

# Classification random forest with exact conditioning for spatial prediction of categorical variables



Francky Fouedjio

*AngloGold Ashanti Australia Ltd., Growth and Exploration, 140 St. Georges Terrace, Perth, WA, 6000, Australia*

## ARTICLE INFO

**Keywords:**

Categorical variable  
Classification  
Exact conditioning  
Principal component analysis  
Signed distance  
Spatial prediction  
Quadratic programming

## ABSTRACT

Machine learning methods are increasingly used for spatially predicting a categorical target variable when spatially exhaustive predictor variables are available within the study region. Even though these methods exhibit competitive spatial prediction performance, they do not exactly honor the categorical target variable's observed values at sampling locations by construction. On the other side, competitor geostatistical methods perfectly match the categorical target variable's observed values at sampling locations by essence. In many geoscience applications, it is often desirable to perfectly match the observed values of the categorical target variable at sampling locations, especially when the categorical target variable's measurements can be reasonably considered error-free. This paper addresses the problem of exact conditioning of machine learning methods for the spatial prediction of categorical variables. It introduces a classification random forest-based approach in which the categorical target variable is exactly conditioned to the data, thus having the exact conditioning property like competitor geostatistical methods. The proposed method extends a previous work dedicated to continuous target variables by using an implicit representation of the categorical target variable. The basic idea consists of transforming the ensemble of classification tree predictors' (categorical) resulting from the traditional classification random forest into an ensemble of signed distances (continuous) associated with each category of the categorical target variable. Then, an orthogonal representation of the ensemble of signed distances is created through the principal component analysis, thus allowing to reformulate the exact conditioning problem as a system of linear inequalities on principal component scores. Then, the sampling of new principal component scores ensuring the data's exact conditioning is performed via randomized quadratic programming. The resulting conditional signed distances are turned out into an ensemble of categorical outputs, which perfectly honor the categorical target variable's observed values at sampling locations. Then, the majority vote is used to aggregate the ensemble of categorical outputs. The effectiveness of the proposed method is illustrated on a simulated dataset for which ground-truth is available and showcased on a real-world dataset, including geochemical data. A comparison with geostatistical and traditional machine learning methods show that the proposed technique can perfectly match the categorical target variable's observed values at sampling locations while maintaining competitive out-of-sample predictive performance.

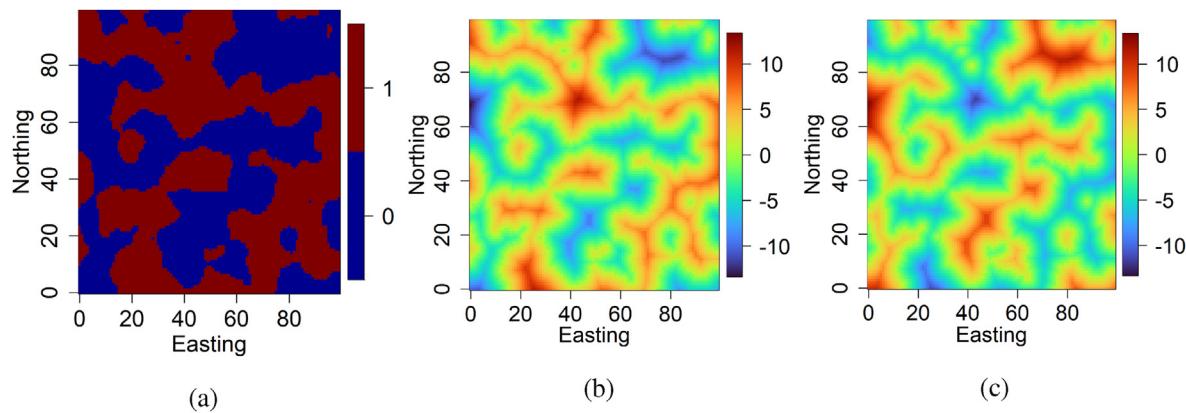
## 1. Introduction

The spatial prediction of a categorical target variable when auxiliary spatial information is available everywhere within the region under study has become ubiquitous in geosciences. Typical examples include predicting land use classes, land cover categories, drainage classes, vegetation species, landslide types, rock types, soil types, lithofacies, hydrofacies, and geological units. The mapping of categorical variables plays an essential role in a wide variety of geoscience applications. It is used to aid in risk-aware decision-making in many areas, such as environmental studies

and natural resource management. Various methods have been proposed for spatially predicting categorical target variables when spatially exhaustive predictor variables are available within the study region. These methods include geostatistical methods (Goovaerts, 2001; Hengl et al., 2004, Hengl et al., 2007), generalized linear mixed models-based approaches (Cao et al., 2011, Cao et al., 2014), and classification machine learning techniques (Kanevski, 2008; Kanevski et al., 2009; Hengl et al., 2018; Maxwell et al., 2018; Du et al., 2020; Giaccone et al., 2021).

Geostatistical methods for spatially predicting a categorical variable in the presence of auxiliary spatial information available everywhere

E-mail addresses: [ffouedjio@anglogoldashanti.com](mailto:ffouedjio@anglogoldashanti.com), [francky.fouedjio@gmail.com](mailto:francky.fouedjio@gmail.com).



**Fig. 1.** Signed distance transform approach - (a) categorical spatial variable with two categories; (b) signed distance function associated with category 0; (c) signed distance function associated with category 1.

**Table 1**  
Simulated data example - simulation parameters.

Mean	Covariance function		
	Type	Scale	Sill
$X_1(\cdot)$	10	Gaussian	12
$X_2(\cdot)$	10	Exponential	7
$X_3(\cdot)$	10	Cardinal Sine	1
$X_4(\cdot)$	10	Cubic	20
$e(\cdot)$	0	Spherical	500

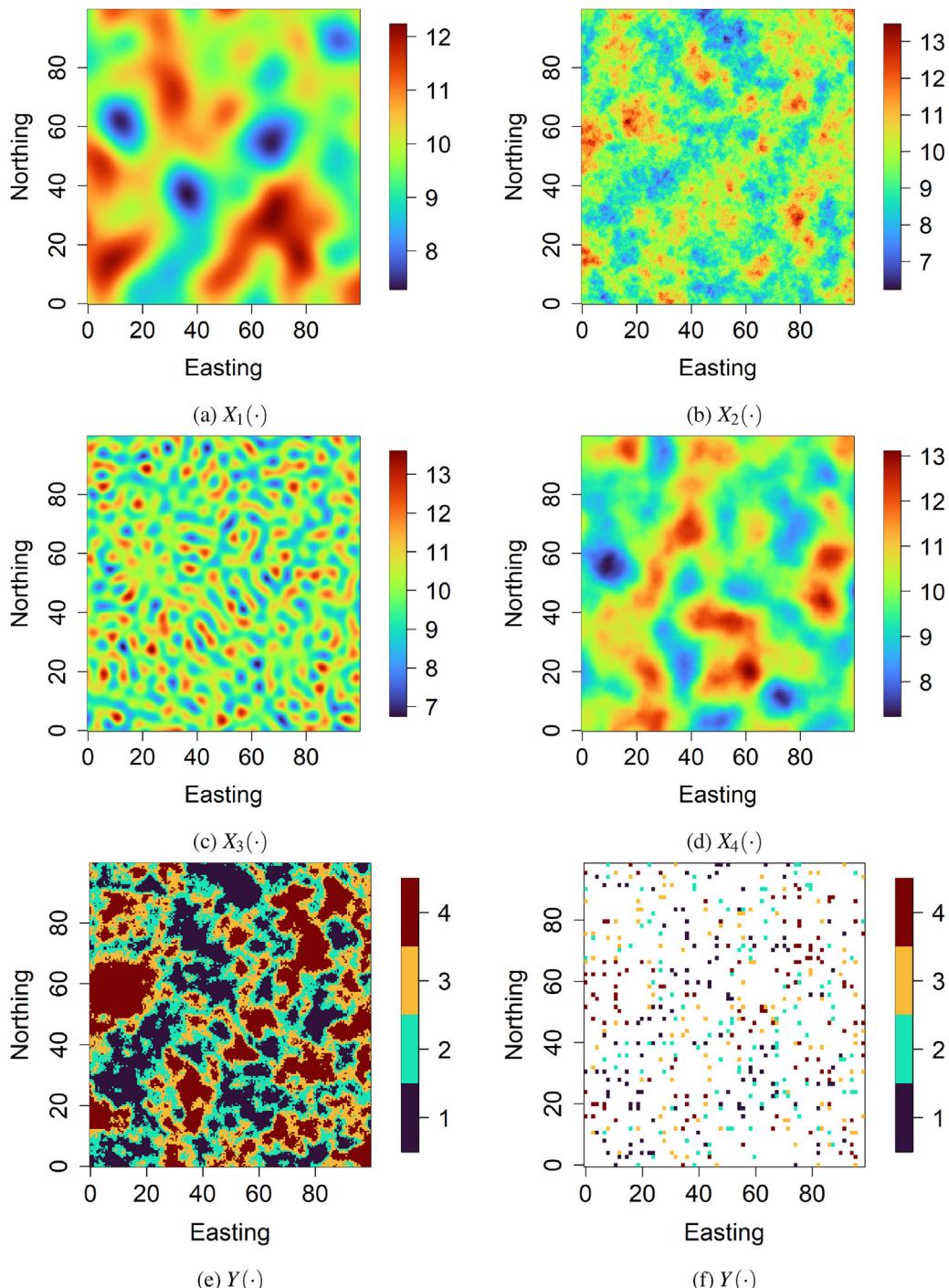
within the region under study include indicator kriging with external drift (IKED), regression-kriging of indicators (RKI), and regression-kriging of memberships (RKFm). Indicator kriging with external drift (IKED) assumes that auxiliary variables are linearly related to the class occurrence of the categorical target variable (Goovaerts, 2001). The auxiliary variables are incorporated into the indicator kriging system as deterministic linear functions. Its implementation is challenging since it is often problematic to simultaneously estimate the parameters of external drift and the covariance function of the stochastic component. Regression-kriging of indicators (RKI) combines multinomial logistic regression of the categorical target variable on predictor variables with kriging of the regression residuals (Hengl et al., 2004, Hengl et al., 2007). Thus, the regression modeling is supplemented with the modeling of variograms for regression residuals, which are then interpolated and added back to the regression estimate. RKI has been adapted to regression-kriging of memberships (RKFm) by substituting crisp indicator values with continuous membership values (Hengl et al., 2007). Indeed, under the RKI method, the interpolation of residuals might lead to values outside the physical range ( $< 0$  or  $> 1$ ). Although easy to implement, these indicator kriging-based methods have some well-known shortcomings. The predicted probabilities of the target variable's categories are not guaranteed to belong to the  $[0, 1]$  interval and sum up to one. Therefore, post-processing methods of the predicted probabilities are required (Bogaert, 2004; Allard et al., 2011). Also, under these methods, the outcome values of the conditional cumulative distribution function may not be monotonic. A posterior correction of the resulting conditional probabilities is often necessary either through a Gaussian transformation or via a logistic regression model (Pardo-Igúzquiza and Dowd, 2005).

Another alternative for spatially predicting a categorical response variable in the context of spatially exhaustive auxiliary information available within the spatial domain of interest consists of using generalized linear mixed models-based approaches. Cao et al. (2011) propose a spatial multinomial logistic mixed model in which spatially correlated latent variables are assumed to account for the spatial dependency in the categorical target variable. The proposed model is represented as a

multinomial logistic function of spatial covariances between target and sampling locations. The sought-after class occurrence probability function for a target location is written as a multinomial logistic linear combination of covariance values between the target and source data locations, which can be analogous to the dual form of kriging methods. This method was later extended to incorporate heterogeneous auxiliary information for spatial prediction of categorical variables (Cao et al., 2014). These generalized linear mixed models-based approaches are free of the aforementioned inherent problems of indicator kriging-based methods. However, they are computationally intensive compared to indicator kriging-based methods.

Machine learning techniques are increasingly used for spatially predicting a categorical response variable when auxiliary information is available everywhere within the study region. Indeed, the number of predictor variables that help explain the spatial variation in the target variable has grown dramatically, making other methods cumbersome to apply. Also, some machine learning methods are well-known for handling complex non-linear relationships and interactions and require less data pre-processing. Classification machine learning methods have proven relevant for spatially predicting categorical variables in many research works, including Albrecht et al. (2021); Kumar et al. (2020); Hengl et al. (2018); Latifovic et al. (2018); Kuhn et al. (2018); Sahoo and Jha (2017); Othman and Gloaguen (2017); Cracknell and Reading (2015, 2014); Yu et al. (2012). Even though classification machine learning methods (e.g., random forest, support vector machines) exhibit competitive spatial prediction performance, they do not exactly honor the categorical target variable's observed values at sampling locations by construction. On the other side, competitor geostatistical techniques (such as regression-kriging of indicators) perfectly match the categorical target variable's observed values at sampling locations by essence. In many geoscience applications, it is desirable to perfectly match the observed values of the categorical target variable at sampling locations, especially when the categorical target variable's measurements can be reasonably considered error-free (hard data).

This work addresses the problem of exact conditioning of machine learning methods for the spatial prediction of categorical variables. It introduces a classification random forest-based method in which the categorical target variable is exactly conditioned to the data, thus having the exact conditioning property like competitor geostatistical methods. The random forest popularity for spatial prediction relies on its ability to efficiently deal with many predictor variables, handle complex nonlinear relationships and interactions, and require less data pre-processing, and be a non-parametric method. The proposed approach extends a previous work dedicated to continuous target variables by using an implicit representation of the categorical target variable (Fouedjio, 2020). The exact conditioning to the data is achieved through a step-by-step approach. First, classification

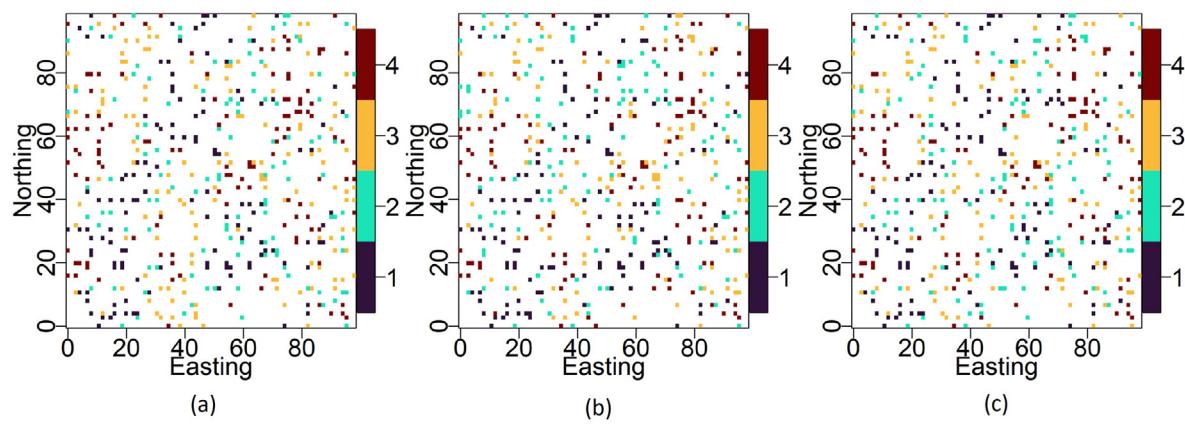


**Fig. 2.** Simulated data example - (a), (b), (c), (d) predictor variables, (e) exhaustive categorical target variable, and (f) sampled categorical target variable.

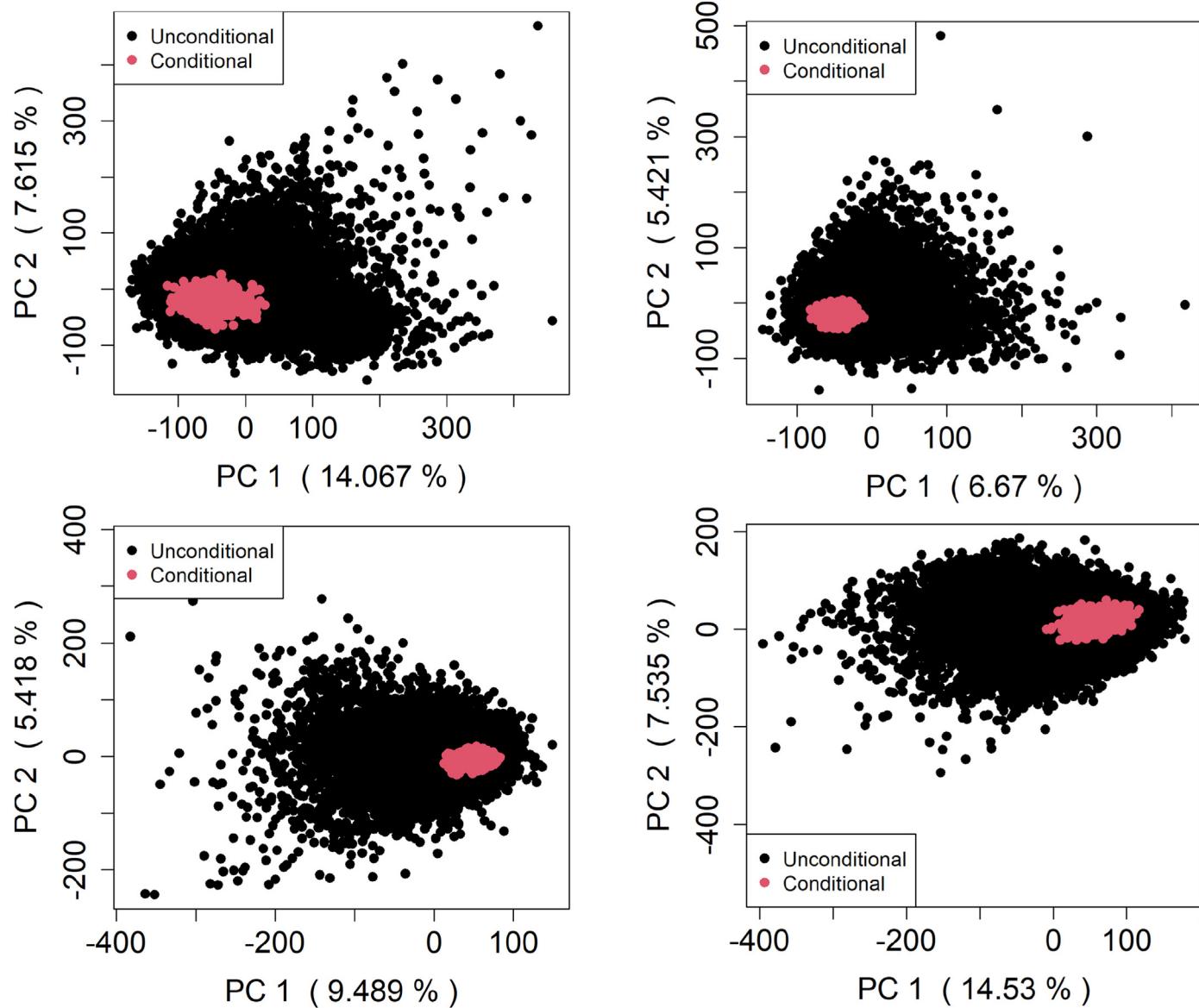
tree predictors' ensemble (categorical) resulting from the traditional classification random forest is transformed into an ensemble of signed distances (continuous) corresponding to each category of the categorical target variable. Second, an orthogonal representation of the ensemble of signed distances is created through the principal component analysis. Third, the exact conditioning problem is reformulated as a system of linear inequalities on principal component scores, thus allowing the sampling of new principal component scores (via the randomized quadratic programming), ensuring the exact conditioning to the data. Fourth, the resulting conditional signed distances are turned out into an ensemble of categorical outputs, which perfectly honor the categorical target variable's observed values at sampling

locations. Finally, the majority vote is used to aggregate the ensemble of categorical outputs. The final output also matches the target variable's observed values at sampling locations by construction. On the one hand, the proposed method's effectiveness is illustrated on a simulated dataset for which the ground truth is available. On the other hand, the proposed technique is exhibited on a real-world dataset comprising geochemical data. A comparison is also made with geostatistical and classical machine learning methods (regression-kriging of indicators, random forest, and support vector machines).

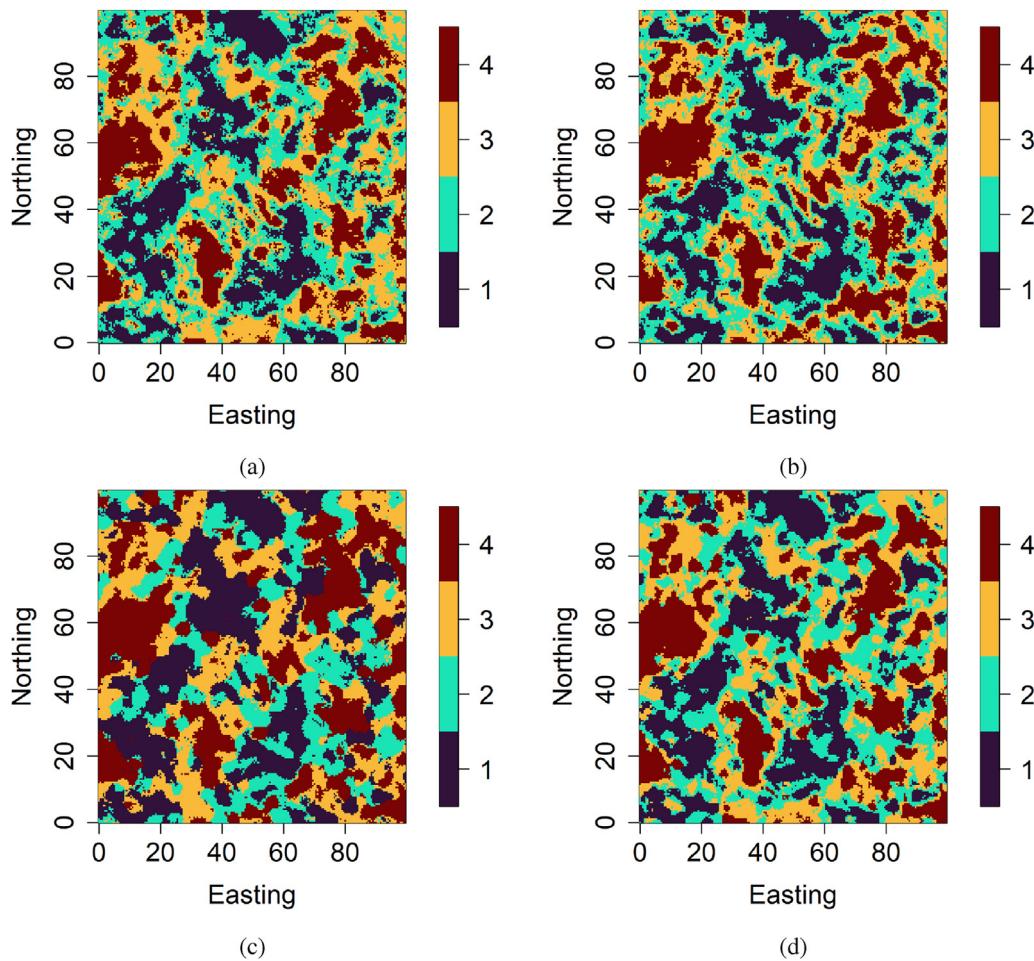
The remainder of the paper is structured as follows. Section 2 describes the different ingredients required to apply the proposed method.



**Fig. 3.** Simulated data example - predicted categorical target variable at training locations by (a) classical random forest and (b) support vector machines; (c) categorical target variable' observed values at training locations. The misclassification rate in the training data is 18.20% and 35.00%, respectively, for the traditional random forest and support vector machines.



**Fig. 4.** Simulated data example -  $B = 10\,000$  unconditional first PC scores and  $T = 1000$  conditional first PC scores.



**Fig. 5.** Simulated data example - prediction map for (a) traditional classification random forest, (b) support vector machines, (c) regression-kriging of indicators, and (d) classification random forest with exact conditioning.

Section 3 demonstrates the proposed approach's effectiveness on a synthetic dataset as well as a real-world dataset. A comparison with geo-statistical and classical machine learning methods is provided. Section 4 offers concluding remarks.

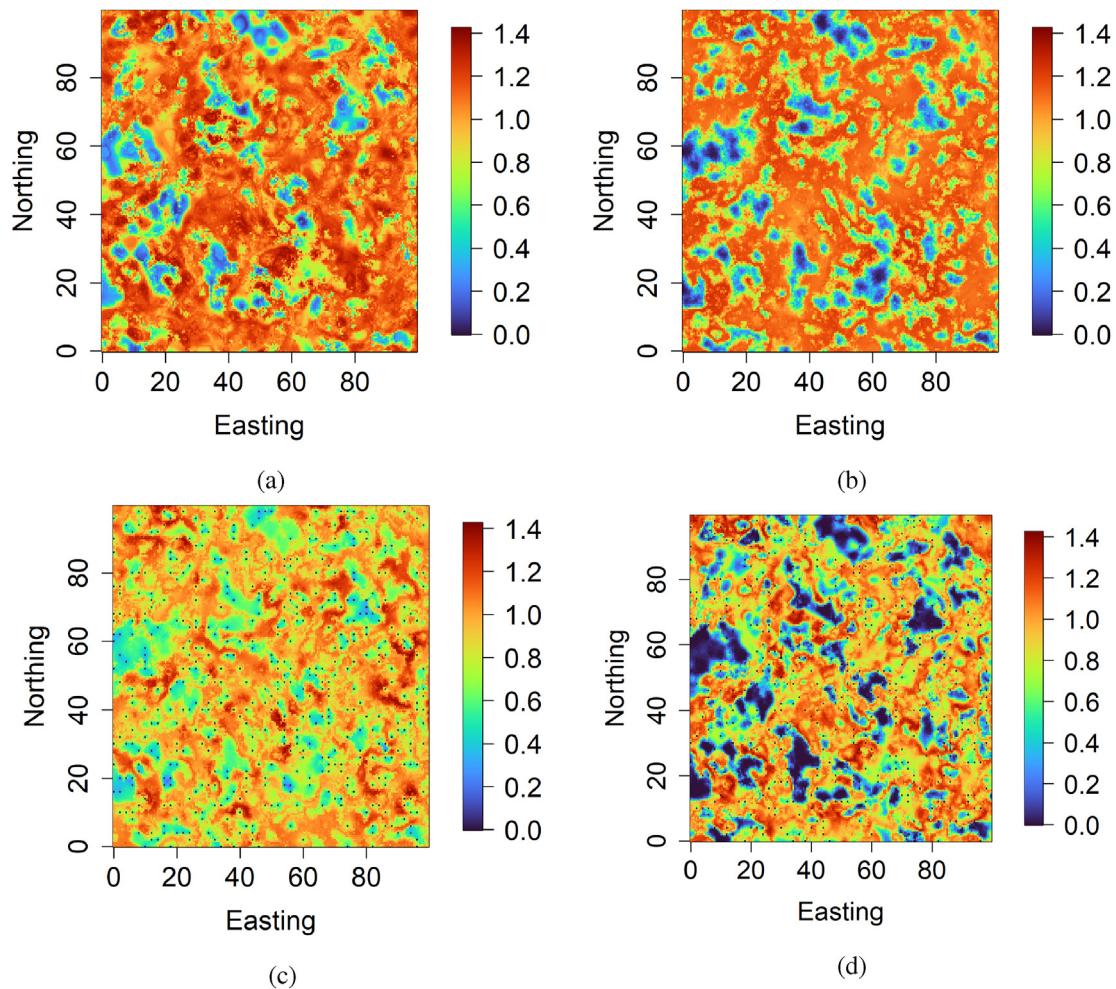
## 2. Methodology

Let  $\{C(\mathbf{s}): \mathbf{s} \in D\}$  be the categorical target variable defined on a fixed continuous spatial domain of interest  $D \subset \mathbb{R}^d$ , with a finite set of possible categorical outputs (categories)  $\{c_1, \dots, c_K\}$  which are mutually exclusive and collectively exhaustive. There exist  $n$  categorical target variable's observed values  $\{C(\mathbf{s}_i)\}_{i=1,\dots,n}$  (hard data) corresponding to sampling locations  $\{\mathbf{s}_i \in D\}_{i=1,\dots,n}$ . In addition to the categorical target variable, there is a set of predictor variables  $\{x^1(\mathbf{s}), \dots, x^p(\mathbf{s}): \mathbf{s} \in D\}$  exhaustively known in the spatial domain  $D$ . We address the problem of predicting the categorical target variable over the spatial domain  $D$  represented by  $N$  grid locations using the categorical target variable's observed values and predictor variables' data. In addition, the categorical target variable's predicted values at sampling locations must be the same as the categorical target variable's observed values at sampling locations, i.e.,  $\hat{C}(\mathbf{s}_i) = C(\mathbf{s}_i)$ ,  $i = 1, \dots, n$ . The description of the different ingredients needed to implement the proposed exact conditioning method is given in this section. The implementation is carried out in the R platform ([R Core Team, 2020](#)).

### 2.1. Random forest classifier

The first step of the proposed method consists of training the traditional random forest (RF) classifier on the data. Random forest classifier is an ensemble method where several individual decision trees are trained on various subsets of the training dataset (bootstrap samples) using different subsets of available predictor variables, followed by an aggregation ([Breiman, 2001](#)). The bootstrapping of the training data and the random selection of subsets of predictor variables ensure that each decision tree in the random forest is unique, which reduces the overall variance of the random forest classifier. For the final decision, the RF classifier aggregates the decisions of individual trees through a voting scheme such as the majority voting, i.e., for each observation, each decision tree votes for one category, and RF chooses the category with the highest number of votes.

Classification random forest has some tuning parameters that can be optimized. There are, among others, the number of trees, number of predictor variables randomly selected at each node, proportion of observations to sample in each decision tree, and minimum number of observations in a decision tree's terminal node. These hyperparameters are optimized via cross-validation. In practice, there is no need to tune the number of decision trees; it is usually recommended to set it to a large number, allowing the convergence of the prediction error to a stable minimum ([Hengl et al., 2018](#)). The implementation of the classification random forest is carried out using the R packages *ranger* ([Wright and](#)



**Fig. 6.** Simulated data example - prediction uncertainty (entropy) map for (a) traditional classification random forest, (b) support vector machines, (c) regression-kriging of indicators, and (d) classification random forest with exact conditioning.

**Table 2**

Simulated data example - predictive performance statistics in the testing dataset containing 39 500 observations.

Methods	Accuracy	Rand index
Random Forest	0.619	0.734
Support Vector Machines	0.602	0.729
Regression-Kriging of Indicators	0.623	0.746
Random Forest with Exact Conditioning	0.639	0.747

Ziegler, 2017) and *tuneRanger* (Probst et al., 2018).

The training of the classical random forest classifier results in an ensemble of classification tree predictors  $\{\tilde{C}_b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1,\dots,B}$ , where  $B$  is the number of decision trees. As the traditional random forest classifier is not explicitly designed to match the data perfectly, classification tree predictors and the aggregated classification tree predictors do not necessarily match the categorical target variable's observed values at sampling locations. Since  $\{\tilde{C}_b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1,\dots,B}$  do not match data perfectly, they will be called “unconditional classification tree predictors”. The next steps aim to generate conditional classification tree predictors that perfectly fit the categorical target variable's observed values at sampling locations.

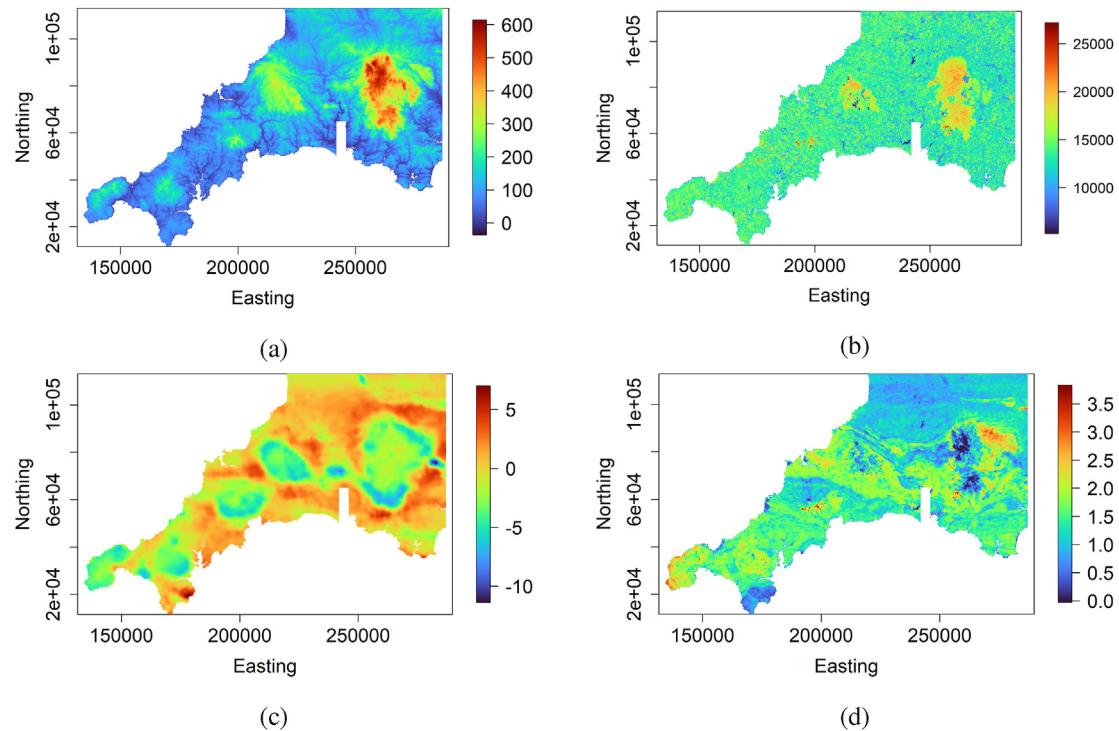
## 2.2. Signed distance transform

The second step of the proposed method includes transforming the unconditional classification tree predictors' ensemble (categorical) into

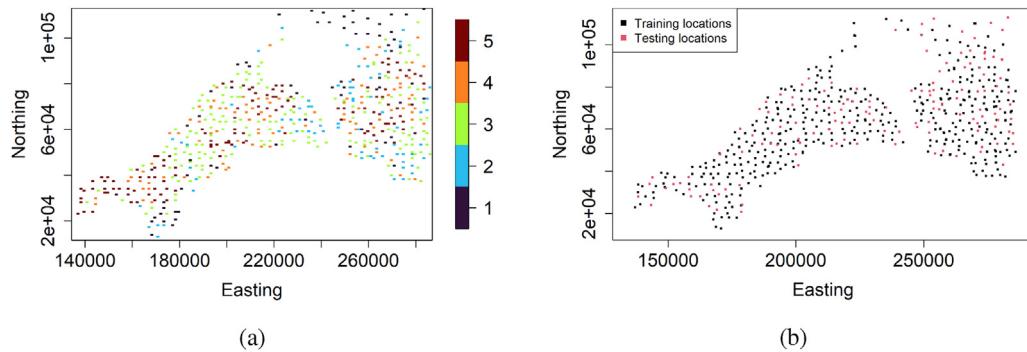
an ensemble of unconditional signed distance functions (continuous) corresponding to each category of the categorical target variable. The categorical target variable  $\{C(\mathbf{s}) : \mathbf{s} \in D\}$  with  $K$  categories  $\{c_k\}_{k=1,\dots,K}$  can be viewed as a variable that creates distinct boundaries in the study region  $D$ . Each category  $c_k (k = 1, \dots, K)$  can be codified by a binary variable indicating their presence or absence:  $I_k(\mathbf{s}) = 1$  if  $C(\mathbf{s}) = c_k$ , and  $I_k(\mathbf{s}) = 0$  if  $C(\mathbf{s}) \neq c_k, \forall \mathbf{s} \in D$ . Each category  $c_k$  can be represented by a signed distance function  $\varphi_k(\cdot)$  such that  $c_k = \{\mathbf{s} \in D, \varphi_k(\mathbf{s}) \leq 0\}$ . The signed distance transform approach (Grevera, 2007; Davies, 2012) can be used to transform each category  $c_k (\{I_k(\mathbf{s}) : \mathbf{s} \in D\})$  into a signed distance function  $\varphi_k(\cdot)$ . Indeed, each category  $c_k$  defines a  $p$ -dimensional binary image  $\{I_k(\mathbf{s}) : \mathbf{s} \in D\}$  where each point (pixel) has either a value of 1 indicating the presence of the category  $c_k$  or a value of 0 indicating the absence of the category  $c_k$ . For every point (pixel) set to 1, a distance transform assigns a value indicating the negatively signed distance from that point (pixel) to the nearest point (pixel) set to 0. Similarly, for every point (pixel) set to 0, a distance transform assigns a value indicating the positively signed distance from that point (pixel) to the nearest point (pixel) set to 1. An illustration of the signed distance transform method is given in Fig. 1. Additionally, the signed distance transformation is one-to-one. The bijectivity is obtained using the following rule:

$$C(\mathbf{s}) = \underset{c_1,\dots,c_K}{\operatorname{argmin}}(\varphi_1(\mathbf{s}), \dots, \varphi_K(\mathbf{s})), \quad \forall \mathbf{s} \in D. \quad (1)$$

Thus, the ensemble of unconditional classification tree predictors  $\{\tilde{C}_b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1,\dots,B}$  is converted to an ensemble of unconditional



**Fig. 7.** Real-world data example - some predictor variables: (a) elevation, (b) Landsat 8 band 6, (c) gravity survey high-pass filtered Bouguer anomaly, (d) potassium counts from gamma ray spectrometry.



**Fig. 8.** Real-world data example - (a) categorical target variable and (b) training and testing locations.

signed distance functions  $\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1,\dots,B}$  for each category  $c_k (k = 1, \dots, K)$  via the signed distance transform approach previously described. The following idea uses principal component analysis to create an orthogonal representation of the ensemble of unconditional signed distance functions. Then, the exact conditioning problem is reformulated as a linear inequality problem on the principal component scores. Then, new principal component scores that ensure the exact conditioning to the hard data are generated through the randomized quadratic programming. Since the principal component orthogonalization is bijective, conditional signed distance functions are obtained by reconstruction. The combination rule defined in Eq. (1) is applied to obtain conditional classification tree predictors (categorical) that exactly match the categorical target variable's observed values at sampling locations.

### 2.3. Principal component analysis

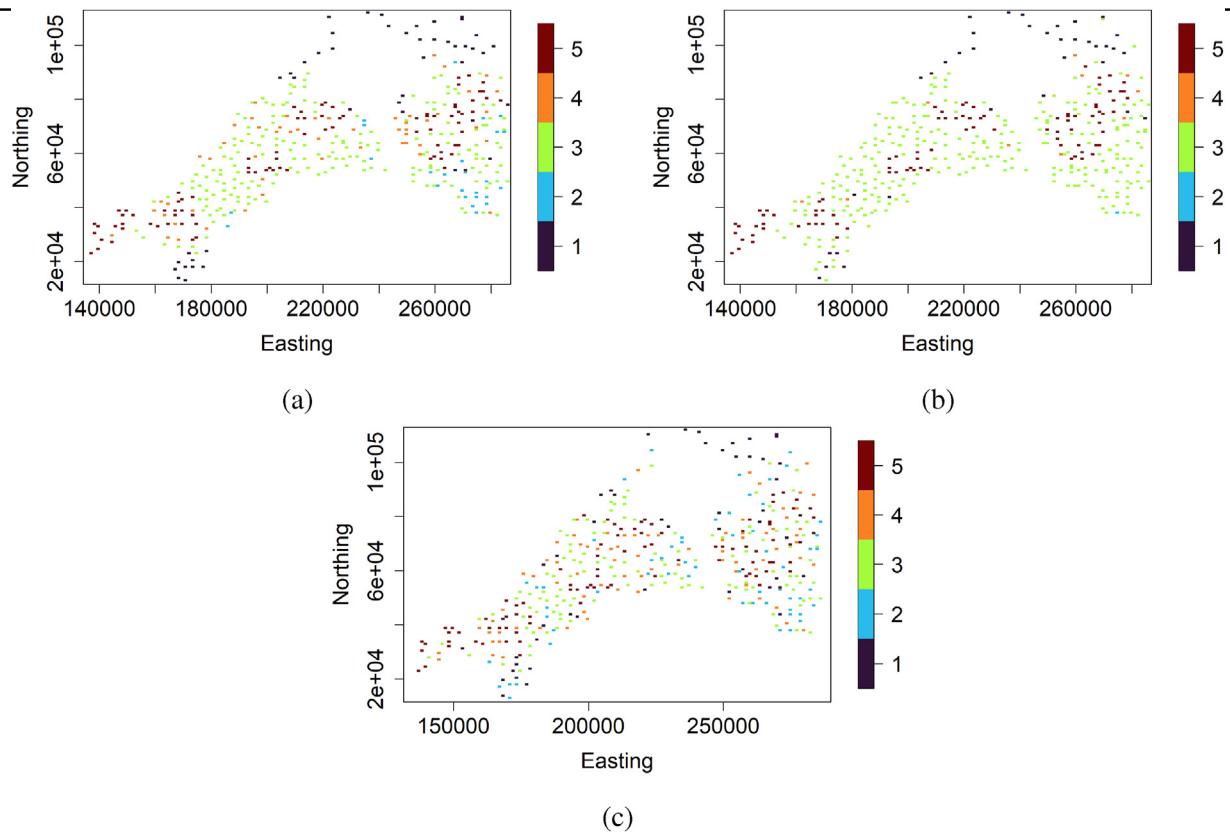
This step consists of performing principal component analysis (PCA) on each ensemble of unconditional signed distance functions

$\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1,\dots,B} (k = 1, \dots, K)$ . This results in the following decomposition in finite dimensions:

$$\tilde{\varphi}_k^b(\mathbf{s}) = \sum_{l=1}^L a_{l,k}^b \psi_{l,k}(\mathbf{s}), \quad \forall \mathbf{s} \in D, \quad b = 1, \dots, B; \quad k = 1, \dots, K; \quad (2)$$

where  $\{a_{l,k}^b\}_{l=1,\dots,L}$  are principal component scores (coefficients) and  $\{\psi_{l,k}(\mathbf{s}) : \mathbf{s} \in D\}_{l=1,\dots,L}$  are principal components factors (eigen-functions);  $L = \min(B, N)$ .

For each category  $c_k (k = 1, \dots, K)$ , PCA is applied to a matrix  $\Gamma_k (B \times N)$  arranged as a set of  $B$  row vectors, each representing a single unconditional signed distance function  $\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in D\}$ . PCA is parallelizable for each matrix  $\Gamma_k (k = 1, \dots, K)$ . In Eq. (2), the ensemble  $\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1,\dots,B}$  can be viewed as a set of images and  $\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in D\}$  as an image. Thus, the resulting principal components factors  $\{\psi_{l,k}(\mathbf{s}) : \mathbf{s} \in D\}_{l=1,\dots,L}$  are images as well. Hence, Eq. (2) provides a decomposition of the images into a set of eigen-images and a set of coefficients. It is important to note that in the PCA framework, the eigenfunctions are considered fixed, while the coefficients are considered



**Fig. 9.** Real-world data example - predicted categorical target variable at training locations by (a) classical random forest and (b) support vector machines; (c) categorical target variable' observed values at training locations

random. The PCA is used here more as an orthogonal decomposition method than a dimension reduction technique since all the principal component factors are kept, as shown in Eq. (2). The bijective nature of PCA allows the reconstruction of signed distance functions from coefficients. In other words, an image can be reconstructed back once all the principal component factors and scores are used.

#### 2.4. Randomized quadratic programming

Given the PCA decomposition of the ensemble of unconditional signed distance functions as described in Sect. 2.3, this step consists of generating new principal component scores such that signed distance functions defined in Eq. (2) perfectly match the data. Let

$$\varphi_k(\mathbf{s}) = \sum_{l=1}^L \theta_{l,k} \psi_{l,k}(\mathbf{s}), \quad \forall \mathbf{s} \in D, \quad k = 1, \dots, K, \quad (3)$$

where  $\{\theta_{l,k}\}_{l=1,\dots,L}$  are random coefficients and  $\{\psi_{l,k}(\mathbf{s}) : \mathbf{s} \in D\}_{l=1,\dots,L}$  are principal component factors derived from the PCA decomposition of unconditional signed distance functions as given in Eq. (2).

The categorical target variable's observed values at sampling locations (hard data) inform the sign of the signed distance function associated with a category at sampling locations. Thus, the set of hard data can be converted into a set of inequality constraints using Eq. (3). Let assume that at the sampling location  $\mathbf{s}_1$ , the category  $c_2$  is observed, i.e.,  $C(\mathbf{s}_1) = c_2$ . This means that the signed distance function associated with the

category  $c_2$  should be negative at location  $\mathbf{s}_1$  ( $\varphi_2(\mathbf{s}_1) \leq 0$ ), and the signed distance functions associated with other categories should be positive at location  $\mathbf{s}_1$  ( $\varphi_k(\mathbf{s}_1) \geq 0, \forall k \neq 2; k = 1, \dots, K$ ). For each  $\varphi_k (k = 1, \dots, K)$ , the conditioning to all data locations is expressed by the following inequalities:

$$\begin{cases} \varphi_k(\mathbf{s}_1) = \theta_{1,k} \psi_{1,k}(\mathbf{s}_1) + \theta_{2,k} \psi_{2,k}(\mathbf{s}_1) + \dots + \theta_{L,k} \psi_{L,k}(\mathbf{s}_1) \leq 0 \text{ or } \geq 0 \\ \dots \\ \varphi_k(\mathbf{s}_n) = \theta_{1,k} \psi_{1,k}(\mathbf{s}_n) + \theta_{2,k} \psi_{2,k}(\mathbf{s}_n) + \dots + \theta_{L,k} \psi_{L,k}(\mathbf{s}_n) \leq 0 \text{ or } \geq 0 \end{cases}. \quad (4)$$

In Eq. (4), the  $n$  inequalities corresponding to  $n$  hard data can be summarized as:

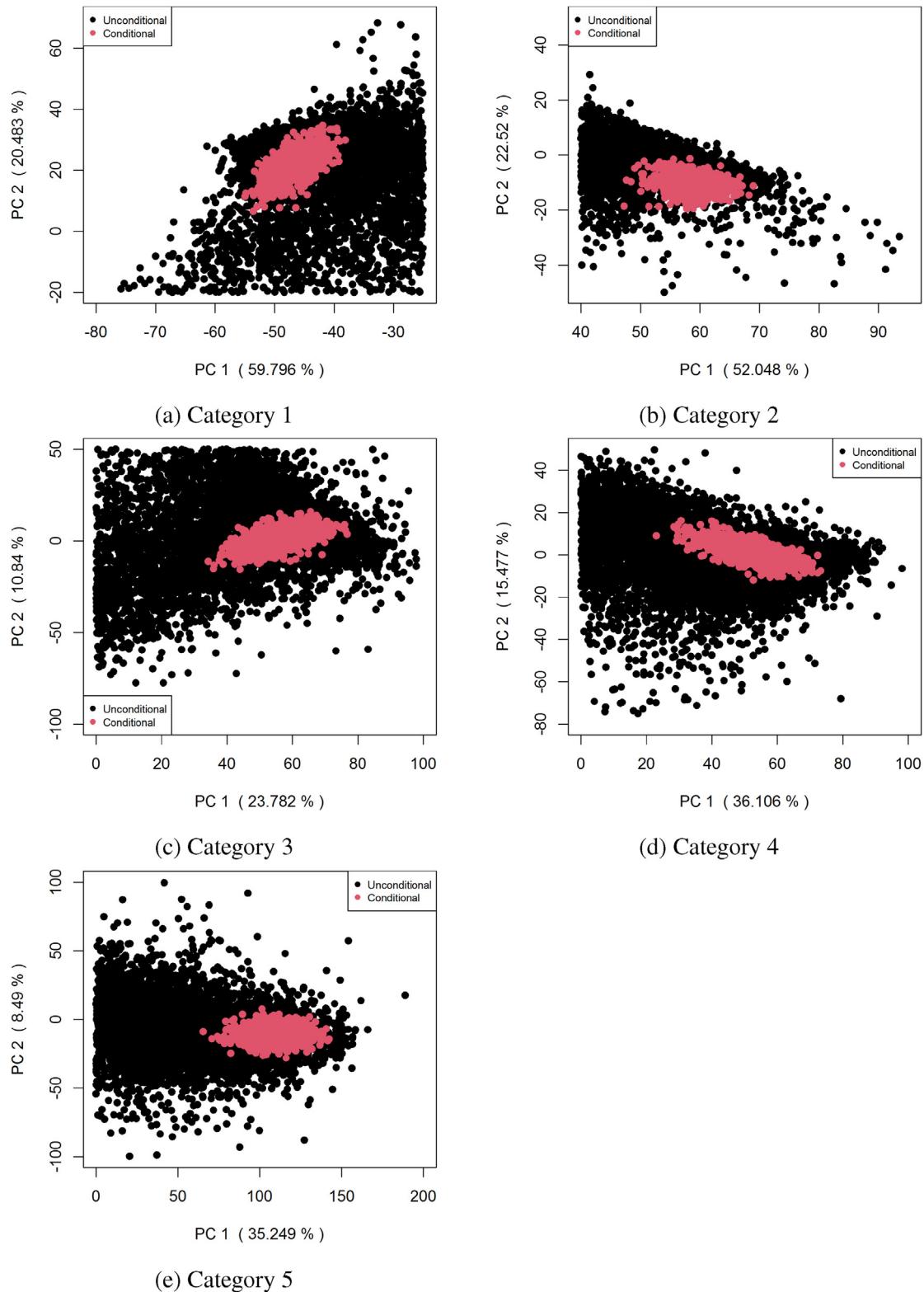
$$\tilde{\Psi}_k \boldsymbol{\theta}_k \leq \mathbf{0}, \quad k = 1, \dots, K. \quad (5)$$

New PC scores vector  $\boldsymbol{\theta}_k = (\theta_{1,k}, \dots, \theta_{L,k})^T$  that matches data are generated by solving the following randomized quadratic optimization problem (Fouedjio et al., 2021a):

$$\min_{\boldsymbol{\theta}_k \in \mathbb{R}^L} ((\boldsymbol{\theta}_k - \boldsymbol{\beta}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\theta}_k - \boldsymbol{\beta}_k)) \quad \text{subject to} \quad \tilde{\Psi}_k \boldsymbol{\theta}_k \leq \mathbf{0}, \quad k = 1, \dots, K. \quad (6)$$

where  $\boldsymbol{\beta}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ , and the mean  $\boldsymbol{\mu}_k$  and the covariance matrix  $\boldsymbol{\Sigma}_k$  are computed using unconditional PC scores  $\{\alpha_{l,k}^b\}_{l=1,\dots,L}$  derived from the PCA of unconditional signed distance functions given in Eq. (2). Specifically,

$$\boldsymbol{\mu}_k = \left[ \frac{1}{B} \sum_{b=1}^B \alpha_{l,k}^b \right]_{l=1,\dots,L} ; \quad \boldsymbol{\Sigma}_k = \frac{1}{B-1} \sum_{b=1}^B (\boldsymbol{\alpha}_k^b - \boldsymbol{\mu}_k)(\boldsymbol{\alpha}_k^b - \boldsymbol{\mu}_k)^T, \quad \text{with} \quad \boldsymbol{\alpha}_k^b = [\alpha_{l,k}^b]_{l=1,\dots,L}. \quad (7)$$



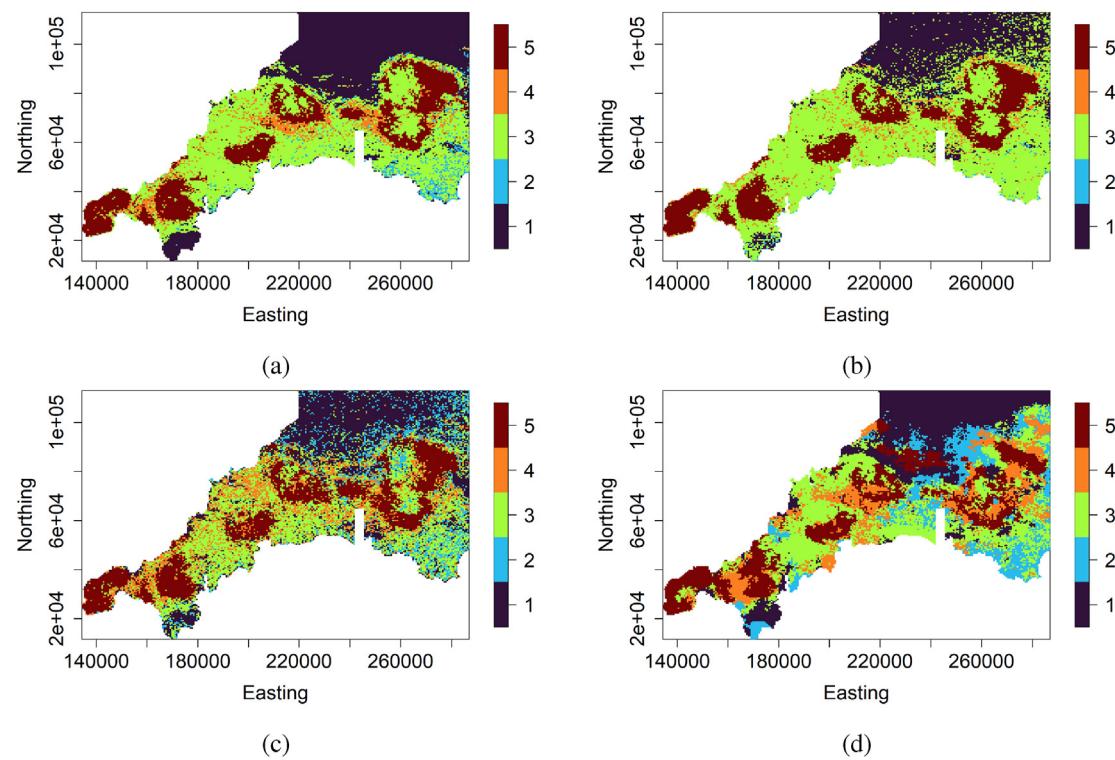
**Fig. 10.** Real-world data example - unconditional and conditional first two PC scores associated with each category.

For each sample  $\beta_k^t \sim \mathcal{N}(\mu_k, \Sigma_k)$  ( $t = 1, \dots, T$ ), quadratic programming (Goldfarb and Idnani, 1983) is performed to find a solution  $\theta_k^t$  that satisfies the inequality constraints and minimizes the quadratic objective function in Eq. (6). The covariance matrix  $\Sigma_k$  in Eq. (7) is a diagonal matrix because the PC scores are uncorrelated by construction. Coefficients  $\theta_k^t$  can be also generated via the Gibbs sampling method

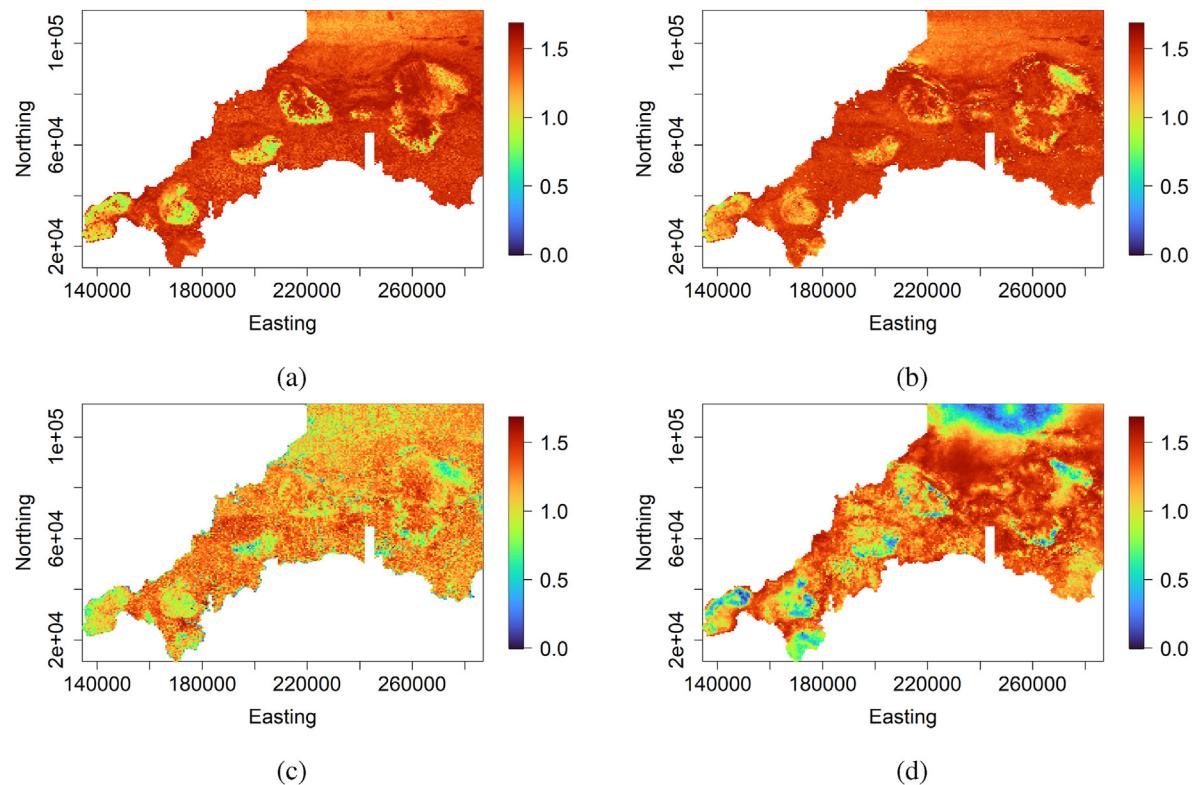
(Fouedjio et al., 2021b). However, this approach can be time-consuming for very large datasets since Gibbs samples are highly correlated.

Given conditional PC scores  $\{\theta_k^t\}_{t=1, \dots, T}$ , conditional signed distance functions are obtained by reconstruction:

$$\varphi_k^t(\mathbf{s}) = \sum_{l=1}^L \theta_{l,k}^t \psi_{l,k}(\mathbf{s}), \quad \forall \mathbf{s} \in D. \quad (8)$$



**Fig. 11.** Real-world data example - prediction map provided by (a) traditional classification random forest, (b) support vector machines, and (c) regression-kriging of indicators, and (d) classification random forest with exact conditioning.



**Fig. 12.** Real-world data example - prediction uncertainty (entropy) map provided by (a) traditional classification random forest, (b) support vector machines, (c) regression-kriging of indicators, and (d) classification random forest with exact conditioning.

Conditional classification tree predictors are given by applying the combination rule defined in Eq. (1):

$$C_t(\mathbf{s}) = \underset{c_1, \dots, c_K}{\operatorname{argmin}} (\varphi_1^t(\mathbf{s}), \dots, \varphi_K^t(\mathbf{s})), \quad \forall \mathbf{s} \in D, \quad t = 1, \dots, T. \quad (9)$$

predictors  $B$  under randomized quadratic programming. That is to say,  $T$  can be smaller or greater than  $B$ .

To summarize, the proposed classification random forest with exact conditioning is performed using the following pseudo algorithm:

---

**Algorithm 1** Classification Random Forest with Exact Conditioning

---

**Input:** categorical target variable' data  $\{C(\mathbf{s}_i) : \mathbf{s}_i \in D\}_{i=1, \dots, n}$  and predictor variables' data

$\{\mathbf{x}^1(\mathbf{s}), \dots, \mathbf{x}^p(\mathbf{s}) : \mathbf{s} \in D\}$ ; the spatial domain  $D$  is represented by a number of grid locations  $N$ .

1. Perform traditional classification random forest on the data; the outcome is an ensemble of unconditional classification tree predictors  $\{\tilde{C}_b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1, \dots, B}$ ;
2. Transform the ensemble of unconditional classification tree predictors  $\{\tilde{C}_b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1, \dots, B}$  into an ensemble of unconditional signed distance functions  $\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in D\}_{b=1, \dots, B}$  for each category  $c_k (k = 1, \dots, K)$  of the target variable using the signed distance transform approach;
3. Apply PCA to each matrix  $\Gamma_k (B \times N)$  arranged as a set of  $B$  row vectors, each representing a single unconditional signed distance function  $\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in D\}$  ( $k = 1, \dots, K$ ); the PCA outcomes are unconditional PC scores  $\{\alpha_{l,k}^b\}_{l=1, \dots, L}$  and PC factors  $\{\psi_{l,k}(\mathbf{s}) : \mathbf{s} \in D\}_{l=1, \dots, L}$ ;  $L = \min(B, N)$ ;
4. Generate conditional PC scores  $\{(\theta_{1,k}^t, \dots, \theta_{L,k}^t)\}_{t=1, \dots, T}$  using the randomized quadratic programming method ( $k = 1, \dots, K$ ) defined in Eq. (6);
5. Derive conditional signed distance functions by reconstruction  $\tilde{\varphi}_k^t(\mathbf{s}) = \sum_{l=1}^L \theta_{l,k}^t \psi_{l,k}(\mathbf{s}), \quad \forall \mathbf{s} \in D$  ( $k = 1, \dots, K$ );
6. Turn out conditional signed distance functions into conditional classification tree predictors by applying the following rule:  $\check{C}_t(\mathbf{s}) = \arg \min_{c_1, \dots, c_K} (\tilde{\varphi}_1^t(\mathbf{s}), \dots, \tilde{\varphi}_K^t(\mathbf{s})), \quad \forall \mathbf{s} \in D, \quad t = 1, \dots, T$ ;
7. Apply the majority vote rule on the ensemble  $\{\check{C}_t(\mathbf{s}) : \mathbf{s} \in D\}_{t=1, \dots, T}$  to get the final result  $\{\hat{C}(\mathbf{s}) : \mathbf{s} \in D\}$ .

**Output:** predicted categorical target variable  $\{\hat{C}(\mathbf{s}) : \mathbf{s} \in D\}$ .

---

Since all the individual reconstructed signed distance functions  $\{\varphi_k^t(\mathbf{s}) : \mathbf{s} \in D\}_{t=1, \dots, T}$  ( $k = 1, \dots, K$ ) perfectly match the hard data, conditional classification tree predictors  $\{C_t(\mathbf{s}) : \mathbf{s} \in D\}_{t=1, \dots, T}$  do also. The aggregation of the conditional classification tree predictors using the majority vote rule leads to the final outcome  $\{\hat{C}(\mathbf{s}) : \mathbf{s} \in D\}$ . This latter coincides with the categorical target variable observed values at sampling locations.

The number of unconditional classification tree predictors  $B$  should be large enough to allow good coverage of the solution space when performing the exact conditioning. Indeed, the number of categorical target variable' observations defines the number of inequalities constraints as shown in Eq. (4). The larger is the number of unconditional classification tree predictors, the wider is the solution space of Eq. (6). So, too many constraints (hard data) relative to too few unconditional classification tree predictors will lead to too small uncertainty. It is worth mentioning that the number of conditional classification tree predictors  $T$  does not depend on the number of unconditional classification tree

### 3. Empirical examples

The proposed classification random forest with exact conditioning is illustrated using simulated and real-world datasets. A prediction performance comparison is carried out with a geostatistical method (regression-kriging of indicators) and traditional machine learning techniques (random forest and support vector machines). Hyper-parameters associated with each machine learning method have been optimized through cross-validation. The proposed classification random forest with exact conditioning uses the same ensemble of decision trees generated by the classical random forest.

The predictive performance of each method is assessed on a testing dataset using the first evaluation statistic, i.e., the accuracy. The accuracy corresponds to the percentage of observations that are correctly classified. It has a value between 0 and 1. The higher is the accuracy, the better is the model. In addition to the accuracy, the Rand index is calculated. The Rand index measures the similarity between the predicted classification and true classification on the testing data by considering all pairs of points and counting pairs that are assigned in the same or different category in the predicted and true classifications. The Rand index has a value between 0 and 1, with 0 indicating that the two classifications do not agree on any pair of points and 1 indicating the same classifications.

#### 3.1. Simulated data example

In this simulated case study, we consider a categorical target variable with four categories, and four continuous predictor variables defined

**Table 3**

Real-world data example - predictive performance statistics in the testing dataset containing 140 observations.

Methods	Accuracy	Rand index
Random Forest	0.386	0.649
Support Vector Machines	0.342	0.575
Regression-Kriging of Indicators	0.336	0.672
Random Forest with Exact Conditioning	0.421	0.710

over the spatial domain  $[0,100]^2$ . The categorical target variable is generated according to the following model:

$$\forall \mathbf{s} \in [0, 100]^2, \quad C(\mathbf{s}) = c_k \quad \text{if} \quad Y(\mathbf{s}) \in [q_k, q_{k+1}[, \quad k = 1, \dots, 4; \quad (10)$$

with  $Y(\mathbf{s}) = 50\sin(X_1(\mathbf{s})) + 3X_1(\mathbf{s})X_2(\mathbf{s}) + X_3(\mathbf{s})^2 + 50\sin(X_4(\mathbf{s})) + \epsilon(\mathbf{s})$ ;  $X_1(\cdot)$ ,  $X_2(\cdot)$ ,  $X_3(\cdot)$ , and  $X_4(\cdot)$  are predictor variables; and  $\epsilon(\cdot)$  is a latent (non-observed) variable; the limits  $q_j$ 's are taken as the 0, 0.25, 0.50, 0.75, and 1 quantiles of the random function  $Y(\cdot)$ , so that the complete system of events  $c_k$  ( $k = 1, \dots, 4$ ) can be defined.

The four predictor variables and the latent variable are simulated on the spatial domain  $[0,100]^2$  based on five independent Gaussian isotropic stationary random functions (Chiles and Delfiner, 2012) with different specification of means and covariance functions as given in Table 1. The simulation is performing using the R package *RGeostats* package (Renard et al., 2020). Fig. 2 presents the simulated data over a  $200 \times 200$  regular grid. The map of the categorical target variable displayed in Fig. 2e is considered as the reference map.

To demonstrate the proposed method's ability to exactly match the categorical target variable's observed values at sampling locations,  $n = 500$  stratified random samples are taken as the training data (Fig. 2). The set of  $n = 500$  stratified random samples amounts to 1.25% of total locations in the reference map and each category contains 125 samples. The rest of data (39 500 samples) is kept aside for the testing. The goal is to reconstruct the reference map of the categorical target variable (Fig. 2e) using the sampled categories (Fig. 2f) with an aid of the observed four spatial auxiliary variables (Fig. 2a - d) such that the categorical target variable's predicted values coincide with the categorical target variable's observed values at sampling locations.

Fig. 3 shows the categorical target variable's observed values at training locations and those predicted by the traditional random forest and support vector machines. There is a significant disagreement between the observed values and the predicted values of categorical target variable at training locations. The misclassification rate in the training data is 18.20% and 35.00%, respectively, for the traditional random forest and support vector machines. For the traditional random forest, the number of decisions trees has been set to 10 000, and the hyper-parameters have been optimized through cross-validation. Concerning the support vector machines, the kernel function and the hyper-parameters have been selected using cross-validation.

The traditional classification random forest is performed on the training data with a large number of decision trees set to  $B = 10\,000$ . Thus, an ensemble of  $B = 10\,000$  unconditional classification tree predictors  $\{\tilde{C}_b(\mathbf{s}) : \mathbf{s} \in [0, 100]^2\}_{b=1, \dots, 10\,000}$  is constructed. This latter ensemble is transformed into an ensemble of unconditional signed distance functions  $\{\tilde{\varphi}_k^b(\mathbf{s}) : \mathbf{s} \in [0, 100]^2\}_{b=1, \dots, 10\,000}$  for each category  $c_k$  ( $k = 1, \dots, 4$ ), according to the methodology described in Sect. 2. Principal component analysis is performed on this ensemble, followed by the sampling of new PC scores ensuring the data's exact conditioning.  $T = 1000$  new PC scores are generated, thus giving an ensemble of  $T = 1000$  conditional signed distance functions  $\{\varphi_k^t(\mathbf{s}) : \mathbf{s} \in [0, 100]^2\}_{t=1, \dots, 1000}$  ( $k = 1, \dots, 4$ ). This latter are turned into conditional classification tree predictors  $\{C_t(\mathbf{s}) : \mathbf{s} \in [0, 100]^2\}_{t=1, \dots, 1000}$  that perfectly match the categorical target variable's observed values at sampling locations. The majority vote scheme is then used to derive the predicted categorical target variable  $\{\hat{C}(\mathbf{s}) : \mathbf{s} \in [0, 100]^2\}$ . This latter also perfectly matches the categorical target variable's observed values at sampling locations by construction.

Fig. 4 shows PC scores before the conditioning (unconditional PC scores  $\{\alpha_k^l\}_{l=1, \dots, 10\,000}$ ) and after the conditioning (conditional PC scores  $\{\theta_k^t\}_{t=1, \dots, 1000}$ ) for each category  $c_k$  ( $k = 1, \dots, 4$ ). One can notice that the points cloud of conditional PC scores is less scattered than those from unconditional PC scores due effectively to the exact conditioning

expressed as a system of linear inequalities.

Fig. 5 presents prediction maps provided by the traditional classification random forest, support vector machines, regression-kriging of indicators, and the proposed classification random forest with exact conditioning. One can notice the prediction maps provided by regression-kriging of indicators (Fig. 5c) and the proposed classification random forest (Fig. 5d) are exactly conditioned to the training data (Fig. 2f), which is not the case for the ones provided by traditional classification random forest (Fig. 5a) and support vector machines (Fig. 5b). The general appearance of prediction maps resulting from the traditional classification random forest and the proposed classification random forest with exact conditioning looks similar. However, there are some local differences due to the exact conditioning nature of the proposed method. It is important to highlight that the proposed classification random forest with exact conditioning uses the ensemble of decision trees generated by the classical random forest as starting point. Overall, the prediction map of the proposed classification random forest exhibits more similar spatial patterns present in the reference map (Fig. 2e) than the regression-kriging of indicators.

The traditional classification random forest, support vector machines, regression-kriging of indicators, and proposed classification random forest with exact conditioning provide the probabilities for each possible outcome of the categorical target variable at any spatial location. Thus, the prediction uncertainty can be quantified through information entropy (Wellmann and Regenauer-Lieb, 2012). Fig. 6 presents the prediction uncertainty maps associated with each method. The regression-kriging of indicators and the proposed classification random forest provide zero entropy at sampling location by construction while the traditional classification random forest and support vector machines do not. The traditional classification random forest, support vector machines, and proposed classification random forest provide lower entropy in local neighborhoods dominated by a single category compared to the regression-kriging of indicators.

The predictive performance statistics in the testing dataset (containing 39 500 observations) for the traditional classification random forest, support vector machines, regression-kriging of indicators, and proposed classification random forest with exact conditioning are reported in Table 2. In addition to exactly fitting the categorical target variable's observed values at sampling locations, the proposed classification random forest maintains a competitive out-of-sample predictive performance.

### 3.2. Real-world data example

In this real case study, the categorical target variable is Tl (Thallium) geochemical concentration transformed into five categories through quantiles and observed at 568 locations over the study region in southwest England (Kirkwood et al., 2016). Predictor variables include elevation, gravity, magnetic, Landsat, radiometric, and their derivatives, totaling 26 predictor variables. Some predictor variables are displayed in Fig. 7. Fig. 8a shows the categorical target variable's observations. The observations are partitioned into a training set (~ 75%) and testing set (~ 25%) as shown in Fig. 8b. The testing set is built such that all the categories have roughly the same number of observations.

Fig. 9 shows the categorical target variable's observed values at training locations and those predicted by the classical random forest and support vector machines. There is a considerable disagreement between the observed values and the predicted values of categorical target variable at training locations. The misclassification rate in the training data is 33.80% and 52.58%, respectively, for the classical random forest and support vector machines. For the classical random forest, the number of decisions trees has been set to 10 000, and the hyper-parameters have been optimized through cross-validation. Regarding the support vector machines, the kernel function and the hyper-parameters have been selected using cross-validation.

The learned conventional random forest model consists of an ensemble of  $B = 10\,000$  classification tree predictors. This latter ensemble is transformed into an ensemble of unconditional signed distance functions for each category, following by PCA and randomized quadratic programming as described in the methodology section (Sect. 2). This results to unconditional and conditional PC scores as shown in Fig. 10;  $T = 1000$  conditional PC scores are generated. As mentioned in the simulated data example, the points cloud of conditional PC scores is less spread out than those from unconditional PC scores because of the exact conditioning.

Prediction maps provided by the traditional classification random forest, support vector machines, regression-kriging of indicators, and proposed classification random forest with exact conditioning are depicted in Fig. 11. The prediction map provided by each method differs notably. In particular, the prediction map of the proposed classification random forest is different from the one provided by the traditional classification random forest. This is explained by the exact conditioning nature of the proposed method. It is important to highlight that the proposed classification random forest with exact conditioning uses the same ensemble of decision trees generated by the traditional classification random forest. Although the regression-kriging of indicators provides exact conditioning, its prediction map shows a noisier spatial distribution of categories (Fig. 11c). In contrast, the prediction map provided by the proposed classification random forest with exact conditioning depicts more regular and continuous contours (Fig. 11d) and is consistent with the training data shown in Fig. 9c.

Fig. 12 presents the prediction uncertainty (entropy) map under the traditional classification random forest, support vector machines, regression-kriging of indicators, and proposed classification random forest with exact conditioning. The prediction uncertainty map resulting from the proposed classification random forest with exact conditioning differs significantly from the others. In particular, the prediction uncertainty map provided by the traditional and proposed classification random forest differ notably due to the exact conditioning in the proposed method and not in the conventional method. Under the proposed classification random forest, local neighborhoods dominated by only one category show lower entropy than local neighborhoods dominated by several categories of the target variable.

Table 3 provides the predictive performance of the traditional classification random forest, support vector machines, regression-kriging of indicators, and proposed classification random forest with exact conditioning in the testing dataset (containing 140 observations). The proposed classification random forest shows better predictive performance than the three other methods according to the accuracy. The cost of not using the proposed classification random forest in this case is not negligible. There is an accuracy improvement of 25% and 9% respectively, compared to the regression-kriging of indicators and traditional classification random forest. Thus, the proposed approach can exactly match the categorical target variable's observed values at sampling locations while achieving good out of sample predictive performance.

#### 4. Conclusion

This paper proposed a classification random forest-based method for the spatial prediction of categorical variables in which the categorical target variable is exactly conditioned to the data. The exact conditioning means that the predicted values of the categorical target variable at sampling locations are the same as those observed at sampling locations. This property is well-known in geostatistical methods. The proposed method combines classification random forest, signed distance functions, principal component analysis, and randomized quadratic programming to achieve the exact conditioning of the categorical target variable to the data. The effectiveness of the proposed method has been demonstrated on simulated and real datasets.

Typical characteristics of the proposed method are the following. It can perfectly match the categorical target variable's observed values at

sampling locations while achieving good out-of-sample predictive performance compared to competitor geostatistical methods such as regression-kriging of indicators. It is easy to implement since it combines well-known existing statistical and machine learning methods. It can easily handle a large number of categories consistently through the signed distance representation. The proposed method can provide realistic prediction uncertainties of the categorical target variable. It has the advantage of not producing noisy spatial prediction maps, as one can observe for regression-kriging of indicators. The proposed method is free of the inherent problems of regression-kriging of indicators such as the predicted probabilities of the target variable's categories that are not guaranteed to belong to the  $[0, 1]$  interval and sum up to one. Updating the categorical target variable predictive map when few observations are added can be carried out quickly. Only the last part of the proposed method, i.e., the randomized quadratic programming, should be performed.

The proposed method is computationally intensive compared to regression-kriging of indicators and conventional classification random forest. However, it comprises components that can be performed in parallel according to the target variables' categories, including conditional principal component scores generation. The proposed method requires that the number of unconditional classification tree predictors should be large enough to allow good coverage of the solution space when performing the exact conditioning. Indeed, the number of categorical target variable's observations defines the number of inequalities constraints. The larger is the number of unconditional classification tree predictors, the wide is the solution space for the exact conditioning. So, too many constraints (hard data) relative to too few unconditional classification tree predictors will lead to too small uncertainty. Nonetheless, it will always be possible to meet this constraint because the number of unconditional classification tree predictors is a free parameter. Although the proposed method uses the random forest as the base learner, it can be used with other ensemble machine learning methods (e.g., boosting).

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### References

- Albrecht, T., González-Álvarez, I., Klump, J., 2021. Using machine learning to map Western Australian landscapes for mineral exploration. *ISPRS Int. J. Geo-Inf.* 10.
- Allard, D., D'Or, D., Froidevaux, R., 2011. An efficient maximum entropy approach for categorical variable prediction. *Eur. J. Soil Sci.* 62, 381–393.
- Bogaert, P., 2004. Spatial prediction of categorical variables: the bme approach. In: Sanchez-Vila, X., Carrera, J., Gómez-Hernández, J.J. (Eds.), *geoENV IV — Geostatistics for Environmental Applications*. Springer Netherlands, Dordrecht, pp. 271–282.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Cao, G., Kyriakidis, P., Goodchild, M., 2011. A multinomial logistic mixed model for prediction of categorical spatial data. *Int. J. Geogr. Inf. Sci.* 25, 2071–2086.
- Cao, G., Yoo, E.H., Wang, S., 2014. A statistical framework of data fusion for spatial prediction of categorical variables. *Stoch. Environ. Res. Risk Assess.* 28, 1785–1799.
- Chiles, J.P., Delfiner, P., 2012. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons.
- Cracknell, M., Reading, A., 2015. Spatial-contextual supervised classifiers explored: a challenging example of lithostratigraphic classification. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* 8, 1–14.
- Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* 63, 22–33.
- Davies, E., 2012. Chapter 9 - binary shape analysis. In: Davies, E. (Ed.), *Computer and Machine Vision*, fourth ed. Academic Press, Boston, pp. 229–265.
- Du, P., Bai, X., Tan, K., Xue, Z., Samat, A., Xia, J., Li, E., Su, H., Liu, W., 2020. Advances of four machine learning methods for spatial data handling: a review. *J. Geovisual. Spatial Anal.* 4.
- Fouedjio, F., 2020. Exact conditioning of regression random forest for spatial prediction. *Artifi. Intel. Geosci.* 1, 11–23.
- Fouedjio, F., Scheidt, C., Yang, L., Achtziger-Zupančič, P., Caers, J., 2021a. A geostatistical implicit modeling framework for uncertainty quantification of 3D

- geo-domain boundaries: application to lithological domains from a porphyry copper deposit. *Comput. Geosci.* 157, 104931.
- Fouedjio, F., Scheidt, C., Yang, L., Wang, Y., Caers, J., 2021b. Conditional simulation of categorical spatial variables using Gibbs sampling of a truncated multivariate normal distribution subject to linear inequality constraints. *Stoch. Environ. Res. Risk Assess.* 35, 457–480.
- Giaccone, E., Oriani, F., Tonini, M., Lambiel, C., Mariéthoz, G., 2021. Using data-driven algorithms for semi-automated geomorphological mapping. *Stoch. Environ. Res. Risk Assess.* 1–17.
- Goldfarb, D., Idnani, A., 1983. A numerically stable dual method for solving strictly convex quadratic programs. *Math. Program.* 27, 1–33.
- Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma* 103, 3–26.
- Grevera, G.J., 2007. Distance Transform Algorithms and Their Implementation and Evaluation. Springer New York, New York, NY, pp. 33–60.
- Hengl, T., Heuvelink, G.B., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120, 75–93.
- Hengl, T., Nussbaum, M., Wright, M., Heuvelink, G., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6, e5518.
- Hengl, T., Toomarian, N., Reuter, H.I., Malakouti, M.J., 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma* 140, 417–427. *Pedometrics* 2005.
- Kanevski, M., 2008. Advanced Mapping of Environmental Data: Geostatistics, Machine Learning and Bayesian Maximum Entropy. John Wiley & Sons.
- Kanevski, M., Pozdnoukhov, A., Timonin, V., 2009. Machine Learning for Spatial Environmental Data: Theory, Applications, and Software. EPFL press.
- Kirkwood, C., Everett, P., Ferreira, A., Lister, B., 2016. Stream sediment geochemistry as a tool for enhancing geological understanding: an overview of new data from south west England. *J. Geochem. Explor.* 163, 28–40.
- Kuhn, S., Cracknell, M.J., Reading, A.M., 2018. Lithologic mapping using random forests applied to geophysical and remote-sensing data: a demonstration study from the Eastern Goldfields of Australia. *Geophysics* 83, B183–B193.
- Kumar, C., Chatterjee, S., Oommen, T., Guha, A., 2020. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using aviris-ng hyperspectral data in gold-bearing granite-greenstone rocks in Huttı, India. *Int. J. Appl. Earth Obs. Geoinf.* 86, 102006.
- Latifovic, R., Pouliot, D., Campbell, J., 2018. Assessment of convolution neural networks for surficial geology mapping in the South Rae geological region, Northwest territories, Canada. *Rem. Sens.* 10.
- Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Rem. Sens.* 39, 2784–2817.
- Othman, A.A., Gloaguen, R., 2017. Integration of spectral, spatial and morphometric data into lithological mapping: a comparison of different machine learning algorithms in the Kurdistan region, NE Iraq. *J. Asian Earth Sci.* 146, 90–102.
- Pardo-Igúzquiza, E., Dowd, P.A., 2005. Multiple indicator cokriging with application to optimal sampling for environmental monitoring. *Comput. Geosci.* 31, 1–13.
- Probst, P., Wright, M., Boulesteix, A.L., 2018. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplin. Rev.: Data Min. Knowl. Discov.* <https://doi.org/10.1002/widm.1301>.
- R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Renard, D., Bez, N., Desassis, N., Beucher, H., Ors, F., Freulon, X., 2020. RGeostats: geostatistical package. URL: <http://cg.ensmp.fr/rgeostats>. r package version 12.0.1.
- Sahoo, S., Jha, M.K., 2017. Pattern recognition in lithology classification: modeling using neural networks, self-organizing maps and genetic algorithms. *Hydrogeol. J.* 25, 311–330.
- Wellmann, J.F., Regenauer-Lieb, K., 2012. Uncertainties Have a Meaning: Information Entropy as a Quality Measure for 3-D Geological Models. *Tectonophysics*, vols. 526–529. Modelling in Geosciences, pp. 207–216.
- Wright, M.N., Ziegler, A., 2017. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Software* 77, 1–17.
- Yu, L., Porwal, A., Holden, E.J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. *Comput. Geosci.* 45, 229–239.