknown quantities $\left\{\mathbf{A}^{(u)}(\mathbf{x}_0)\right\}_{u=0}^r$, $\left\{\lambda_1^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$, $\left\{\lambda_2^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$, and $\left\{\psi^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$ characterizing the local stationary multivariate spatial dependence structure $\mathbf{C}_S(\cdot; \mathbf{x}_0)$ at the target location $\mathbf{x}_0 \in G$. These quantities are determined by minimizing the following local weighted sum of squares criterion:

$$LWSS(\mathbf{x}_0) = \sum_{l=1}^L \omega(\mathbf{h}_l; \mathbf{x}_0)\|\widehat{\mathbf{\Gamma}}_S(\mathbf{h}_l; \mathbf{x}_0) - \mathbf{\Gamma}_S(\mathbf{h}_l; \mathbf{x}_0)\|_F^2, \quad \forall \mathbf{x}_0 \in G, \quad (8)$$

where $\{\mathbf{h}_l \in \mathbb{R}^d, \|\mathbf{h}_l\| \leq \eta\}_{l=1}^L$ is a finite set of spatial lag vectors; similarly to the traditional stationary setting (Chilés and Delfiner 2012), $\omega(\mathbf{h}_l; \mathbf{x}_0) =$

$\left(\sum_{\mathcal{V}(\mathbf{h}_l)} \tilde{K}_\xi(\mathbf{x}_0, \mathbf{s}_k)\tilde{K}_\xi(\mathbf{x}_0, \mathbf{s}_{k\prime})\right)\|\mathbf{h}_l\|^{-1}$ are weights which are chosen to be proportional to the amount of information in the non-parametric kernel estimator defined in Eq. (7) and inversely proportional to the norm of the spatial lag vector (in order to improve the fitting at small distances); $\|\mathbf{X}\|_F^2 = tr(\mathbf{X}\mathbf{X}^T)$ denotes the Frobenius norm of a square matrix $\mathbf{X}$. The minimization is subject to the constraint that the local coregionalization matrices $\{\mathbf{B}^{(u)}(\mathbf{x}_0) = \mathbf{A}^{(u)}(\mathbf{x}_0)\mathbf{A}^{(u)}(\mathbf{x}_0)^T\}_{u=0}^r$ are all positive semi-definite.

The optimization problem defined in Eq. (8) can be solved using existing algorithms developed in the stationary framework (Goulard and Voltz 1992; Lark and Papritz 2003; Emery 2010; Desassis and Renard 2013). The basic idea of these algorithms is to minimize the weighted sum of squares by optimizing each coregionalization matrix successively, and to repeat the process until weighted sum of squares cannot decrease any more.

It is worth pointing out that the estimation of unknown quantities $\left\{\mathbf{A}^{(u)}(\mathbf{x}_0)\right\}_{u=0}^r$, $\left\{\lambda_1^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$, $\left\{\lambda_2^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$, and $\left\{\psi^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$ does not require the prior estimation of the mean vector $\boldsymbol{\mu}(\mathbf{x}_0)$. The variogram as a first difference operator filters out constants. The vector-valued mean function being approximately equal to a constant vector in the quasi-stationarity neighborhood, the matrix-valued local stationary variogram kernel estimator defined in Eq. (7) filters out the vector-valued mean function at short distances. For distances up to the radius of the quasi-stationarity neighborhood, the matrix-valued local stationary variogram kernel estimator thus estimates well the underlying local stationary multivariate spatial dependence structure of the data whose parameters are estimated from Eq. (8).

The vector-valued mean function $\boldsymbol{\mu}(\cdot)$ being considered constant within the quasi-stationarity neighborhood, its estimation at a target location $\mathbf{x}_0 \in G$ is performed through a local stationary cokriging of the mean (Wackernagel 2003). This latter is based on parameters $\left\{\mathbf{A}^{(u)}(\mathbf{x}_0)\right\}_{u=0}^r$, $\left\{\lambda_1^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$, $\left\{\lambda_2^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$, and $\left\{\psi^{(u)}(\mathbf{x}_0)\right\}_{u=1}^r$ characterizing the local stationary multivariate spatial dependence structure $\mathbf{C}_S(\cdot; \mathbf{x}_0)$ and estimated according to Eq. (8). Therefore, no model is specified for the vector-valued mean function $\boldsymbol{\mu}(\cdot)$. More specifically, we have:

$$\widehat{\boldsymbol{\mu}}(\mathbf{x}_0) = \sum_{k=1}^{n_0} \boldsymbol{\Pi}_k(\mathbf{x}_0)^T \mathbf{Z}(\mathbf{s}_k^0), \quad \forall \mathbf{x}_0 \in G, \tag{9}$$

where $\{\mathbf{s}_k^0, k = 1, \ldots, n_0\}$ are data locations belonging to the quasi-stationarity neighbourhood $\mathcal{V}_{\mathbf{x}_0}$; $\{\boldsymbol{\Pi}_k(\mathbf{x}_0), k = 1, \ldots, n_0\}$ are $p \times p$ matrices of weights which are solution of the following system of equations:

$$\begin{bmatrix} \mathbf{C}_S(\mathbf{s}_1^0 - \mathbf{s}_1^0; \mathbf{x}_0) & \ldots & \mathbf{C}_S(\mathbf{s}_1^0 - \mathbf{s}_{n_0}^0; \mathbf{x}_0) & \mathbf{I} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{C}_S(\mathbf{s}_{n_0}^0 - \mathbf{s}_1^0; \mathbf{x}_0) & \ldots & \mathbf{C}_S(\mathbf{s}_{n_0}^0 - \mathbf{s}_{n_0}^0; \mathbf{x}_0) & \mathbf{I} \\ \mathbf{I} & \ldots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Pi}_1(\mathbf{x}_0) \\ \vdots \\ \boldsymbol{\Pi}_{n_0}(\mathbf{x}_0) \\ \mathbf{M} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I} \end{bmatrix}, \tag{10}$$

with $\mathbf{M}$ being $p \times p$ matrix of Lagrange multipliers, $\mathbf{0}$ being $p \times p$ matrix of zeros, and $\mathbf{I}$ being the identity matrix of size $p \times p$. The matrix-valued local stationary covariance function $\mathbf{C}_S(\cdot; \mathbf{x}_0)$ is evaluated using the parameter estimates computed at Eq. (8).

3.3 Smoothing parameter raw estimates

Parameter estimates are required at any location of interest, especially at unsampled and sampled locations in order to perform cokriging or cosimulation. Ideally, parameters should be inferred at each data location and each location to be predicted or simulated. However, this would be very demanding in computer resources. In practice, solving the optimization problem defined in Eq. (8) for all target locations is computationally extensive. In addition, it may be redundant for close target locations due to the high correlation of their associated estimates. To reduce the computational burden, the basic idea consists in performing the parameter estimation procedure described in Sect. 3.2 only at a reduced set of locations referred to as anchor locations defined over the study domain. Using the parameter estimates at anchor locations, a kernel smoothing technique is performed to make available parameter estimates at any location of interest. To do this, parameters being supposed to be regular functions varying slowly from one end of the study domain to the other end, the Nadaraya-Watson kernel estimator (Wand and Jones 1995) with a Gaussian kernel is used, in addition to being relatively simple. However, other smoothers can be used as well (e.g., local polynomials, splines).

The Nadaraya-Watson kernel estimator of the vector-valued mean function $\boldsymbol{\mu}(\cdot)$ at a location of interest $\mathbf{x}_0 \in G$ is given by:

$$\widetilde{\boldsymbol{\mu}}(\mathbf{x}_0) = \sum_{k=1}^{m} w_k(\mathbf{x}_0)\widehat{\boldsymbol{\mu}}(\mathbf{x}_k), \quad w_k(\mathbf{x}_0) = \frac{K_\delta(\|\mathbf{x}_0 - \mathbf{x}_k\|)}{\sum_{k=1}^{m} K_\delta(\|\mathbf{x}_0 - \mathbf{x}_k\|)}, \tag{11}$$

where $K_\delta(\cdot)$ is the Gaussian kernel with smoothing parameter $\delta > 0$; $\{\widehat{\boldsymbol{\mu}}(\mathbf{x}_k)\}_{k=1,\ldots,m}$ are raw estimates of the

parameter $\boldsymbol{\mu}(\cdot)$ at anchor locations $\{\mathbf{x}_k\}_{k=1,\dots,m}$ obtained in Sect. 3.2.

Similarly to the Nadaraya-Watson kernel estimator of the vector-valued mean function $\widetilde{\boldsymbol{\mu}}(\cdot)$ described by Eq. (11), Nadaraya-Watson kernel estimators $\left\{\widetilde{\mathbf{A}}^{(u)}(\cdot)\right\}_{u=0}^{r}$, $\left\{\widetilde{\lambda}_1^{(u)}(\cdot)\right\}_{u=1}^{r}$, $\left\{\widetilde{\lambda}_2^{(u)}(\cdot)\right\}_{u=1}^{r}$ are defined. Regarding anisotropy angles $\left\{\psi^{(u)}(\cdot)\right\}_{u=1}^{r}$, they can not be interpolated as scalars because anisotropy angles θ, $\theta + \pi$, and $\theta - \pi$ give the same direction of anisotropy. Noting that $\widetilde{\mu}_i(\mathbf{x}_0)$ a component of $\widetilde{\boldsymbol{\mu}}(\mathbf{x}_0)$ defined in Eq. (11) is the solution of the minimization problem

$\arg \min_{\mu_0 \in \mathbb{R}} \sum_{k=1}^{m} w_k(\mathbf{x}_0)(\mu_0 - \widehat{\mu}_i(\mathbf{x}_k))^2$, we can similarly define the Nadaraya-Watson kernel estimator of $\psi^{(u)}(\cdot)$ as follows:

$$\widetilde{\psi}^{(u)}(\mathbf{x}_0) = \arg \min_{\psi_0 \in]-\pi;\pi[} \sum_{k=1}^{m} w_k(\mathbf{x}_0) d^2(\psi_0, \widehat{\psi}^{(u)}(\mathbf{x}_k)), \quad u = 1,\dots,r,$$
(12)

where $\left\{\widehat{\psi}^{(u)}(\mathbf{x}_k)\right\}_{k=1,\dots,m}$ are the raw estimates of the parameter $\psi^{(u)}(\cdot)$ at anchor locations $\{\mathbf{x}_k\}_{k=1,\dots,m}$; $d(\psi_0, \widehat{\psi}^{(u)}(\mathbf{x}_k))$ is a distance between two angles defined by $d(\psi_0, \widehat{\psi}^{(u)}(\mathbf{x}_k)) = \min(|\psi_0 - \widehat{\psi}^{(u)}(\mathbf{x}_k)|, |\psi_0 - \widehat{\psi}^{(u)}(\mathbf{x}_k) - \pi|, |\psi_0 - \widehat{\psi}^{(u)}(\mathbf{x}_k) + \pi|)$.

Note that the estimation of parameters at an anchor location is performed independently of other anchor locations. Likewise, the kernel smoothing of parameter raw estimates at a target location is carried out independently of other target locations. Thus, if parallelization is utilized then the computational time could be reduced. Regarding the choice of anchor locations, the set of anchor locations is chosen as a grid covering the study domain. The number of anchor locations or the spacing of the anchor locations must be such that the smoothed parameters closely follow those that would be inferred directly at every target location. The number of anchor locations or the spacing of the anchor locations may depend on the complexity of the true underlying non-stationarity and especially it is a trade-off between computational efficiency and the accuracy of the estimated parameters. As we will see on the practical examples in Sect. 5, parameters inferred directly at each target location can be closely reconstructed by smoothing the values obtained at a moderate set of anchor locations.

3.4 Tuning hyper-parameters

An important aspect in the proposed estimation method is the selection of the bandwidth parameter η entering in the computation of the matrix-valued local stationary variogram kernel estimator (Eq. 7). The size of the quasi-

stationarity neighborhood is controlled by this bandwidth parameter. Another key points are the selection of the smoothing parameter δ intervening in the interpolation of the raw estimates of parameters (Eq. 11) and the choice of the number of basic local stationary univariate covariance structures $(r + 1)$ (Eq. 4). The selection of the appropriate values of the bandwidth parameter η and the smoothing parameter δ is data-driven.

Spatial dependence structure modeling and estimating are rarely goals per se but intermediate steps before the spatial prediction which is the ultimate goal. The data-driven approach consists in taking the bandwidth value that gives the best one leave-out cross-validation mean square prediction error (Wackernagel 2003):

$$CV_1(\eta) = \frac{1}{np} \sum_{i=1}^{p} \sum_{k=1}^{n} \left(Z_i(\mathbf{s}_k) - \widehat{Z}_i^{-k}(\mathbf{s}_k) \right)^2,$$
(13)

where $\widehat{Z}_i^{-k}(\mathbf{s}_k)$ is the spatial predictor computed at location $\mathbf{s}_k$ using all observations except $\{Z_i(\mathbf{s}_k)\}$. The spatial prediction method is described in Sect. 4.1.

The value of the smoothing bandwidth δ associated with the vector-valued mean function $\boldsymbol{\mu}(\cdot)$ is selected using the following cross-validation criterion (Wand and Jones 1995):

$$CV_2(\delta) = \frac{1}{m} \sum_{l=1}^{m} \left(\frac{\widehat{\boldsymbol{\mu}}(\mathbf{x}_l) - \widetilde{\boldsymbol{\mu}}(\mathbf{x}_l)}{1 - w_l(\mathbf{x}_l)} \right)^2,$$
(14)

where $\{\widehat{\boldsymbol{\mu}}(\mathbf{x}_l)\}_{l=1,\dots,m}$ and $\{\widetilde{\boldsymbol{\mu}}(\mathbf{x}_l)\}_{l=1,\dots,m}$ are respectively the raw and smoothed estimates of the vector-valued mean function $\boldsymbol{\mu}(\cdot)$ at anchor locations $\{\mathbf{x}_k\}_{l=1,\dots,m}$. Theoretically, smoothing bandwidths associated with each parameter may be different. In practice, choosing the same smoothing bandwidth for all parameters in order to reduce the computational burden, makes little difference in terms of prediction performance.

In practice, the classical stationary linear coregionalization model is applied by usually selecting two or three basic stationary univariate correlation structures representing different scales of variation (Chilés and Delfiner 2012): a first structure modeling a discontinuity at the origin (the so-called nugget effect); a second structure describing short range (small scale) variation; a third structure of the same type as the previous for accounting long range (large scale) variation. In the local stationarity setting, accounting for large scale variation is irrelevant. Thus for estimating model parameters, using two basic local stationary univariate covariance structures (i.e., $r = 1$ or $r = 2$ depending whether the nugget effect is accounted or not) in Eq. (6) should be sufficient to capture the local spatial variability.

4 Multivariate spatial prediction

The main goals of modeling and estimating the multivariate spatial dependence structure of the data are the prediction and the simulation of variables at target locations. In this section, a description of cokriging and conditional cosimulation in this non-stationary setting is given.

4.1 Cokriging

The goal is to predict the vector-valued random field $\mathbf{Z}(\cdot)$ at a target location $\mathbf{s}_0 \in G$ based on data $[\mathbf{Z}(\mathbf{s}_1), \ldots, \mathbf{Z}(\mathbf{s}_n)]^T$. The point predictor for the unknown value of the vector-valued random field $\mathbf{Z}(\cdot)$ at a location of interest $\mathbf{s}_0 \in G$ is given by the simple cokriging estimator:

$$\widehat{\mathbf{Z}}(\mathbf{s}_0) = \boldsymbol{\mu}(\mathbf{s}_0) + \sum_{k=1}^{n} \boldsymbol{\Pi}_k(\mathbf{s}_0)^T [\mathbf{Z}(\mathbf{s}_k) - \boldsymbol{\mu}(\mathbf{s}_k)], \quad (15)$$

where $p \times p$ matrices of weights $\{\boldsymbol{\Pi}_k(\mathbf{s}_0), k = 1, \ldots, n\}$ are found from the following simple cokriging system:

$$\begin{bmatrix} \mathbf{C}(\mathbf{s}_1, \mathbf{s}_1) & \ldots & \mathbf{C}(\mathbf{s}_1, \mathbf{s}_n) \\ \vdots & \ddots & \vdots \\ \mathbf{C}(\mathbf{s}_n, \mathbf{s}_1) & \ldots & \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Pi}_1(\mathbf{s}_0) \\ \vdots \\ \boldsymbol{\Pi}_n(\mathbf{s}_0) \end{bmatrix} = \begin{bmatrix} \mathbf{C}(\mathbf{s}_1, \mathbf{s}_0) \\ \vdots \\ \mathbf{C}(\mathbf{s}_n, \mathbf{s}_0) \end{bmatrix}. \quad (16)$$

The vector-valued mean function $\boldsymbol{\mu}(\cdot)$ and the matrix-valued covariance function $\mathbf{C}(\cdot, \cdot)$ are evaluated using the parameter estimates found at Sect. 3. The variance-covariance matrix of the prediction errors correspond to $\mathbf{Q}(\mathbf{s}_0) = \mathbf{C}(\mathbf{s}_0, \mathbf{s}_0) - \sum_{k=1}^{n} \boldsymbol{\Pi}_k(\mathbf{s}_0)^T \mathbf{C}(\mathbf{s}_k, \mathbf{s}_0)$.

4.2 Conditional cosimulation

The aim is to simulate the vector-valued random field $\mathbf{Z}(\cdot)$ assumed to be Gaussian, at a large number of locations such that the realization honors the data $[\mathbf{Z}(\mathbf{s}_1), \ldots, \mathbf{Z}(\mathbf{s}_n)]^T$. This can be done from a non-conditional simulation of the vector-valued random field $\mathbf{Z}(\cdot)$ following by the method of conditioning by cokriging (Lantuejoul 2002).

The vector-valued random field $\mathbf{Z}(\cdot)$ being second order non-stationary, traditional non-conditional simulation techniques developed in the stationary framework can not be used. The representation of the vector-valued random field $\mathbf{Z}(\cdot)$ in Eq. (1) suggests that the non-conditional simulation of $\mathbf{Z}(\cdot)$ involves: the non-conditional simulation of p standard Gaussian white noises, and the non-conditional simulation of $p \times r$ independent Gaussian univariate random fields with zero mean, unit variance and closed-form non-stationary correlation function. The non-conditional simulation of these latter can be performed

efficiently either using the propagative version of the Gibbs sampler proposed by Lantuejoul and Desassis (2012) or using the spectral method proposed by Emery and Arroyo (2017).

5 Practical examples

In this section, the proposed non-stationary multivariate spatial modeling approach is illustrated using two real bivariate data examples: geophysical and geochemical bivariate data. Prediction performance comparison is carried out with the traditional stationary multivariate spatial modeling approach using some well-known predictive scores (Gneiting and Raftery 2007; Zhang and Wang 2010; Chilés and Delfiner 2012): mean absolute error (MAE), root mean square error (RMSE), logarithmic score (LogS) and continued ranked probability score (CRPS). For all these scores, smaller values indicate better predictions.

5.1 Geophysical bivariate data example

The first real bivariate data example comes from a gamma radiometric soil survey conducted in the region of the Hunter Valley, NSW, Australia (Stockmann et al. 2012). Variables of interest are gamma-ray emission from Potassium (K, cps) and Thorium (Th, cps) occurring naturally in the soil. We have a training dataset containing 537 observations for model estimation and a validation dataset of 1000 observations for prediction performance assessment. Figure 1 shows spatial plots of the variables in the training dataset and clearly reveals the high correlation between the two variables; hence the need for a multivariate spatial process model. There is no apparent global geometric anisotropy in the data. This was confirmed when computing directional experimental variograms in the stationary framework.

Figure 2 displays some parameter raw estimates (mean, standard deviation, cross-correlation, and geometric anisotropy) at anchor locations according to the estimation procedure described in Sect. 3. Maps of different parameter raw estimates at anchor locations indicate the presence of the non-stationarity in the data. Means of Potassium (K) and Thorium (Th) vary substantially across space as well as their standard deviations. The cross-correlation coefficient between Potassium (K) and Thorium (Th) varies spatially taking values ranging from approximately 0.4 to 0.9. The locally varying geometric anisotropy depicted by ellipses is quite visible. Such directional effects are also quite apparent in the data. Parameter raw estimates at anchor locations are obtained using two basic local univariate stationary structure models (nugget effect model and exponential model with geometric anisotropy).

Fig. 1 a Soil gamma-radiometric potassium (K) training data; **b** soil gamma-radiometric thorium (Th) training data; **c** scatter plot of K against Th in the training data
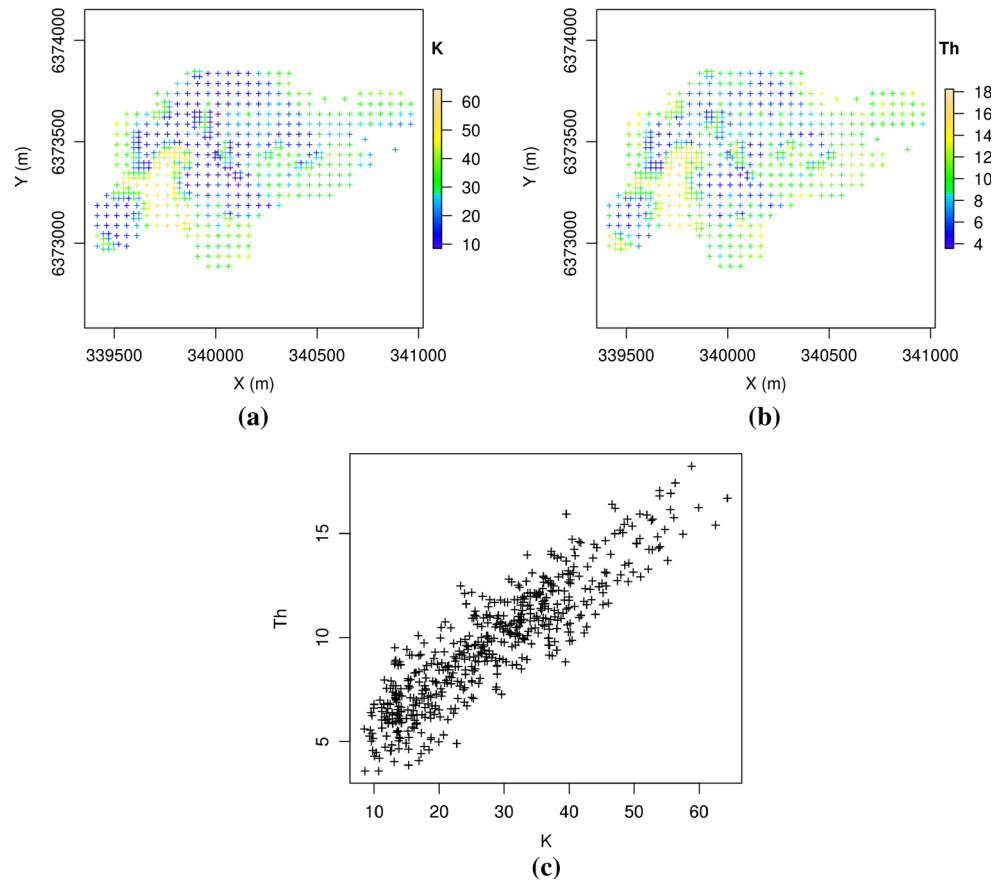

(a)

(b)

(c)

Figure 3 presents the maps of smoothed parameter estimates over the study domain (mean, standard deviation, cross-correlation, geometric anisotropy ratio, and geometric anisotropy direction). Following the hyper-parameters selection procedure described in Sect. 3.4, the optimal bandwidth associated with the matrix-valued local stationary variogram kernel estimator is $\eta = 141$ m. Regarding the smoothing bandwidth related to the interpolation of parameter raw estimates over study domain, its optimal value is $\delta = 43$ m. Figure 4 shows maps of some parameters inferred directly at each target location (parameter raw estimates). It can be observed that the exhaustively inferred parameters can be closely reconstructed by smoothing the values obtained at anchor locations. The number of anchor locations is 475 while the number of target locations is 15,492.

A visualization of direct and cross covariance functions at some reference locations through level contours under the estimated non-stationary and stationary linear coregionalization models is given in Fig. 5. This plot illustrates the fact that the non-stationary linear coregionalization model allows the spatial dependence structure to change from one location to another, while the stationary linear coregionalization model estimates a constant spatial dependence structure. The stationary linear

coregionalization model is estimated using two basic univariate isotropic stationary structure models (nugget effect and exponential models). It can be observed that the stationary model provides an elliptical correlation pattern, while the non-stationary model produces non-elliptical correlation pattern.

Table 1 reports the predictive scores computed on a validation dataset (1000 observations) for non-stationary and stationary linear coregionalization models. It emerges that the non-stationary linear coregionalization model outperforms the stationary linear coregionalization model in terms of prediction accuracy and prediction uncertainty accuracy. The cost of non-using the non-stationary modeling approach is substantial. For example, the non-stationary modeling approach reduces the RMSE by 23% and the CRPS by 40% according to the stationary modeling approach.

Figure 6 shows predictions and prediction standard deviations based on estimated non-stationary and stationary linear coregionalization models. The overall look of cokriging maps associated with each model differs notably, in particular the cokriging standard deviation maps. Under the non-stationary model, cokriging standard deviations tend to be low in regions of low variability, while they tend to be high in regions of high variability. Thus, cokriging
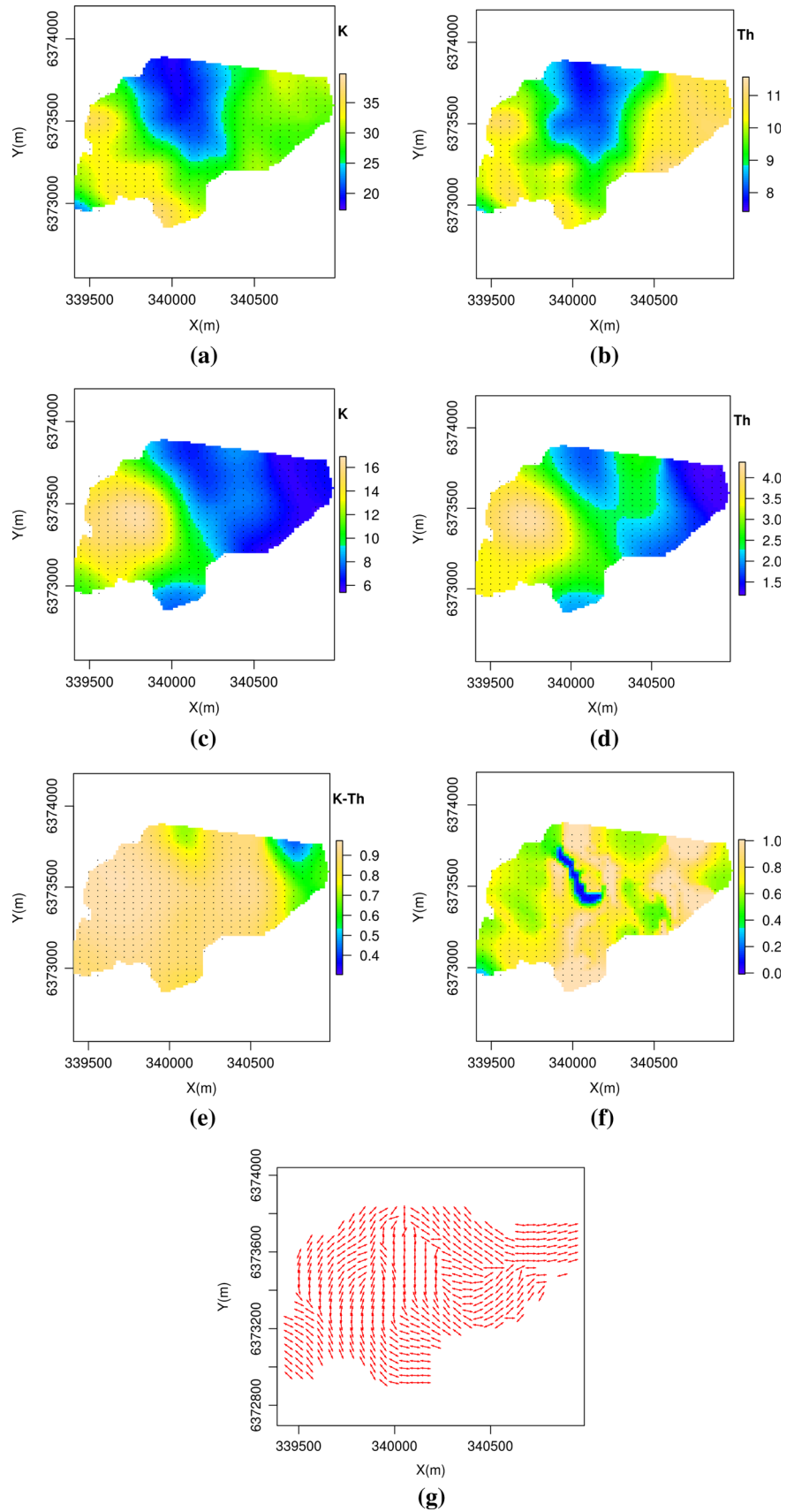
Fig. 2 Parameter raw estimates at anchor locations. **a** mean of K; **b** mean of Th; **c** standard deviation of K; **d** standard deviation of Th; **e** cross-correlation between K and Th; **f** geometric anisotropy
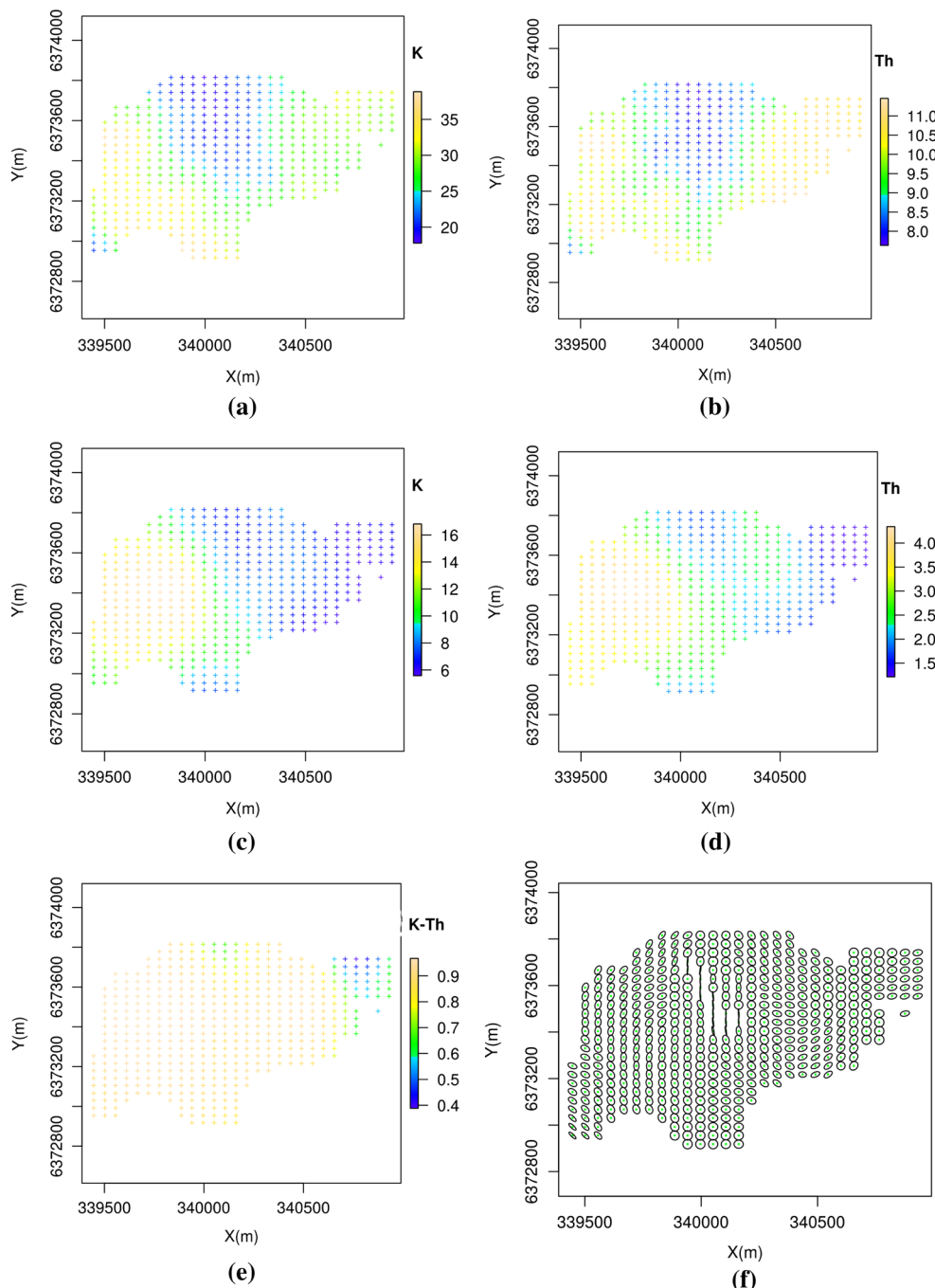
standard deviations reflect not only the samples configuration and availability around target locations, but also the local variability. On the other hand, cokriging standard deviations related to the stationary modeling approach shows slight differences throughout the study domain, due to the sampling intensity. Such a pattern was expected as the stationary modeling approach assumes the same spatial dependence structure over the region of interest.

Gaussian conditional simulations based on estimated non-stationary and stationary linear coregionalization

models are presented in Fig. 7. Both simulations are generated from the same random number seed to facilitate comparisons. As one can see, conditional simulations under the non-stationary model differ from one under the stationary model, especially in terms of anisotropy.

5.2 Geochemical bivariate data example

The second real bivariate data example is the well-known Meuse data (Burrough et al. 1998). Data set consists of 155

Fig. 3 Smoothed parameter estimates over the region of interest. **a** Mean of K; **b** mean of Th; **c** standard deviation of K; **d** standard deviation of Th; **e** cross-correlation between K and Th; **f** geometric anisotropy ratio; **g** geometric anisotropy direction
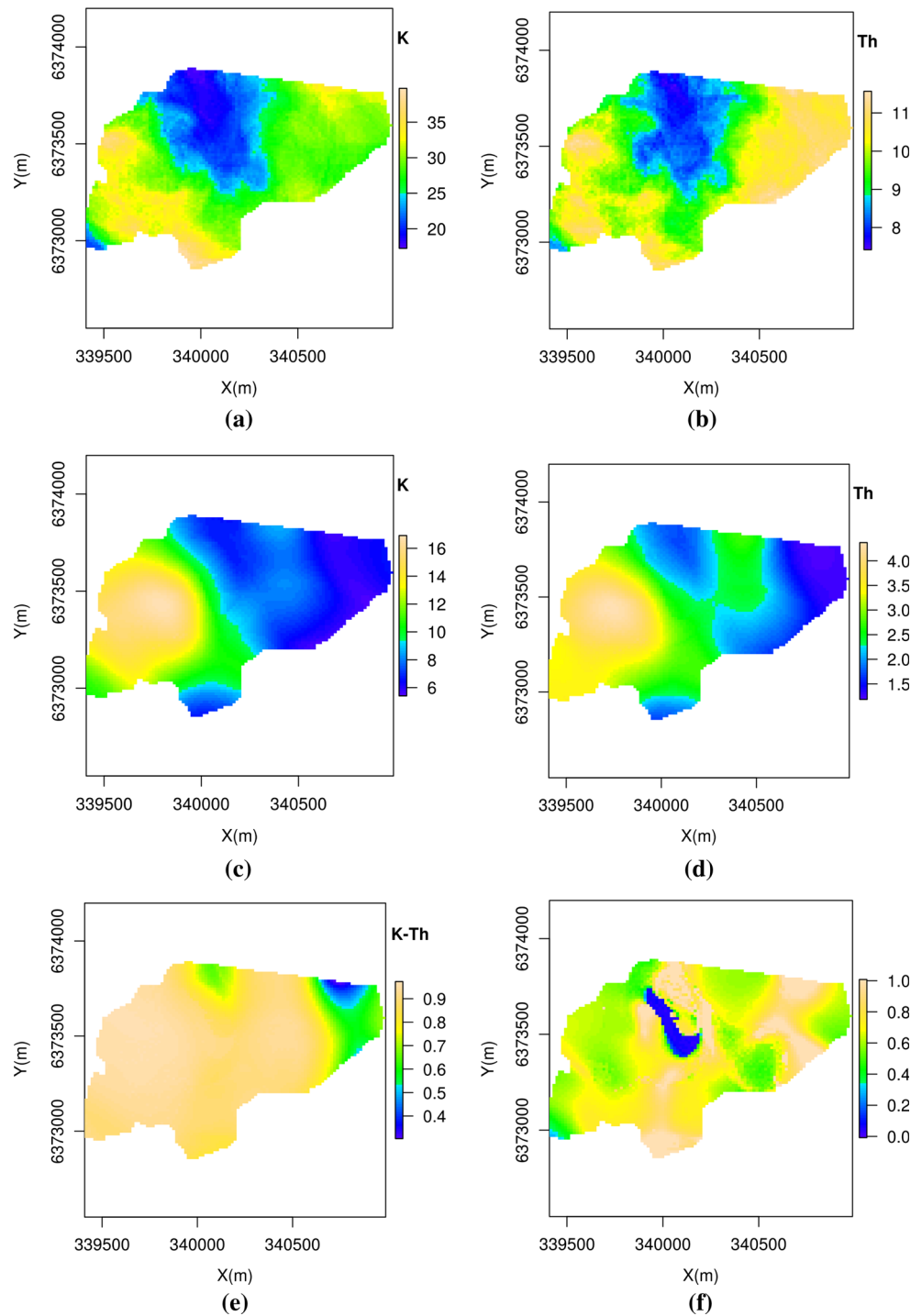
Fig. 4 Parameter raw estimates over the study domain. **a** Mean of K; **b** mean of Th; **c** standard deviation of K; **d** standard deviation of Th; **e** cross-correlation between K and Th; **f** geometric anisotropy ratio

samples of top soil heavy metal concentrations (ppm), collected in a flood plain of the river Meuse, near the village Stein, Netherlands. We focus on two variables Cd and Zn. Both variables are positively skewed, hence a log transformation was applied to each. Spatial plots of transformed variables are depicted in Fig. 8 where one can observe a high correlation between the two transformed variables. The representation of data suggests the presence

of a global zonal anisotropy direction in the data. This is confirmed when computing directional experimental variograms in the stationary setting. The orientation of maximum continuity is along the direction South/West - North/East.

Maps of different parameter raw estimates at anchor locations are shown in Fig. 9, revealing the non-stationarity in the data. Means of log Cd and log Zn vary substantially

Fig. 5 **a**, **b**, **c**, **d** Direct and **e**, **f** cross covariance function level contours at some reference locations. **a** Non-stationary model; **b** stationary model; **c** non-stationary model; **d** stationary model; **e** non-stationary model; **f** stationary model

(a)

(b)

(c)

(d)

(e)

(f)
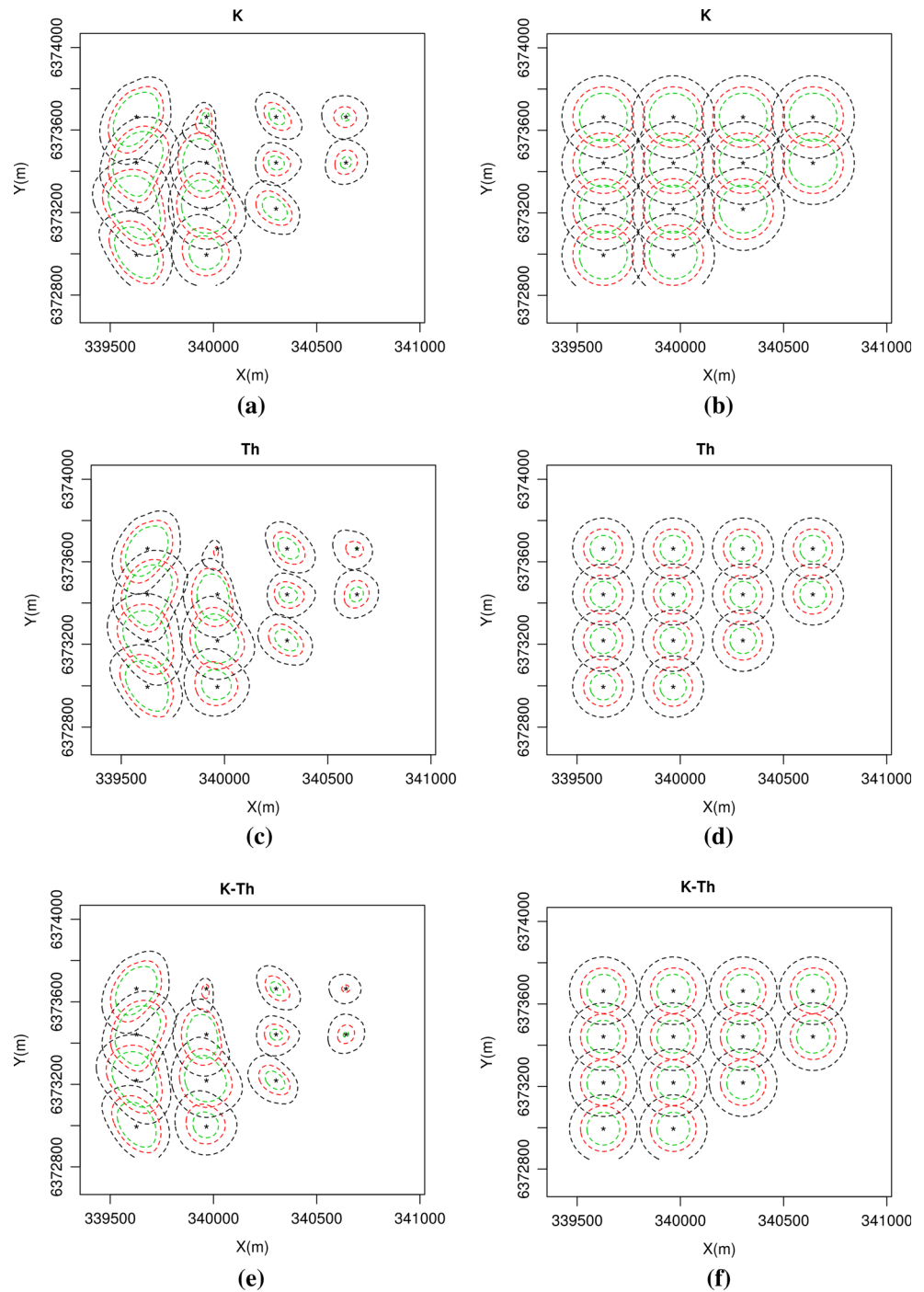
throughout the study domain as well as their standard deviations. The strength of correlation between the two variables evolves across the study domain, ranging from 0.86 to 0.93. The estimated geometric anisotropy function at anchor locations reveals a locally varying geometric anisotropy including the main direction of continuity discovered by the stationary modeling approach. Note that global anisotropy cannot describe local directional features of a spatial surface, only global ones. Parameter raw

Table 1 External validation scores with 1000 hold-out observations

Criteria	Non-stationary model		Stationary model	
	K	Th	K	Th
MAE	2.438	0.820	3.283	1.306
RMSE	3.548	1.098	4.309	1.612
LogS	5.427	2.936	6.266	4.058
CRPS	3.896	1.165	5.249	1.837

Fig. 6 **a**, **b**, **c**, **d** Predictions and **e**, **f**, **g**, **h** prediction standard deviations. **a** Non-stationary model; **b** non-stationary model; **c** stationary model; **d** stationary model; **e** non-stationary model; **f** non-stationary model; **g** stationary model; **h** stationary model
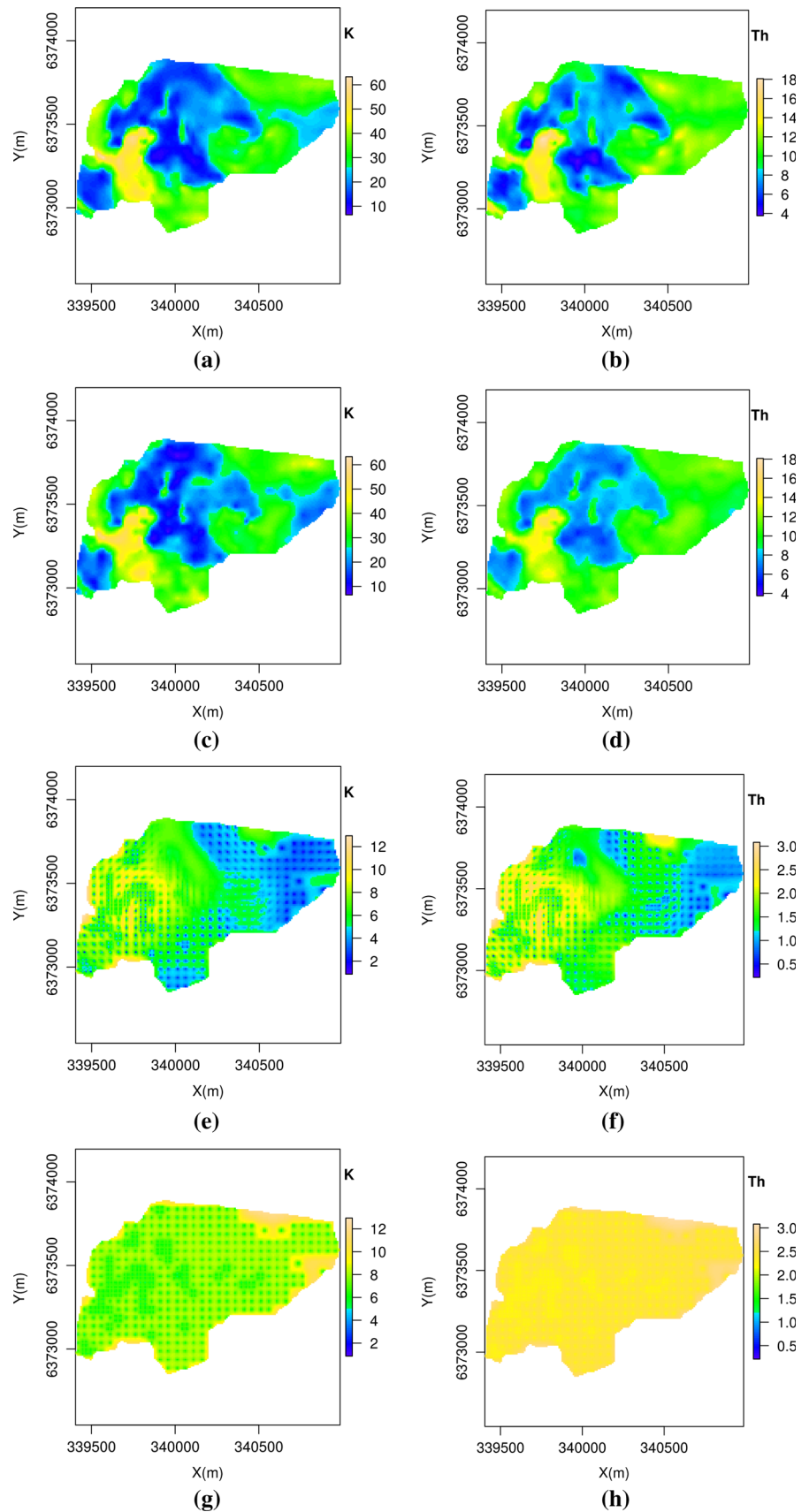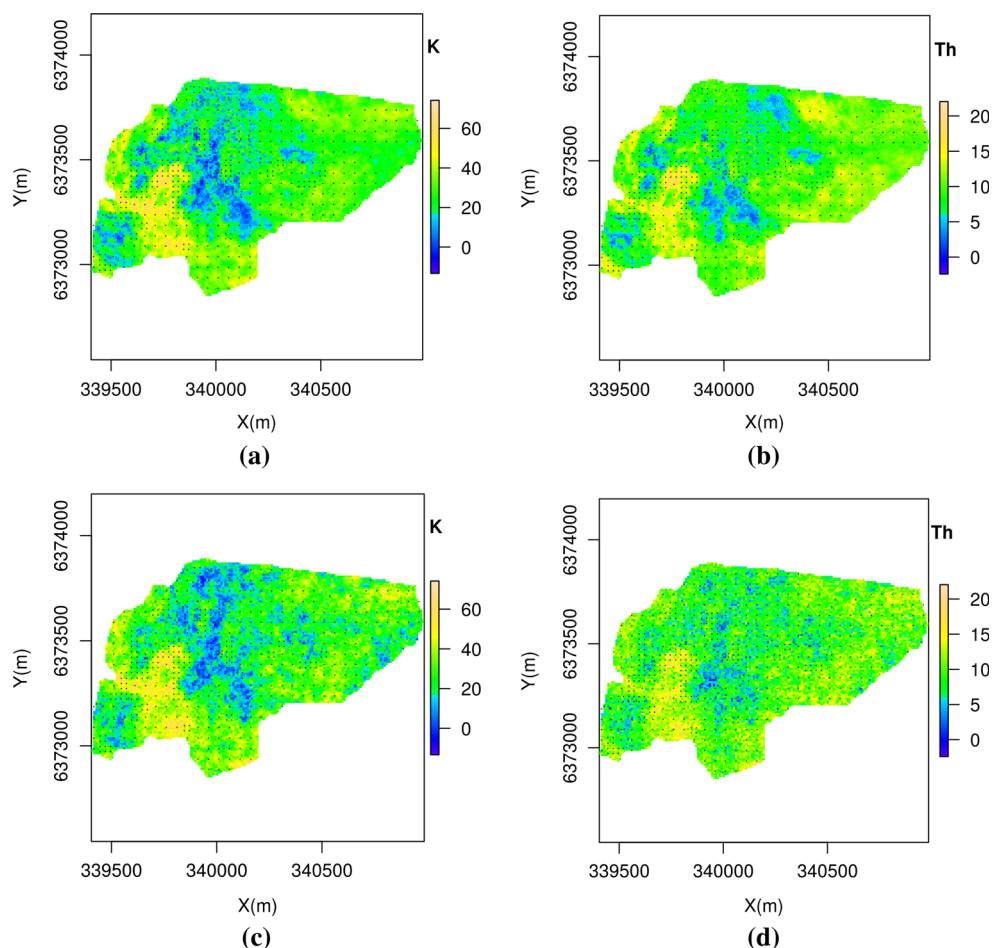
Fig. 7 Conditional simulations based on **a**, **b** the non-stationary model and **c**, **d** the stationary model. **a** Non-stationary model; **b** non-stationary model; **c** stationary model; **d** stationary model

(a)

(b)

(c)

(d)

estimates at anchor locations are obtained using two basic local univariate stationary structure models (nugget effect model and exponential model with geometric anisotropy).

Maps of smoothed parameter estimates over the study domain (mean, standard deviation, cross-correlation, geometric anisotropy ratio, and geometric anisotropy direction) are given in Fig. 10. According to the bandwidth selection approach described in Sect. 3.4, the optimal bandwidth associated with the matrix-valued local stationary variogram kernel estimator is $\eta = 767$ m. The optimal value of the smoothing bandwidth related to the interpolation of parameter raw estimates over study domain corresponds to $\delta = 56$ m. Maps of some parameters inferred directly at each target location (parameter raw estimates) are given in Fig. 11. It can be observed that the completely inferred parameters can be closely recovered by interpolating the values obtained at anchor locations. The number of anchor locations is 505 while the number of target locations is 13,500.

A representation of the estimated non-stationary and stationary direct and cross covariance functions at some reference locations via level contours is given in Fig. 12. One can see how the non-stationary spatial dependence

structure changes the shape from one location to another compared to the stationary one. Correlation patterns provided by the two models are quite different. The stationary linear coregionalization model is estimated using three basic local univariate stationary structure models (nugget effect model, isotropic exponential model, and exponential model with a zonal anisotropy along the direction South/ West–North/East).

To assess the predictive ability of the proposed non-stationary modeling approach in this dataset, a pseudo cross-validation is considered instead of an external validation due to the relatively small size of the dataset. The pseudo cross-validation consists in leaving out a randomly selected 10% of locations (15 locations), and cokrige the remaining bivariate observations to these held-out locations, without re-estimate the model. This procedure is repeated 1000 times. Table 2 contains the averaged predictive scores and standard deviations from this procedure. The analysis of the table shows that the proposed non-stationary linear coregionalization model performs better than the stationary one in terms of prediction accuracy and prediction uncertainty accuracy. The global improvement is about 4% in terms of RMSE and 6% in terms of CRPS

Fig. 8 a log Cd concentration data; **b** log Zn concentration data; **c** scatter plot of log Cd against log Zn
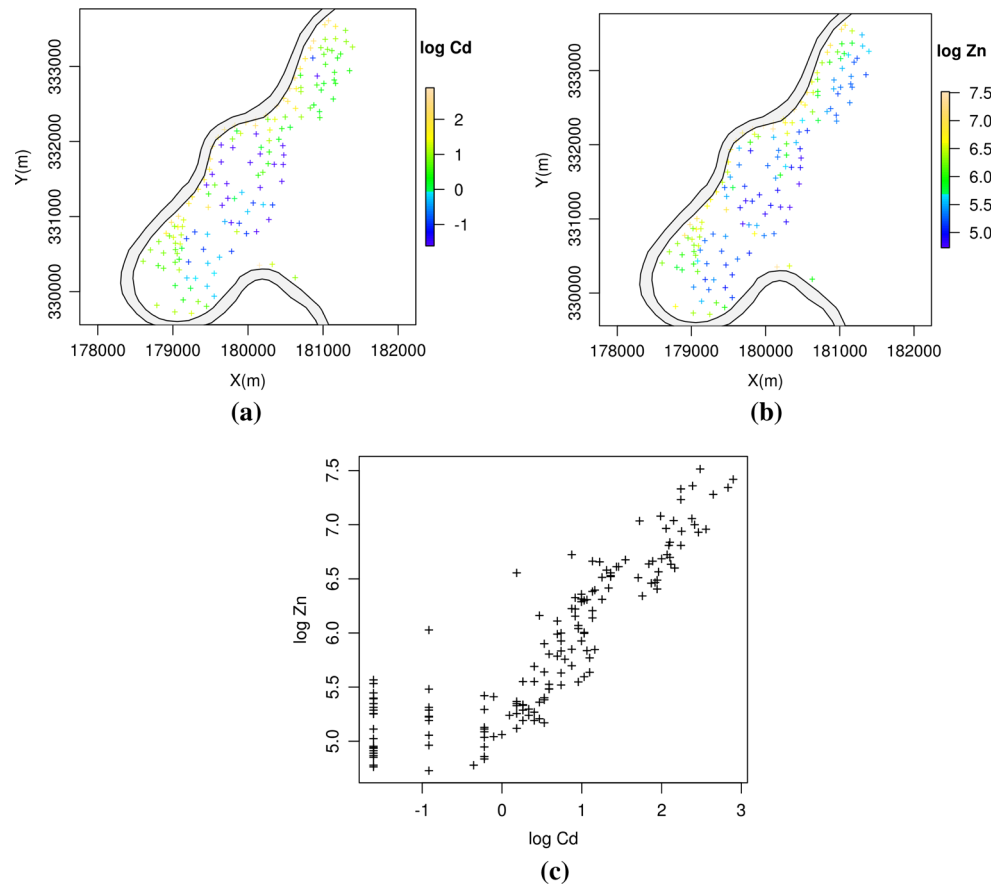
with respect to the stationary linear coregionalization model.

The cokriging results for the estimated non-stationary and stationary linear coregionalization models are shown in Fig. 13. The general appearance of the maps of cokriged values associated with each model differs. Moreover, the non-stationary and stationary linear coregionalization models differ also in describing the spatial uncertainty associated with the predictions. One can see that under the non-stationary multivariate spatial modeling approach, prediction standard deviations reflect not only the samples configuration and availability around estimates, but also the local variability. However, cokriging standard deviation maps under the stationary multivariate spatial modeling approach shows slight differences in the prediction standard deviations over the study domain, which were dependent on the sampling intensity. As previously mentioned in the first example in Sect. 5.1 such pattern was expected for a stationary modeling approach because it is based on identical global structural parameters throughout the study domain, while the non-stationary approach adapts to locally varying structure of data.

Figure 14 shows Gaussian conditional simulations performed under the estimated non-stationary and stationary linear coregionalization models. In order to facilitate

comparisons, simulations are generated from the same random number seed. It appears that conditional simulations based on the non-stationary linear coregionalization model differ from one based on the stationary linear coregionalization model, especially regarding the anisotropy.

6 Discussion and conclusion

In this paper, a fully non-stationary linear coregionalization model is introduced as an extension of the conventional stationary linear coregionalization model to handle the lack of stationarity. The proposed non-stationary linear coregionalization model is more flexible than the spatially varying linear coregionalization model of Gelfand et al. (2004). It lets coregionalization matrices to change with space and basic univariate correlation functions belonging to the class of closed-form non-stationary univariate correlation functions with locally varying geometric anisotropy proposed by Paciorek and Schervish (2006). Thus, some varying spatial features of the coregionalization as locally varying geometric anisotropy can be captured.

Fig. 9 Parameter raw estimates at anchor locations. **a** Mean of log Cd; **b** mean of log Zn; **c** standard deviation of log Cd; **d** standard deviation of log Zn; **e** cross-correlation between log Cd and log Zn; **f** geometric anisotropy
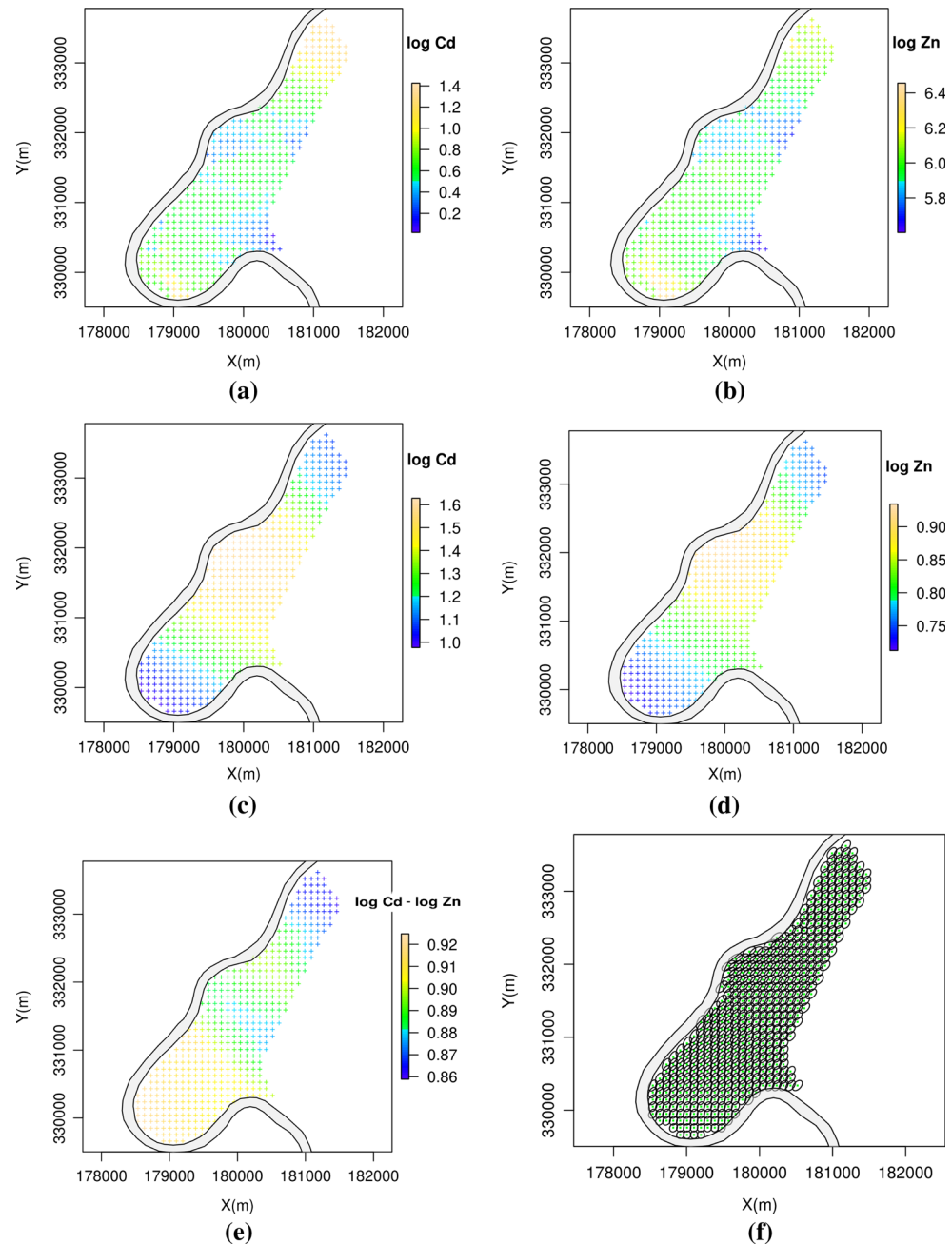
The proposed estimation framework offers an integrated treatment of all aspects of non-stationarity: mean, variance, and spatial continuity. It relies on the mild hypothesis of local stationarity and does not impose any distributional assumptions except the existence of the two first moments. It does not require any calculation of inverse or determinant of matrices, and it is parallelizable. The proposed non-stationary multivariate spatial modeling approach has the advantage to retain ease of interpretation as well computational tractability. It allows using tools already developed in the stationary multivariate framework as well as in the non-stationary univariate setting. The advantage in terms of

prediction has been demonstrated on two real bivariate data examples. Beyond the spatial prediction, it can serve as an exploratory tool for the non-stationarity.

The proposed non-stationary multivariate spatial modeling approach relies on the matrix-valued local stationary variogram kernel estimator defined under the local stationarity setting. This kernel-type estimator can also be used under the intrinsically locally stationary setup where the matrix-valued local stationary variogram is unbounded. Thus, it is important to verify that the matrix-valued local stationary variogram kernel estimator used to describe the local multivariate spatial variation presents a sill or is

Fig. 10 Smoothed parameter estimates over the study domain. **a** Mean of log Cd; **b** mean of log Zn; **c** standard deviation of log Cd; **d** standard deviation of log Zn; **e** cross-correlation between log Cd and log Zn; **f** geometric anisotropy ratio; **g** geometric anisotropy direction
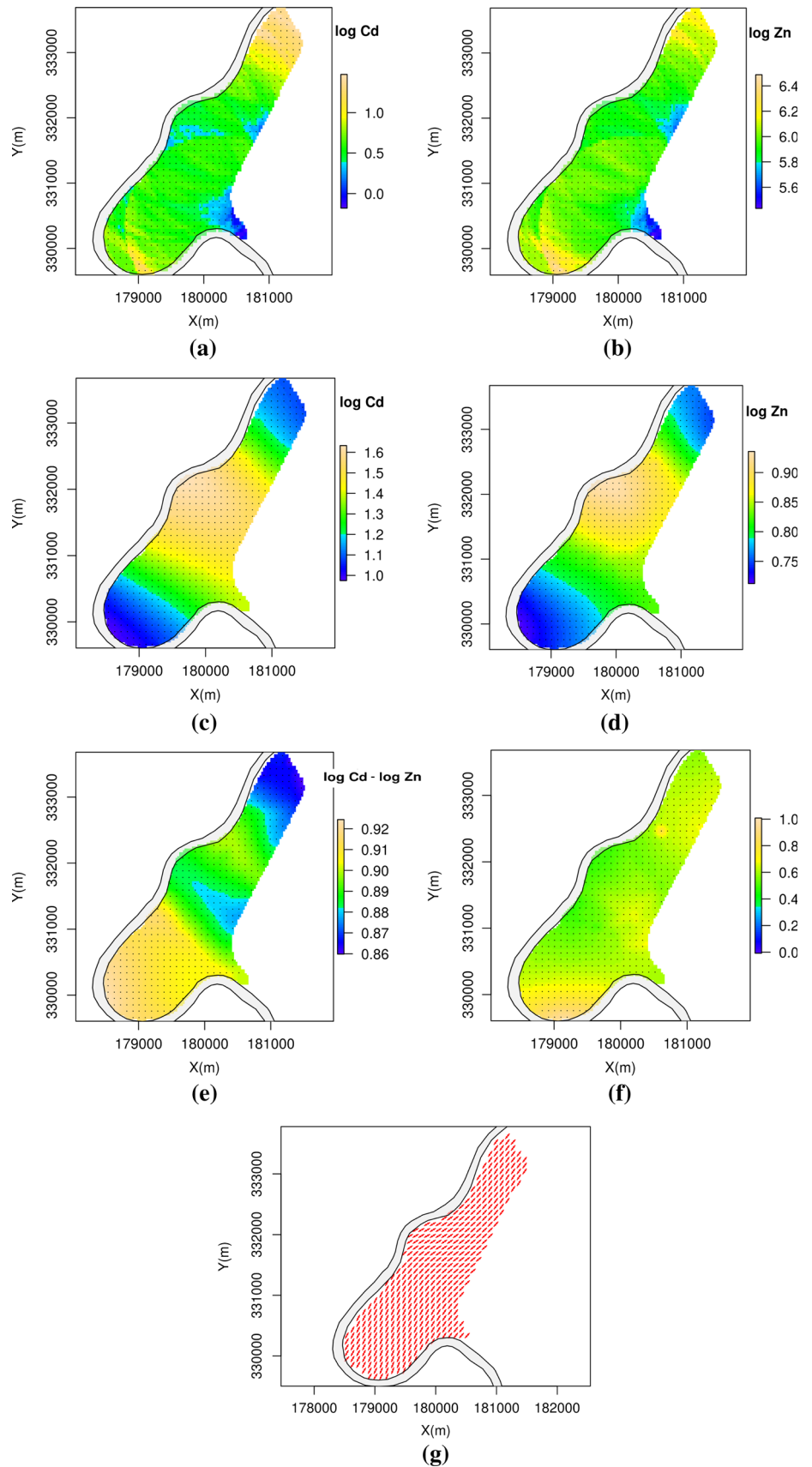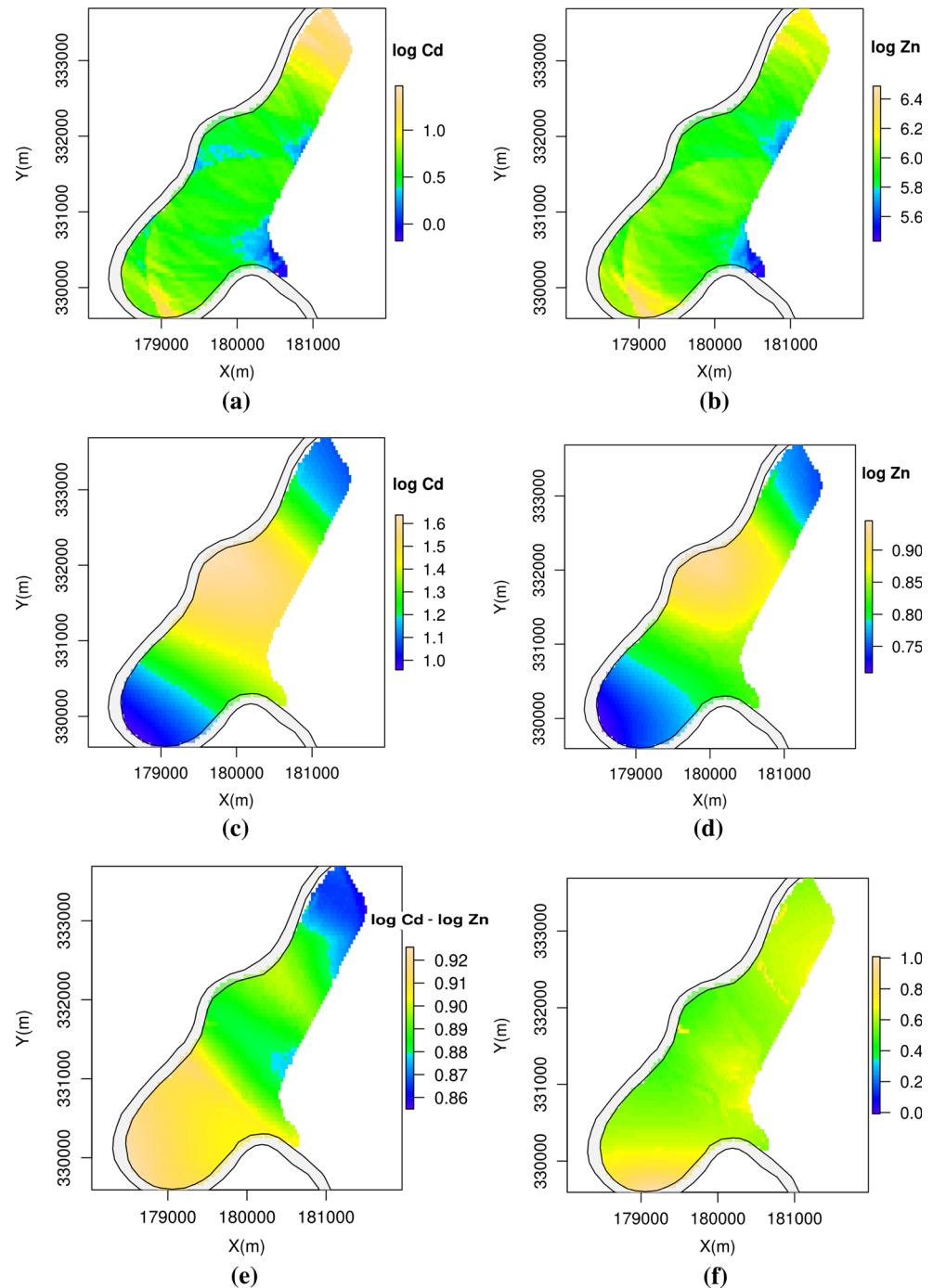
Fig. 11 Parameter raw estimates over the study domain. **a** Mean of log Cd; **b** mean of log Zn; **c** standard deviation of log Cd; **d** standard deviation of log Zn; **e** cross-correlation between log Cd and log Zn; **f** geometric anisotropy ratio

bounded before deciding to use the proposed non-stationary multivariate spatial modeling approach. This can be accomplished by visualizing the matrix-valued local stationary variogram kernel estimator at anchor locations. To better adapt to the variable sampling density in the study domain, it would be interesting to work with a locally adaptive kernel estimator. The basic idea is to increase the bandwidth in low sample density regions and to narrow it in highly sampled regions.

The proposed non-stationary multivariate spatial modeling can be applied to partially heterotopic datasets (some variables share some sample locations) although it was described for isotopic datasets (data are available for each variable at all sampling locations). In the particular case of entirely heterotopic datasets (variables have been measured on different sets of sample locations and have no sample locations in common), the matrix-valued local stationary variogram cannot be computed. However, in the same way

Fig. 12 **a**, **b**, **c**, **d** Direct and **e**, **f** cross covariance function level contours at some reference locations. **a** Non-stationary model; **b** stationary model; **c** non-stationary model; **d** stationary model; **e** non-stationary model; **f** stationary model
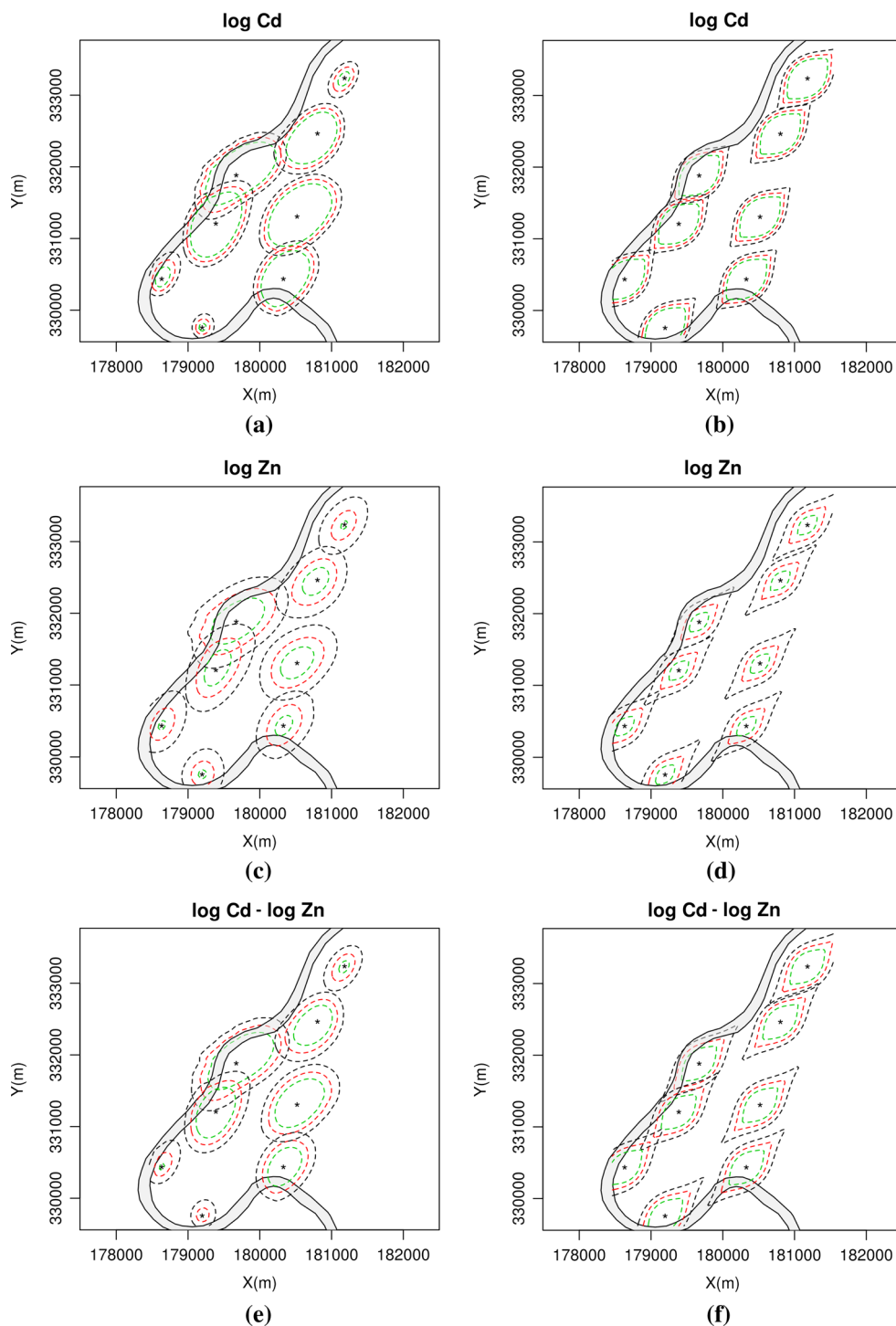

Table 2 Pseudo cross-validation scores averaged over 1000 cross-validation replications and associated standard deviations in parentheses

Criteria	Non-stationary model		Stationary model	
	log Cd	log Zn	log Cd	log Zn
MAE	0.602 (0.142)	0.289 (0.066)	0.655 (0.142)	0.300 (0.067)
RMSE	0.813 (0.191)	0.377 (0.093)	0.857 (0.197)	0.390 (0.102)
LogS	2.264 (0.512)	0.938 (0.406)	2.594 (0.559)	0.981 (0.467)
CRPS	0.754 (0.168)	0.357 (0.097)	0.858 (0.175)	0.377 (0.098)

Fig. 13 **a**, **b**, **c**, **d** Predictions and **e**, **f**, **g**, **h** prediction standard deviations. **a** Non-stationary model; **b** non-stationary model; **c** stationary model; **d** stationary model; **e** non-stationary model; **f** non-stationary model; **g** stationary model; **h** stationary model
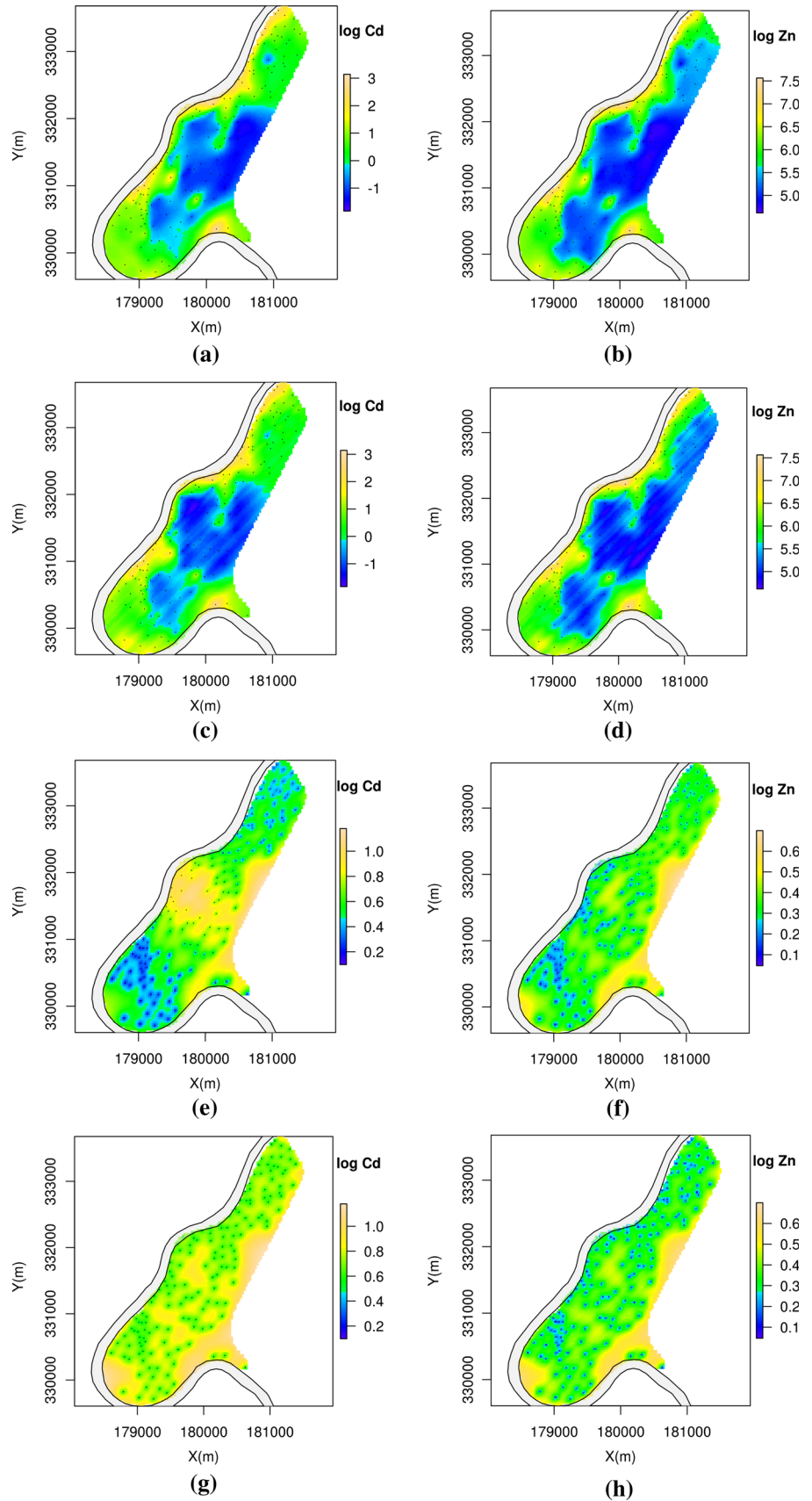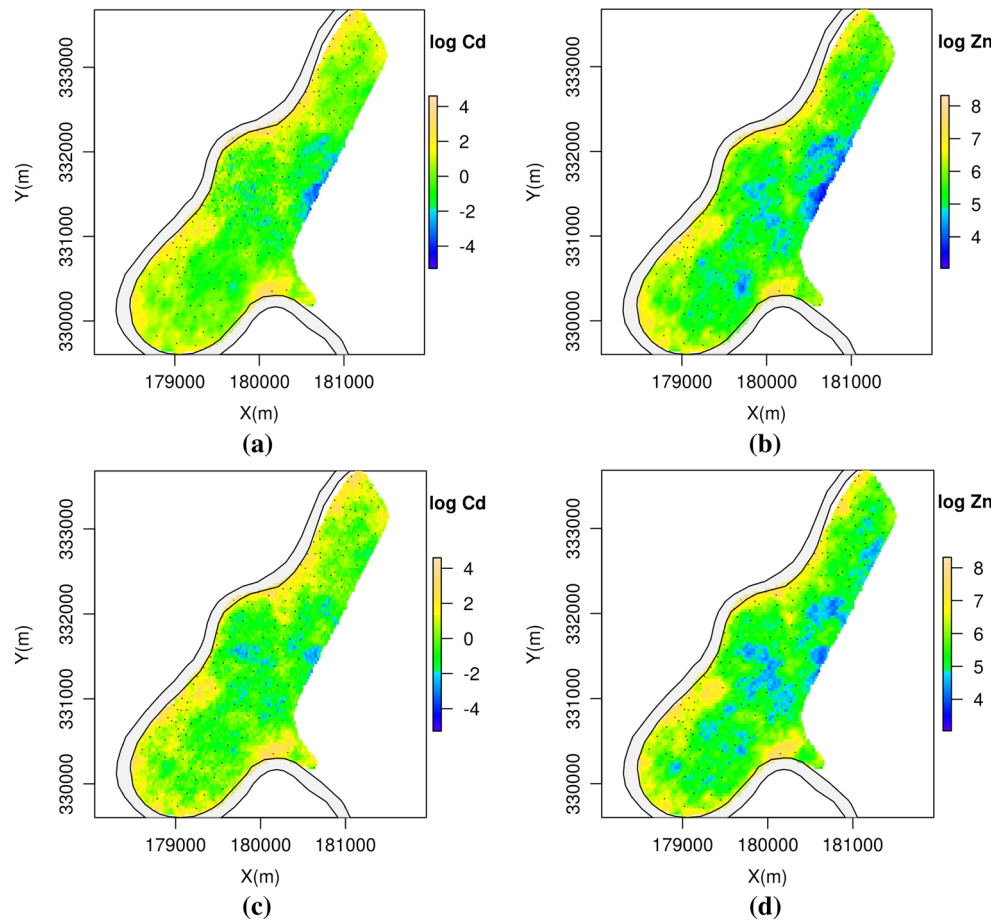
Fig. 14 Conditional simulations based on **a**, **b** the non-stationary model and **c**, **d** the stationary model. **a** Non-stationary model; **b** non-stationary model; **c** stationary model; **d** stationary model

as the matrix-valued local stationary variogram kernel estimator has been defined, one can define a non-parametric kernel estimator of the local mean and a matrix-valued local stationary covariance kernel estimator. These can be used to estimate the proposed non-stationary linear coregionalization model in the completely heterotopic case.

The proposed non-stationary multivariate spatial modeling approach which is based on the local stationarity assumption works well for smoothly varying non-stationarity. It also requires enough data to be able to capture the non-stationarity adequately as any non-stationary spatial modeling approach. As a result, it may not work well for small and sparse data or data with abrupt spatial structure changes. In these cases, it may be advisable proceeding under the stationary framework or partitioning the study domain if possible.

As a linear coregionalization model, the proposed non-stationary linear coregionalization model can not handle variables with different degrees of regularity (behavior at the origin). The smoothness is identical for all the variables, imposed by the roughest basic univariate spatial dependence structure. In such situation, a more complex non-stationary multivariate spatial model should be used as the one proposed by Kleiber and Nychka (2012).

Moreover, as a linear coregionalization model, the application of the proposed non-stationary linear coregionalization model assumes implicitly that variables under study are correlated. Thus, if variables under study are almost uncorrelated, it will be better to handle them separately.

Appendix

Proof of Eq. (2)

Under the model specified in Eq. (1) the mean of the vector-valued random field $\mathbf{Z}(\cdot)$ is straightforward to compute:

$$\mathbb{E}(\mathbf{Z}(\mathbf{x})) = \boldsymbol{\mu}(\mathbf{x}) + \sum_{u=1}^{r} \mathbf{A}^{(u)}(\mathbf{x})\mathbb{E}(\mathbf{W}^{(u)}(\mathbf{x})) + \mathbf{A}^{(0)}(\mathbf{x})\mathbb{E}(\boldsymbol{\epsilon}(\mathbf{x}))$$

$$= \boldsymbol{\mu}(\mathbf{x}), \quad \text{because } \mathbb{E}(\mathbf{W}^{(u)}(\mathbf{x})) = \mathbf{0} \quad \text{and } \mathbb{E}(\boldsymbol{\epsilon}(\mathbf{x})) = \mathbf{0}.$$

Proof of Eq. (3)

The model specified in Eq. (1) can be rewritten in the univariate form as follows:

$$Z_i(\mathbf{x}) = \mu_i(\mathbf{x}) + \sum_{u=0}^{r} \sum_{m=1}^{p} a_{im}^{(u)}(\mathbf{x}) W_m^{(u)}(\mathbf{x}), \tag{17}$$
$$\text{with } W_m^{(0)}(\mathbf{x}) = \epsilon_m(\mathbf{x}), \quad i = 1, \dots, p.$$

The covariance between $Z_i(\mathbf{x})$ and $Z_j(\mathbf{y})$ is:

$$C_{ij}(\mathbf{x}, \mathbf{y})$$
$$= \mathbb{C}\mathrm{ov}(Z_i(\mathbf{x}), Z_j(\mathbf{y}))$$
$$= \mathbb{C}\mathrm{ov}\left(\sum_{u=0}^{r} \sum_{m=1}^{p} a_{im}^{(u)}(\mathbf{x}) W_m^{(u)}(\mathbf{x}), \sum_{u'=0}^{r} \sum_{m'=1}^{p} a_{jm'}^{(u')}(\mathbf{y}) W_{m'}^{(u')}(\mathbf{y}) \right)$$
$$= \sum_{u=0}^{r} \sum_{m=1}^{p} \sum_{u'=0}^{r} \sum_{m'=1}^{p} a_{im}^{(u)}(\mathbf{x}) a_{jm'}^{(u')}(\mathbf{y}) \mathbb{C}\mathrm{ov}\left(W_m^{(u)}(\mathbf{x}), W_{m'}^{(u')}(\mathbf{y}) \right)$$
$$= \sum_{u=0}^{r} \sum_{m=1}^{p} a_{im}^{(u)}(\mathbf{x}) a_{jm}^{(u)}(\mathbf{y}) \mathbb{C}\mathrm{ov}\left(W_m^{(u)}(\mathbf{x}), W_m^{(u)}(\mathbf{y}) \right),$$
$$\text{because } W_m^{(u)}(\cdot) \perp W_{m'}^{(u')}(\cdot)$$
$$= \sum_{u=0}^{r} \sum_{m=1}^{p} a_{im}^{(u)}(\mathbf{x}) a_{jm}^{(u)}(\mathbf{y}) \rho_{NS}^{(u)}(\mathbf{x}, \mathbf{y}),$$
$$\text{with } \rho_{NS}^{(0)}(\mathbf{x}, \mathbf{y}) = \rho_S^{(0)}(\|\mathbf{x} - \mathbf{y}\|).$$

Hence $\quad \mathbf{C}(\mathbf{x}, \mathbf{y}) = \sum_{u=1}^{r} \mathbf{A}^{(u)}(\mathbf{x}) \mathbf{A}^{(u)}(\mathbf{y})^T \rho_{NS}^{(u)}(\mathbf{x}, \mathbf{y}) + \mathbf{A}^{(0)}(\mathbf{x}) \mathbf{A}^{(0)}(\mathbf{y})^T \rho_S^{(0)}(\|\mathbf{x} - \mathbf{y}\|)$.

The matrix-valued covariance function $\mathbf{C}(\cdot, \cdot)$ represents a valid covariance structure if and only if $\mathbf{C}(\cdot, \cdot)$ is non-negative definite in the sense that, for any n-dimensional finite system of p-dimensional vectors $\{\boldsymbol{\beta}_k\}_{k=1}^n$, and for any n-dimensional collection of spatial locations $\{\mathbf{x}_k\}_{k=1}^n$, and any integer n, we have $\sum_{i,j=1}^{p} \sum_{k,l=1}^{n} \beta_{ik} C_{ij}(\mathbf{x}_k, \mathbf{x}_l) \beta_{jl} \geq 0$.

$$\sum_{i,j=1}^{p} \sum_{k,l=1}^{n} \beta_{ik} C_{ij}(\mathbf{x}_k, \mathbf{x}_l) \beta_{jl}$$
$$= \sum_{i,j=1}^{p} \sum_{k,l=1}^{n} \sum_{u=0}^{r} \sum_{m=1}^{p} \beta_{ik} a_{im}^{(u)}(\mathbf{x}_k) a_{jm}^{(u)}(\mathbf{x}_l) \beta_{jl} \rho_{NS}^{(u)}(\mathbf{x}_k, \mathbf{x}_l)$$
$$= \sum_{u=0}^{r} \sum_{m=1}^{p} \sum_{k,l=1}^{n} \left(\sum_{i=1}^{p} \beta_{ik} a_{im}^{(u)}(\mathbf{x}_k) \right) \left(\sum_{j=1}^{p} \beta_{jl} a_{jm}^{(u)}(\mathbf{x}_l) \right) \rho_{NS}^{(u)}(\mathbf{x}_k, \mathbf{x}_l)$$
$$= \sum_{u=0}^{r} \sum_{m=1}^{p} \left(\sum_{k,l=1}^{n} v_k^{u,m} v_l^{u,m} \rho_{NS}^{(u)}(\mathbf{x}_k, \mathbf{x}_l) \right),$$
$$\text{with } v_k^{u,m} = \sum_{i=1}^{p} \beta_{ik} a_{im}^{(u)}(\mathbf{x}_k), v_l^{u,m} = \sum_{j=1}^{p} \beta_{jl} a_{jm}^{(u)}(\mathbf{x}_l)$$

≥ 0, because $\left\{ \rho_{NS}^{(u)}(\cdot, \cdot) \right\}_{u=0}^{r}$ are valid univariate correlation functions,

that is $\sum_{k,l=1}^{n} v_k^{u,m} v_l^{u,m} \rho_{NS}^{(u)}(\mathbf{x}_k, \mathbf{x}_l) \geq 0$.

References

Bourgault G, Marcotte D (1991) Multivariable variogram and its application to the linear model of coregionalization. Math Geol 23(7):899–928

Burrough PA, McDonnell R, McDonnell RA, Lloyd CD (1998) Principles of geographical information systems. Oxford University Press, Oxford

Chilés JP, Delfiner P (2012) Geostatistics: modeling spatial uncertainty. Wiley, Hoboken

Desassis N, Renard D (2013) Automatic variogram modeling by iterative least squares: univariate and multivariate cases. Math Geosci 45(4):453–470

Emery X (2010) Iterative algorithms for fitting a linear model of coregionalization. Comput Geosci 36(9):1150–1160

Emery X, Arroyo D (2017) On a continuous spectral algorithm for simulating non-stationary gaussian random fields. Stoch Environ Res Risk Assess 1–15

Fouedjio F (2016) Second-order non-stationary modeling approaches for univariate geostatistical data. Stoch Environ Res Risk Assess 31(8):1887–1906

Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. Test 13(2):263–312

Genton MG, Kleiber W et al (2015) Cross-covariance functions for multivariate geostatistics. Stat Sci 30(2):147–163

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378

Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. Math Geol 24(3):269–286

Grzebyk M, Wackernagel H (1994) Multivariate analysis and spatial/temporal scales: real and complex models. Proc XVIIth Int Biomet Conf 1:19–33

Journel A, Huijbregts C (1978) Mining geostatistics. Academic Press, London

Kleiber W, Nychka D (2012) Nonstationary modeling for multivariate spatial processes. J Multivar Anal 112:76–91

Kleiber W, Porcu E (2015) Nonstationary matrix covariances: compact support, long range dependence and quasi-arithmetic constructions. Stoch Env Res Risk Assess 29(1):193–204

Lantuejoul C (2002) Geostatistical simulation: models and algorithms. Springer, New York

Lantuejoul C, Desassis N (2012) Simulation of a gaussian random vector: a propagative version of the Gibbs sampler. In: The 9th international geostatistics congress

Lark R, Papritz A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115(3—4):245–260

Majumdar A, Gelfand AE (2007) Multivariate spatial modeling for geostatistical data using convolved covariance functions. Math Geol 39(2):225–245

Majumdar A, Paul D, Bautista D (2010) A generalized convolution model for multivariate nonstationary spatial processes. Statistica Sinica 675–695

Matheron GF (1971) The theory of regionalized variables and its applications, vol 5 of Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau. École Nationale Supérieure des Mines de Paris

Paciorek CJ, Schervish MJ (2006) Spatial modelling using a new class of nonstationary covariance functions. Environmetrics 17(5):483–506

Stockmann U, Minasny B, Hancock GR, Willgoose GR (2012) Exploring short-term soil landscape formation in the Hunter

Valley, NSW, using gamma ray spectrometry. In: Malone, McBratney, Minasny (eds) Digital soil assessments and beyond. CRC Press, Boca Raton, pp 77–82

Vargas-Guzmán JA, Warrick AW, Myers DE (2002) Coregionalization by linear combination of nonorthogonal components. Math Geol 34(4):405–419

Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Springer, Berlin, Heidelberg

Wand M, Jones C (1995) Kernel smoothing. Monographs on Statistics and Applied Probability. Chapman and Hall

Zhang H, Wang Y (2010) Kriging and cross-validation for massive spatial data. Environmetrics 21(3–4):290–304