



Merging machine learning and geostatistical approaches for spatial modeling of geoenergy resources



Gamze Erdogan Erten ^{a,*}, Oktay Erten ^a, C. Özgen Karacan ^b, Jeff Boisvert ^a, Clayton V. Deutsch ^a

^a Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, Canada

^b U.S. Geological Survey, Geology, Energy & Minerals Science Center, Reston, VA, USA

ARTICLE INFO

Keywords:

Ordinary kriging
Ensemble super learner
Ordinary intrinsic collocated cokriging
Elliptical radial basis neural network

ABSTRACT

Geostatistics is the most commonly used probabilistic approach for modeling earth systems, including quality parameters of various geoenergy resources. In geostatistics, estimates, either on a point or block support, are generated as a spatially-weighted average of surrounding samples. The optimal weights are determined through the stationary variogram model which accounts for the spatial structure of the samples. Recently, efficient modeling workflows using various machine learning algorithms (MLAs) have been expanded to the spatial context for modeling geological heterogeneity. The flexible use of MLAs as a spatial estimation tool stems mainly from the fact that unlike kriging, they do not require any variogram, nor do they depend strongly on a prior stationarity assumption (i.e., second order stationarity). This study evaluates the performance of two MLAs (ensemble super learner and elliptical radial basis neural network), ordinary kriging, and hybrid spatial modeling approaches using ordinary intrinsic collocated cokriging. The aforementioned modeling techniques are compared for estimating resources for four coal variables (wash yield, ash yield, calorific value and thickness) as an example. The results suggest that MLAs, when implemented alone, do not outperform ordinary kriging, but the estimation accuracy of the final model, measured by the root mean squared error tends to subtly improve (1.7% for wash yield, 6.98% for ash yield, 4.94% for calorific value and 0.36% for seam thickness) when MLAs and geostatistical algorithms are merged through the hybrid spatial modeling approaches.

1. Introduction

Geoenergy resources, such as coal and shales, have spatially varying physical and chemical properties, which should be estimated as accurately as possible to evaluate in-situ resources and to plan their production and extraction, as well as designing energy and industrial utilization processes optimally. Geostatistics, which relies on the concept of random function (Rossi and Deutsch, 2013; Chilès and Delfiner, 2012), has often been used to spatially estimate the coal properties in two- or three-dimensional spaces (Jeuken et al., 2020; Ertunç et al., 2013; Tercan et al., 2013; Webber et al., 2013; Heriawan and Koike, 2008), and to assess their geological uncertainties through multiple realizations generated by conditional simulation (Karacan and Olea, 2018; Olea and Luppens, 2015; Geboy et al., 2013; Pardo-Igúzquiza et al., 2013; Heriawan and Koike, 2008b).

In a typical geostatistical estimation context, the kriging algorithms require a stationarity assumption which allows inference of the stationary covariance (or auto-covariance), correlogram (auto-correlation)

or variogram. The optimal weights required by kriging can only be calculated through the variogram or covariance values at any lag distance obtained from the positive definite functions (i.e., spherical, exponential, Gaussian, hole-effect) fitted to the experimental variogram or covariance. Furthermore, there is usually a nonuniform smoothing in kriging estimates, where the degree of smoothness tends to increase at estimated locations away from the conditioning data (Deutsch, 2003).

Due to the aforementioned stationarity requirement and the nonuniform smoothing of kriging, machine learning methods (James et al., 2013; Kanevski, 2013; Hastie et al., 2009; Kanevski, 2009) have become popular as an alternative approach for classification and spatial modeling of coal variables (Chatterjee et al., 2022; Ibrahim, 2022; Maxwell et al., 2021; Tiwary et al., 2020; Maxwell et al., 2019). The traditional machine learning algorithms (MLAs) make use of a linear or non-linear relationship between the response variable and explanatory variables, where the data are assumed to be independent and identically distributed; in other words, each sample has the same probability distribution. However, earth science data almost always exhibit a spatial

* Corresponding author.

E-mail address: erdogane@ualberta.ca (G.E. Erten).

auto-correlation among the samples, which violates the aforementioned assumption. Therefore, there are two major concerns in relation to the use of MLAs as a spatial estimation tool: (1) the MLAs do not account for the spatial dependency among the samples, and (2) the estimates generated by the MLAs do not reproduce the data values at their locations.

Although there are many applications of MLAs as a spatial estimation tool, there are few studies which address both of the aforementioned shortcomings of the MLAs algorithms. For example, the studies from geography and soil sciences (Talebi et al., 2022; Georganos et al., 2021; Behrens et al., 2018; Hengl et al., 2018) deal with only the first concern by proposing different methodologies to account for the spatial dependency among the samples in specific MLAs, while the studies from environmental sciences (Fouedjio, 2021; Fouedjio, 2020) address only the second concern by ensuring the exact data reproduction in the MLA-based estimation workflow for modeling continuous and categorical variables.

Studies from mining, on the other hand, (da Silva et al., 2022; Erdogan Erten et al., 2022; Samson and Deutsch, 2022; Erdogan Erten et al., 2021) address both above-mentioned concerns by making use of the capabilities of both MLAs and geostatistics within the hybrid spatial modeling frameworks.

The objective of this study is to evaluate the performances of two MLAs (namely ensemble super learner (ESL) and elliptical radial basis neural network (ERBFN)) and hybrid spatial modeling approaches, and compare their efficacy against ordinary kriging (OK) for spatial estimation of the attributes of geoenergy resources. We demonstrate each model's performance in a case study modeling four spatially varying attributes of Pocahontas No. 3 coal. Furthermore, we demonstrate how spatial dependency and data reproduction issues can be resolved through optimal combination of the models generated by MLAs and geostatistics.

2. Review of theory and application of the utilized methods

2.1. Geostatistical modeling approaches

Consider that Z is a coal variable (or any geoenergy variable) that is modeled over the study area D using a second-order stationary random function $Z(\mathbf{u})$, $\{Z(\mathbf{u}) | \mathbf{u} \in D \subseteq \mathbb{R}^p\}$. Let $\{z(\mathbf{u}_i), i = 1, \dots, n\}$ be the data which is a partial realization of $Z(\mathbf{u})$. The second-order stationary $Z(\mathbf{u})$ implies that the mean, m and variance, σ^2 are, on average, constant for all locations within D :

$$E[Z(\mathbf{u})] = m, \quad \forall \mathbf{u} \in D \quad (1)$$

$$E[(Z(\mathbf{u}) - m)^2] = C(0) = \sigma^2, \quad \forall \mathbf{u} \in D \quad (2)$$

The covariance $C(\mathbf{h})$ does not depend on a spatial location \mathbf{u} , but depends only on the separation lag \mathbf{h} which is a vectorial distance in the sampling space (Chilès and Delfiner, 2012; Olea, 1999; Deutsch and Journel, 1998):

$$E[(Z(\mathbf{u}) \cdot Z(\mathbf{u} + \mathbf{h}))] - m^2 = C(\mathbf{u}, \mathbf{u} + \mathbf{h}) = C(\mathbf{h}), \quad \forall \mathbf{u}, \mathbf{h}, \mathbf{u} + \mathbf{h} \in D \quad (3)$$

If the mean cannot be assumed to be constant across D , one can make an intrinsic stationary assumption (Matheron, 2023) which implies that the expected differences are zero; that is, the increments of $Z(\mathbf{u})$ are second-order stationary:

$$E[Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})] = 0 \quad (4)$$

If the covariance of the residuals is replaced by the variance of the differences, the resulting equation (Eq. (5)) is the variogram which is the measure of spatial relations:

$$\text{Var}[Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})] = E[(Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h}))^2] = 2\gamma(\mathbf{h}) \quad (5)$$

where $\gamma(\mathbf{h})$ denotes a semivariance at lag \mathbf{h} . Provided that the random

function model is second-order stationary, the covariance function, $C(\mathbf{h})$ is linearly related to the variogram, $\gamma(\mathbf{h})$.

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}), \quad \forall \mathbf{h} \in D \quad (6)$$

In practice, the experimental variogram is estimated from the sample data using Matheron's method of moments estimator:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2 \cdot N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z(\mathbf{u}_i) - Z(\mathbf{u}_i + \mathbf{h})]^2 \quad (7)$$

where $Z(\mathbf{u}_i)$ and $Z(\mathbf{u}_i + \mathbf{h})$ are the samples at locations \mathbf{u}_i and $\mathbf{u}_i + \mathbf{h}$ and $N(\mathbf{h})$ is the number of pairs at lag \mathbf{h} .

2.1.1. Ordinary Kriging

The ordinary kriging (OK) estimator $Z_{OK}^*(\mathbf{u}_0)$ is given by (Pyrcz and Deutsch, 2014):

$$Z_{OK}^*(\mathbf{u}_0) = \sum_{i=1}^n \lambda_i^{OK}(\mathbf{u}_0) \cdot Z(\mathbf{u}_i) \quad (8)$$

where $\lambda_i^{OK}(\mathbf{u}_0)$ is the weight assigned to the i^{th} conditioning datum $z(\mathbf{u}_i)$. To ensure the unbiasedness ($E[Z_{OK}^*(\mathbf{u}_0) - Z(\mathbf{u}_0)] = 0$), the estimation variance is defined by:

$$\begin{aligned} \text{Var}[Z_{OK}^*(\mathbf{u}_0)] &= E[(Z_{OK}^*(\mathbf{u}_0) - Z(\mathbf{u}_0))^2] \\ &= 2 \sum_{i=1}^n \lambda_i^{OK}(\mathbf{u}_0) \gamma(\mathbf{u}_i - \mathbf{u}_0) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{OK}(\mathbf{u}_0) \lambda_j^{OK}(\mathbf{u}_0) \gamma(\mathbf{u}_i - \mathbf{u}_j) \end{aligned} \quad (9)$$

The minimization of the estimation variance (Eq. (9)) under the constraint on the sum of weights results in the OK system:

$$\sum_{i=1}^n \lambda_i^{OK}(\mathbf{u}_0) \gamma(\mathbf{u}_i - \mathbf{u}_j) + \mu(\mathbf{u}_0) = \gamma(\mathbf{u}_i - \mathbf{u}_0), \quad j = 1, \dots, n \quad (10)$$

where $\gamma(\mathbf{u}_i - \mathbf{u}_j)$ are the variogram values between the data; $\gamma(\mathbf{u}_i - \mathbf{u}_0)$ are the variogram values between the data and the variable at the location being estimated. The estimation variance, $\sigma_{OK}^2(\mathbf{u}_0)$ is defined by:

$$\sigma_{OK}^2(\mathbf{u}_0) = \sum_{i=1}^n \lambda_i^{OK}(\mathbf{u}_0) \gamma(\mathbf{u}_i - \mathbf{u}_0) + \mu(\mathbf{u}_0) \quad (11)$$

As can be seen from Eq. (11), the estimation variance (or kriging error) depends only on the spatial configuration of the samples and its calculation is independent of the values of the conditioning data.

2.1.2. Ordinary Intrinsic Collocated Cokriging (OICCK)

OICCK (Babak and Deutsch, 2009a), which is a variant of ordinary cokriging, makes use of the exhaustive secondary information which are retained both at the locations of the primary data and at the location being estimated. In such a multi-collocated neighborhood (Wackernagel, 2003), the variance can be reproduced more accurately, and variance inflation does not occur (Babak and Deutsch, 2009b; Babak and Deutsch, 2009c). To ensure the unbiasedness of the estimator, the OICCK system uses only one constraint where sum of all weights is equal to one; therefore, both primary and secondary variables should be standardized, with a mean of zero and a unit variance prior to estimation. The OICCK system can then be defined in terms of correlograms; that is, $\rho(\mathbf{h}) = C(\mathbf{h})/C(0)$ (Deutsch, 2003; Goovaerts, 1997).

Consider that Z is a primary variable that is sampled at n locations, and Y is a secondary variable whose values are available at all m grid nodes, where $m \gg n$. The OICCK estimator at the location \mathbf{u}_0 is defined by:

$$Z_{OICCK}^*(\mathbf{u}_0) = \sum_{i=1}^n \lambda_{z,i}(\mathbf{u}_0) \cdot Z(\mathbf{u}_i) + \sum_{i=1}^m \lambda_{y,i}(\mathbf{u}_0) \cdot Y(\mathbf{u}_i) + \lambda_{y,0}(\mathbf{u}_0) \cdot Y(\mathbf{u}_0) \quad (12)$$

where $\lambda_{z,i}$, $\lambda_{y,i}$, and $\lambda_{y,0}$ are the weights assigned to the collocated i^{th} primary and secondary data $\{z(\mathbf{u}_i), y(\mathbf{u}_i)\}$, $i = 1, \dots, n$, and the secondary datum $y(\mathbf{u}_0)$ at the location being estimated, respectively. The unbiasedness constraint implies that all of the aforementioned weights must sum to one:

$$\sum_{i=1}^n \lambda_{z,i}(\mathbf{u}_0) + \sum_{i=1}^n \lambda_{y,i}(\mathbf{u}_0) + \lambda_{y,0}(\mathbf{u}_0) = 1 \quad (13)$$

The following system of linear equations then yields the weights:

$$\begin{aligned} & \sum_{i=1}^n \lambda_{z,i}(\mathbf{u}_0) \cdot \rho_{zz}(\mathbf{u}_i - \mathbf{u}_j) + \sum_{i=1}^n \lambda_{y,i}(\mathbf{u}_0) \cdot \rho_{yz}(\mathbf{u}_i - \mathbf{u}_j) \\ & + \lambda_{y,0}(\mathbf{u}_0) \cdot \rho_{yz}(\mathbf{u}_j - \mathbf{u}_0) + \mu(\mathbf{u}_0) = \rho_{zz}(\mathbf{u}_j - \mathbf{u}_0), \quad j = 1, \dots, n \\ & \sum_{i=1}^n \lambda_{z,i}(\mathbf{u}_0) \cdot \rho_{yz}(\mathbf{u}_i - \mathbf{u}_j) + \sum_{i=1}^n \lambda_{y,i}(\mathbf{u}_0) \cdot \rho_{yy}(\mathbf{u}_i - \mathbf{u}_j) \\ & + \lambda_{y,0}(\mathbf{u}_0) \cdot \rho_{yy}(\mathbf{u}_j - \mathbf{u}_0) + \mu(\mathbf{u}_0) = \rho_{yz}(\mathbf{u}_j - \mathbf{u}_0), \quad j = 1, \dots, n \\ & \sum_{i=1}^n \lambda_{z,i}(\mathbf{u}_0) \cdot \rho_{yz}(\mathbf{u}_i - \mathbf{u}_0) + \sum_{i=1}^n \lambda_{y,i}(\mathbf{u}_0) \cdot \rho_{yy}(\mathbf{u}_i - \mathbf{u}_0) \\ & + \lambda_{y,0}(\mathbf{u}_0) + \mu(\mathbf{u}_0) = \rho_{yz}(0) \end{aligned} \quad (14)$$

where $\rho_{zz}(\mathbf{u}_i - \mathbf{u}_j)$ and $\rho_{zz}(\mathbf{u}_j - \mathbf{u}_0)$ are the spatial correlations between primary data, and between the primary data and the primary variable at the location being estimated, respectively; $\rho_{yz}(\mathbf{u}_i - \mathbf{u}_j)$ and $\rho_{yz}(\mathbf{u}_i - \mathbf{u}_0)$ are the spatial cross-correlations between the primary and secondary data, and between the primary variable at the location being estimated and secondary data, respectively; $\rho_{yy}(\mathbf{u}_i - \mathbf{u}_j)$ and $\rho_{yy}(\mathbf{u}_i - \mathbf{u}_0)$ are the spatial correlations between the secondary data, and between the secondary data and the secondary datum available at the location being estimated, respectively, and $\rho_{yz}(0)$ is the pairwise correlation coefficient between the primary and secondary variables. The estimation variance, $\sigma_{OICCK}^2(\mathbf{u}_0)$ is defined by:

$$\begin{aligned} \sigma_{OICCK}^2(\mathbf{u}_0) = & \sum_{i=1}^n \lambda_{z,i}(\mathbf{u}_0) \cdot \rho_{zz}(\mathbf{u}_i - \mathbf{u}_0) + \sum_{i=1}^n \lambda_{y,i}(\mathbf{u}_0) \cdot \rho_{yz}(\mathbf{u}_i - \mathbf{u}_0) \\ & + \lambda_{y,0}(\mathbf{u}_0) \cdot \rho_{yz}(0) + \mu(\mathbf{u}_0) \end{aligned} \quad (15)$$

In order to simplify the OICCK implementation, a Markov assumption of conditional independence can be used. There are two Markov assumptions to consider: (1) Markov model 1 (MM1) (Almeida and Journel, 1994), and (2) Markov model 2 (MM2) (Journel, 1999). The cross-correlogram model given by MM1 is defined by:

$$\rho_{zy}(\mathbf{h}) \simeq \rho_{zy}(0) \cdot \rho_{zz}(\mathbf{h}) \quad (16)$$

where $\rho_{zz}(\mathbf{h})$ is the correlogram model of the primary variable; $\rho_{zy}(\mathbf{h})$ is the cross-correlogram model between the primary and secondary variables, and $\rho_{zy}(0)$ is the pairwise correlation coefficient. MM1 does not require the correlogram model of the secondary variable $\rho_{yy}(\mathbf{h})$, as it is assumed to be equal to that of the primary variable; that is, $\rho_{yy}(\mathbf{h}) = \rho_{zz}(\mathbf{h})$. In addition, the shape and continuity of the experimental cross-correlogram $\hat{\rho}_{zy}(\mathbf{h})$ are assumed to be identical to those of the experimental correlogram of the primary variable $\hat{\rho}_{zz}(\mathbf{h})$.

The cross-correlogram model given by MM2 is defined by:

$$\rho_{zy}(\mathbf{h}) \simeq \rho_{zy}(0) \cdot \rho_{yy}(\mathbf{h}) \quad (17)$$

The correlogram model of the primary variable is then calculated by:

$$\rho_{zz}(\mathbf{h}) \simeq \rho_{zy}^2(0) \cdot \rho_{yy}(\mathbf{h}) + \left(1 - \rho_{zy}^2(0)\right) \cdot \rho_r(\mathbf{h}) \quad (18)$$

where $\rho_r(\mathbf{h})$ is the residual term so that the correlogram model of the primary variable will be approximately equal to the experimental correlogram of the primary variable; that is, $\rho_{zz}(\mathbf{h}) \approx \hat{\rho}_{zz}(\mathbf{h})$.

2.2. Machine learning-based modeling approaches

2.2.1. Ensemble Super Learner (ESL)

The ESL approach (Erdogan Erten et al., 2022; Erdogan Erten et al., 2021), which employs the super learner model (SLM) (Van der Laan et al., 2023) to construct an estimation model, consists of three main implementation steps: (i) coordinate rotations of the data and grid nodes being estimated, with an incrementally increasing predefined angle (i.e., from 10° to 90° with an increment of 10°), (ii) using each rotated dataset, fit a SLM and make predictions, and (iii) calculating an ensemble prediction by averaging the individual predictions.

In the first step, the system of coordinate rotation given in geostatistical software library (GSLIB) convention (Deutsch and Journel, 1998) is employed. In two-dimensional orientation, the angle (or azimuth) α measured clockwise from the Northing (y-axis), while keeping the z-axis fixed defines the rotation. The vector \mathbf{U} contains the values of the rotated Easting, x' and Northing, y' coordinates (Eq. (19)).

$$\mathbf{U} = \begin{bmatrix} x' \\ y' \\ z \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (19)$$

In the second step, the SLM, which considers the commonly implemented MLAs including support vector regressor (Cortes and Vapnik, 1995), gradient boosting regressor (Friedman, 2001), k -neighbors regressor (Cover and Hart, 1967), random forest regressor (Breiman, 2001), bagging regressor (Breiman, 1996), neural network (Haykin, 2009), extra tree regressor (Geurts et al., 2006), decision tree regressor (Quinlan, 1986), and adaboost regressor (Freund and Schapire, 1996), are employed to create an estimation model for each rotated dataset. The hyperparameters of each MLA are tuned using the grid search algorithm with a 5-fold cross validation (Erdogan Erten et al., 2022).

Finally, the SLM performs a weighted model averaging using the following steps. First, each MLA (M_1, \dots, M_K) is trained using all training data, where K is the number of MLAs used, and a $(n_{\text{training}} \times K)$ matrix \mathbf{Z}' (Eq. (20)) is obtained, whose elements \hat{Y}_{ij} with $i = 1, \dots, n_{\text{training}}$ and $j = 1, \dots, K$ are the estimates generated by each MLA.

$$\mathbf{Z}' = \begin{bmatrix} \hat{Y}_{11} & \dots & \hat{Y}_{1K} \\ \vdots & \ddots & \vdots \\ \hat{Y}_{n_{\text{training}}1} & \dots & \hat{Y}_{n_{\text{training}}K} \end{bmatrix} \quad (20)$$

Then, all training data are randomly split into V equal-sized partitions indexed by $v \in 1, \dots, V$. For each v , which describes two separate sets, training set $T(v)$ contains all of the data except those in a partition v , and corresponding validation set $V(v)$ contains the data in partition v . For each v , each MLA is trained on $T(v)$, and the estimates are made on $V(v)$, which yields a $(n_{\text{training}} \times K)$ matrix \mathbf{Z} (Eq. (21)), whose elements $\hat{Y}_{11}^v \dots \hat{Y}_{n_{\text{training}}K}^v$ with $i = 1, \dots, n_{\text{training}}$ and $j = 1, \dots, K$ correspond to one estimated value for each observation for each MLA.

$$\mathbf{Z} = \begin{bmatrix} \hat{Y}_{11}^v & \dots & \hat{Y}_{1K}^v \\ \vdots & \ddots & \vdots \\ \hat{Y}_{n_{\text{training}}/V1}^v & \dots & \hat{Y}_{n_{\text{training}}/VK}^v \\ \vdots & \ddots & \vdots \\ \hat{Y}_{11}^V & \dots & \hat{Y}_{11}^V \\ \vdots & \ddots & \vdots \\ \hat{Y}_{n_{\text{training}}/V1}^V & \dots & \hat{Y}_{n_{\text{training}}/VK}^V \end{bmatrix} \quad (21)$$

A family of weighted combinations of MLAs indexed by β , $m(z|\beta) = \sum_{k=1}^K \beta_k \hat{Y}_k$, is proposed subject to the constraints $\beta_k \geq 0 \forall k$ and $\sum_{k=1}^K \beta_k = 1$. The contribution of each MLA is computed using a meta learner to the final model $(\hat{\beta}_1, \dots, \hat{\beta}_K)$ by determining $\hat{\beta}$ that minimizes the cross-validated loss function (Eq. (22)).

$$\hat{\beta} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (Y_i - m(z_i | \beta))^2 \quad (22)$$

where Y_i is the training data and z_i is the vector of estimated values $(\hat{Y}_{i1}, \dots, \hat{Y}_{iK})$ for subject i from the i^{th} row of the estimation matrix \mathbf{Z} . The SLM estimates are finally generated by:

$$Y_{SL} = \sum_{k=1}^K \hat{\beta}_k z_k' \quad (23)$$

where $z_k' = (\hat{Y}_{11} \dots \hat{Y}_{n_{\text{training}} 1})$ is a column k of the \mathbf{Z}' estimation matrix.

2.2.2. Elliptical Radial Basis Function Neural Network (ERBFN)

A radial basis function network (RBFN) consists of three layers: an input layer, a hidden layer with a non-linear RBF activation function and an output layer. Activation functions in RBFNs are commonly implemented as Gaussian functions (Ghosh and Nag, 2001). Consider that a dataset D which has N patterns of (x, y) where x is the input of the dataset and y is the actual output. The output of the i^{th} activation function ϕ_i in the hidden layer is computed using the distance between the input pattern x and the center i , as given in Eq. (24):

$$\phi_i(\|x - c_i\|) = \exp(-r_i \|x - c_i\|^2) \quad (24)$$

where $\|\cdot\|$ is the Euclidean norm, c_i and r_i are the center and radius of hidden neuron i , respectively. The output of the node k of the output layer of the network is calculated as a linear summation (Eq. (25)):

$$y_k = \sum_{i=1}^n w_{ik} \phi_i(x) \quad (25)$$

where n is the number of neurons in the hidden layer and w_{ik} is the weight between i^{th} center and k^{th} output. RBFNs are commonly fitted in two steps: (i) the center and radius are determined using unsupervised clustering algorithms, and (ii) the connection weights between the hidden layer and the output layer are updated and calculated such that an error criterion (i.e., mean squared error) is minimized over the dataset.

ERBFN is similar to RBFN, but uses the Mahalanobis distance given in Eq. (26) instead of Euclidean distance in the Gaussian functions (Mak and Li, 1999):

$$d_m(x, S_i, c_i) = \sqrt{(x - c_i)^T S_i^{-1} (x - c_i)} \quad (26)$$

where S_i^{-1} is the inverse covariance matrix between c_i and x . Thus, ERBFN allows the RBF to take an anisotropic elliptical shape rather than a circular isotropic shape, which improves the performance of the RBFN (Beheim et al., 2004; Ibrilci et al., 2002).

Samson and Deutsch (Samson and Deutsch, 2022) implemented ERBFN for spatial estimations by using the Gaussian kernel as a kernel function. In this study, three Gaussian kernel parameters are considered: (i) the location c_i of each kernel center, (ii) the radius r of each kernel, and (iii) the number of Gaussian kernels. The parameter c_i is found using the k -mean clustering, and then the first set of locations are optimized through training of the network. The radius r is initially set to one, and then optimized throughout the training using the ADAM (adaptive estimates of lower-order moments) optimizer (Kingma and Ba, 2023). The last parameter, the number of Gaussian kernels, is of critical importance. Because, using too many kernels may cause overfitting, while using too few kernels may result in underfitting, which affect the overall performance of the estimation model.

To determine the ideal number of Gaussian kernels, (Samson and Deutsch, 2022) generated an ensemble of different node estimates. In this approach, the final estimations are defined as the mathematical averaging of the multiple estimations in the ensemble. The benefits of using an ensemble are to improve the average estimation performance

and remove artefacts due to overfitting or underfitting problems.

2.3. Hybrid spatial modeling approaches

2.3.1. Combination of machine learning algorithms and ordinary kriging

The model generated by ESL or ERBFN may be optimally merged with the model generated by OK for two main reasons: (i) the estimates yielded by ESL or ERBFN do not reproduce the data values at their locations, and (ii) the quality of the OK estimates that are further away from the conditioning data tends to be low due to the nonuniform smoothing effect.

The estimation variance-based weighting procedure, which is proposed by (Erdogan Erten et al., 2022), can be used to rectify the aforementioned shortcomings of MLAs and OK. For example, considering the estimation model generated by ESL, the combination of ESL and OK models is given by:

$$Z_{\text{Combined}}^*(\mathbf{u}) = w(\mathbf{u}) \cdot Z_{\text{ESL}}^*(\mathbf{u}) + (1 - w(\mathbf{u})) \cdot Z_{\text{OK}}^*(\mathbf{u}), \quad \forall \mathbf{u} \in D \quad (27)$$

where $Z_{\text{Combined}}^*(\mathbf{u})$ is the final estimate; $w(\mathbf{u})$ is the weight assigned to the ESL estimate, $Z_{\text{ESL}}^*(\mathbf{u})$, and $1 - w(\mathbf{u})$ is the weight assigned to the OK estimate, $Z_{\text{OK}}^*(\mathbf{u})$. $w(\mathbf{u})$ is calculated as a function of the standardized OK estimation variance, $\sigma_{\text{OK}}^2(\mathbf{u}) = [0, 1]$:

$$w(\mathbf{u}) = (\sigma_{\text{OK}}^2(\mathbf{u}))^b \quad (28)$$

where the exponent b , which regularizes $w(\mathbf{u})$ according to $Z_{\text{OK}}^*(\mathbf{u})$, can be defined by a linear equation:

$$b = b_0 + b_1 Z_{\text{OK}}^*(\mathbf{u}) \quad (29)$$

where the parameter b should be optimized according to the conditioning data, $\sigma_{\text{OK}}^2(\mathbf{u})$, $Z_{\text{ESL}}^*(\mathbf{u})$, and $Z_{\text{OK}}^*(\mathbf{u})$. The optimization problem can then be defined by the minimization of the root mean squared error (RMSE) which is calculated between the collocated conditioning data, $Z_{\text{True}}(\mathbf{u})$ and $Z_{\text{Combined}}^*(\mathbf{u})$, $\forall \mathbf{u} = \mathbf{u}_i$; that is,

$$\min \sqrt{\frac{1}{n} \sum_{i=1}^n [Z_{\text{True}}(\mathbf{u}_i) - Z_{\text{Combined}}^*(\mathbf{u}_i)]^2} \quad (30)$$

subject to : $a_1 \leq b_0 \leq c_1$

$$: a_2 \leq b_1 \leq c_2$$

where b_0 and b_1 are constrained by a_1 , c_1 , and a_2 , c_2 which are the lower and upper limits of the variables, respectively. The aforementioned optimization problem can be solved by the sequential quadratic programming (Yang, 2017).

2.3.2. Incorporating machine learning models through ordinary intrinsic collocated cokriging

OICCK (Section 2.1.2) can be another approach for addressing the shortcomings of MLAs and OK. In combining MLAs with OICCK, the model generated by ESL or ERBFN for an attribute is considered to be exhaustive secondary information, and the available conditioning data from the training set of the same attribute is the primary variable.

The standardized samples of the primary and secondary variables along with the correlogram model of the primary variable and the pairwise linear correlation coefficient between the primary and secondary variables are the required inputs to solve the OICCK system for each grid node to be estimated. It is noted that the inference of the cross-correlogram model between the primary and secondary variables can be made either by MM1 (Eq. (16)) or MM2 (Eq. (17)). After back-transformation of the standardized estimates to the original data unit, the final estimates in the resulting model reproduce the primary conditioning data at their locations, and reduces the degree of the smoothing effect (Samson and Deutsch, 2022).

3. Case study

3.1. Location of the study area and the Pocahontas No.3 coal

The studied area lies within the coalfield of southwestern Virginia in the Central Appalachian Basin (Fig. 1). The area is in Buchanan County and northeast of the southwest-northeast oriented Cumberland Over-thrust Block, which was subjected to detachment, transportation, and rotation along large faults during Alleghanian orogenesis (Henika and Dishner, 1994). The tectonic history of the Central Appalachian Basin affected its structure and its depositional and stratigraphic characteristics, which eventually affected coal beds of the field and their properties. From mining and other coal resources (i.e., coal bed methane) perspectives, the most important formations in this area are the lower Pennsylvanian aged Pocahontas and New River Formations. Along the southeast margin of the Virginia coalfield, the lower Pennsylvanian strata may attain a combined maximum thickness of approximately 899.16 m (Grimm et al., 2012).

The Pocahontas and New River Formations contain 15 to 20 individual coal beds of approximately 4.57 to 9.14 m of combined thickness a 304.8 m interval at depths ranging from 274.32 to 640.08 m. The coal beds are interbedded with other non-coal strata of varying thicknesses. These strata are medium to light gray, fine-grained sandstones and

medium to dark gray, carbonaceous siltstones and shales (Gilliland et al., 2013). Many of the coal beds in the Pocahontas and New River Formations were coal bed methane production targets for many years e.g., (Nolde and Spears, 1998; Nolde, 1994) using vertical wells and multiple-zone completions.

In the general area, the Pocahontas No. 3 coal is the thickest coal with medium to low-volatile bituminous rank. The Pocahontas No. 3 coal is also economically the most valuable due to its favorable properties for coke production, such as low ash and sulfur, high calorific value, and low trace-element impurities. Pocahontas No. 3 coal has been mined in Buchanan County since the late 1960s by different mining operations. Presently, it is being mined in the study area presented in this paper by a longwall operation.

3.2. Available data and preparation for modeling

The data used in this study, wash yield (%), ash yield (%), calorific value (Btu/lb) and seam thickness (m), were from Pocahontas No. 3 coal samples recovered from 324 exploration core holes, distributed in an area extending approximately 24 km in Easting and 15 km in Northing directions (Fig. 2). To be able to assess the estimation performances of ESL, ERBFN, OK and the hybrid spatial modeling approaches, an independent validation set (VS) is required. Therefore, the original data is

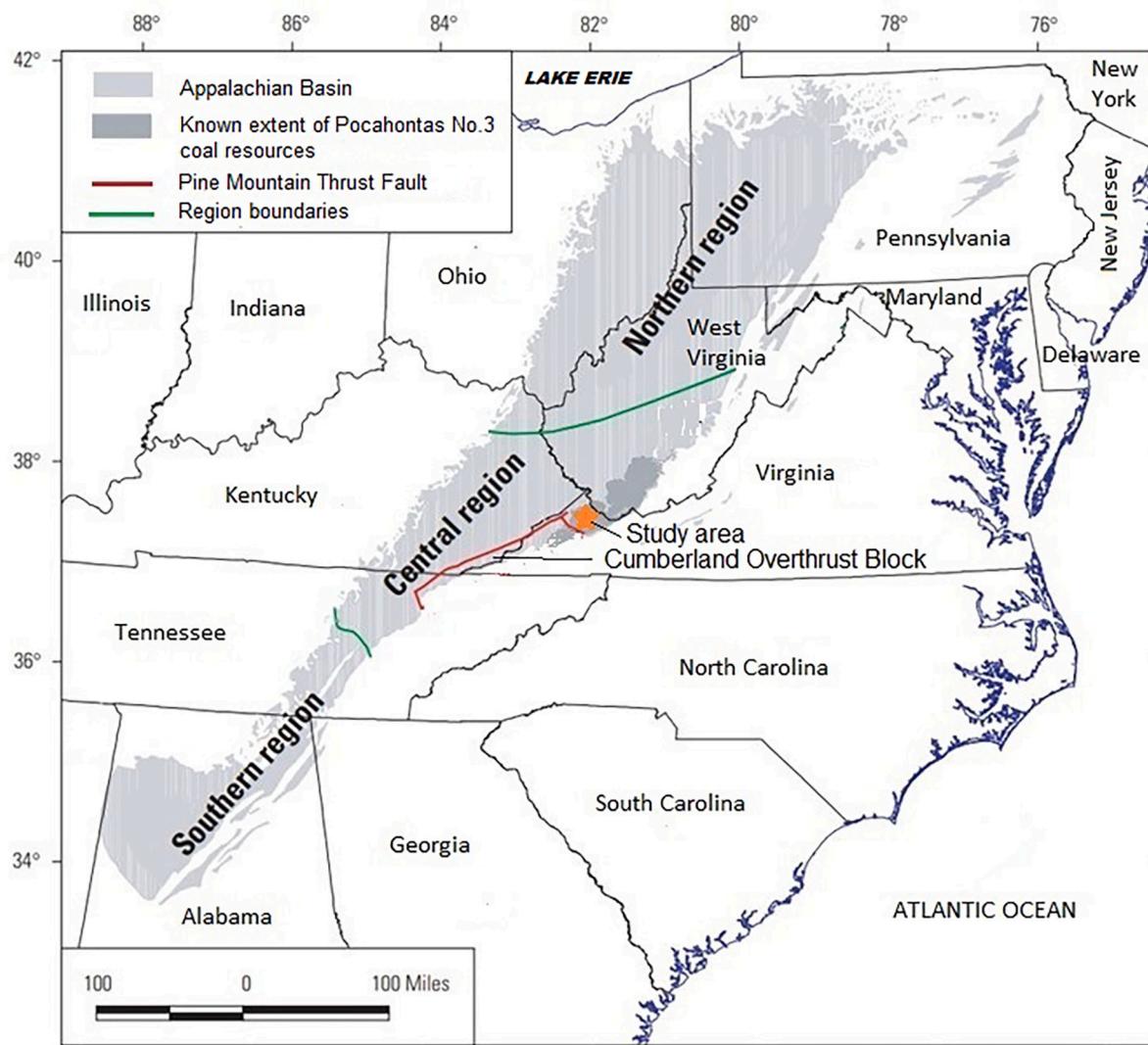


Fig. 1. General location of the study area in the Central Appalachian Basin (Figure modified from (Milici et al., 2000) (1 mile = 1.6093 km).

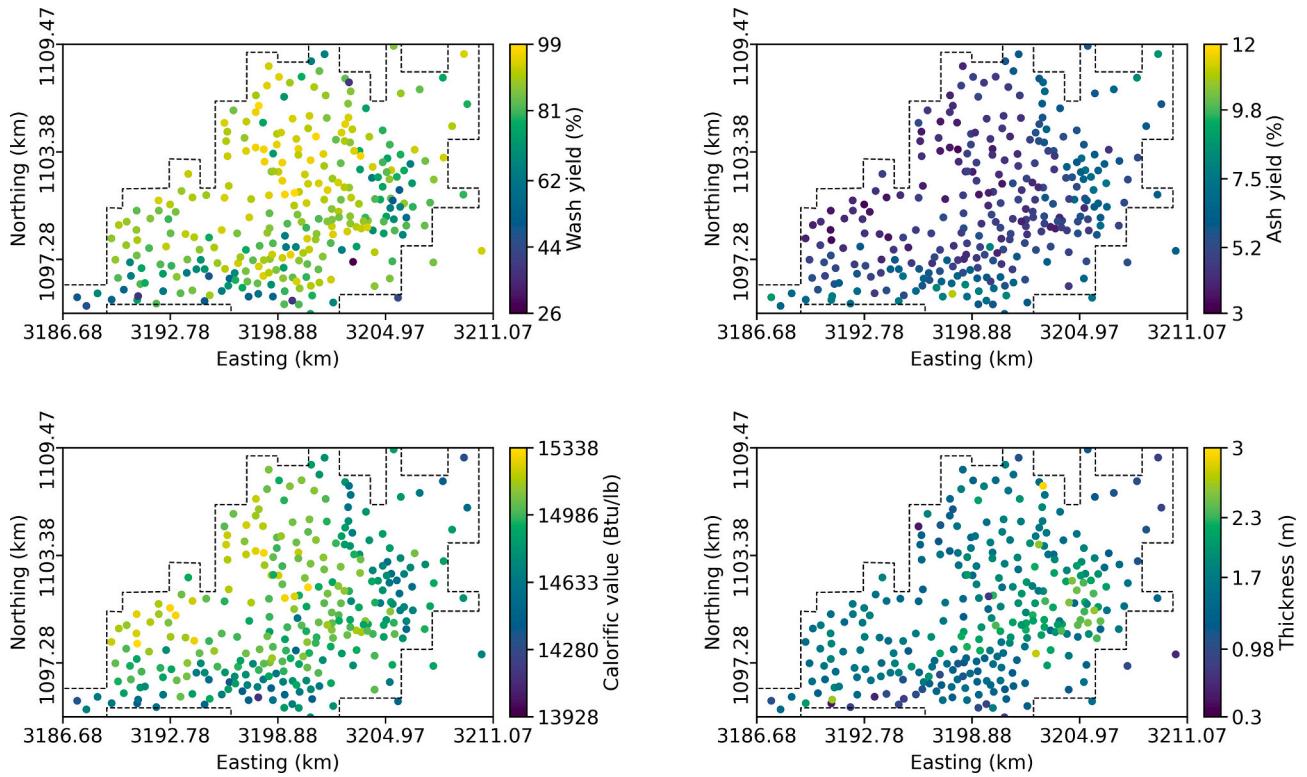


Fig. 2. Sampling locations of the variables in the original dataset (1 Btu/lb. = 0.556 kcal/kg).

randomly split into two subsets, such that 75% of the original data is considered to be the training set (TS), and the remaining 25% is considered to be VS. The sampling locations of the TS and VS are shown in Fig. 3. It can be seen from Fig. 3 that the samples of both TS and VS appear to uniformly cover the domain of interest, which is important in terms of assessing the accuracy of the generated models. The summary statistics of the original data, TS and VS are contained in Table 1. It can be observed from the summary statistics given in Table 1 that the samples of both TS and VS appear to be representative of the original data and that no bias is introduced when splitting the original data into two subsets.

3.3. Modeling of the variables through ordinary kriging

A decision should be made as to whether to estimate the coal variables directly or through the accumulation (i.e., ash value \times seam thickness \times specific gravity) (Journeel and Huijbregts, 1978). If the seam thickness does not significantly vary in the study area, and is not strongly correlated with other variables, then the error associated with

estimating the variables directly from the sample values is expected to be low. The scatter plots of the variables shown in Fig. 4 indicate that the seam thickness appears to be weakly correlated with all other variables. Therefore, the variables can be estimated directly.

The variogram maps for all of the variables do not reveal any distinct anisotropy direction; therefore, the experimental omni-directional variogram of each variable is calculated for 20 lags with the lag separation distance of 1.2 km and the lag tolerance of 0.6 km. A horizontal bandwidth is set to a large value (i.e., 3 km). Each experimental variogram is then fitted by an analytical model using a nugget variance and an isotropic spherical structure (Fig. 5). The parameters of the analytical models are given by:

$$\begin{aligned}\gamma_{\text{Wash yield}}(\mathbf{h}) &= 0.55 \cdot \text{Nug}(\mathbf{h}) + 0.45 \cdot \text{Sph}(\mathbf{h}; 7.3 \text{ km}) \\ \gamma_{\text{Ash Yield}}(\mathbf{h}) &= 0.35 \cdot \text{Nug}(\mathbf{h}) + 0.65 \cdot \text{Sph}(\mathbf{h}; 7.9 \text{ km}) \\ \gamma_{\text{Cal. Val.}}(\mathbf{h}) &= 0.35 \cdot \text{Nug}(\mathbf{h}) + 0.65 \cdot \text{Sph}(\mathbf{h}; 8.5 \text{ km}) \\ \gamma_{\text{Thickness}}(\mathbf{h}) &= 0.18 \cdot \text{Nug}(\mathbf{h}) + 0.82 \cdot \text{Sph}(\mathbf{h}; 9.1 \text{ km})\end{aligned}\quad (31)$$

The validity of the variogram models (Fig. 5 and Eq. (31)) is assessed using cross validation (Isaaks and Srivastava, 1989), where one sample is temporarily removed from the dataset, and it is re-estimated from the remaining data using OK with a competing analytical model. The criteria used to check the suitability of the analytical models fitted include RMSE, R² (coefficient of determination) and slope of regression. The RMSE metric is calculated by:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (z_i - z_i^*)^2}{n}} \quad (32)$$

where z_i is the i^{th} true value that is removed, and z_i^* is the re-estimate of the same value. Consider that the least squares regression line fitted to the pairwise values of the truth z_i , which is the dependent (or response) variable, and estimate z_i^* , which is the explanatory variable, is given by:

$$\hat{z}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot z_i^*, \quad i = 1, \dots, n \quad (33)$$

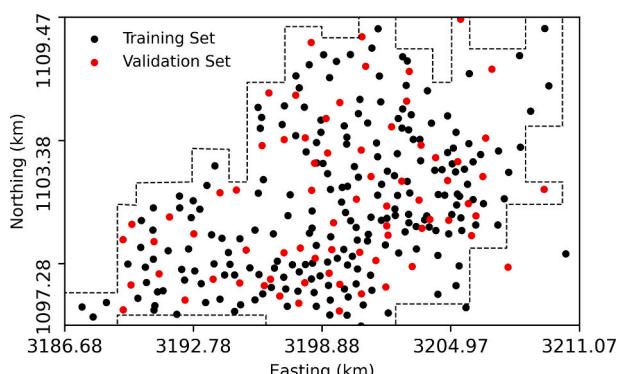
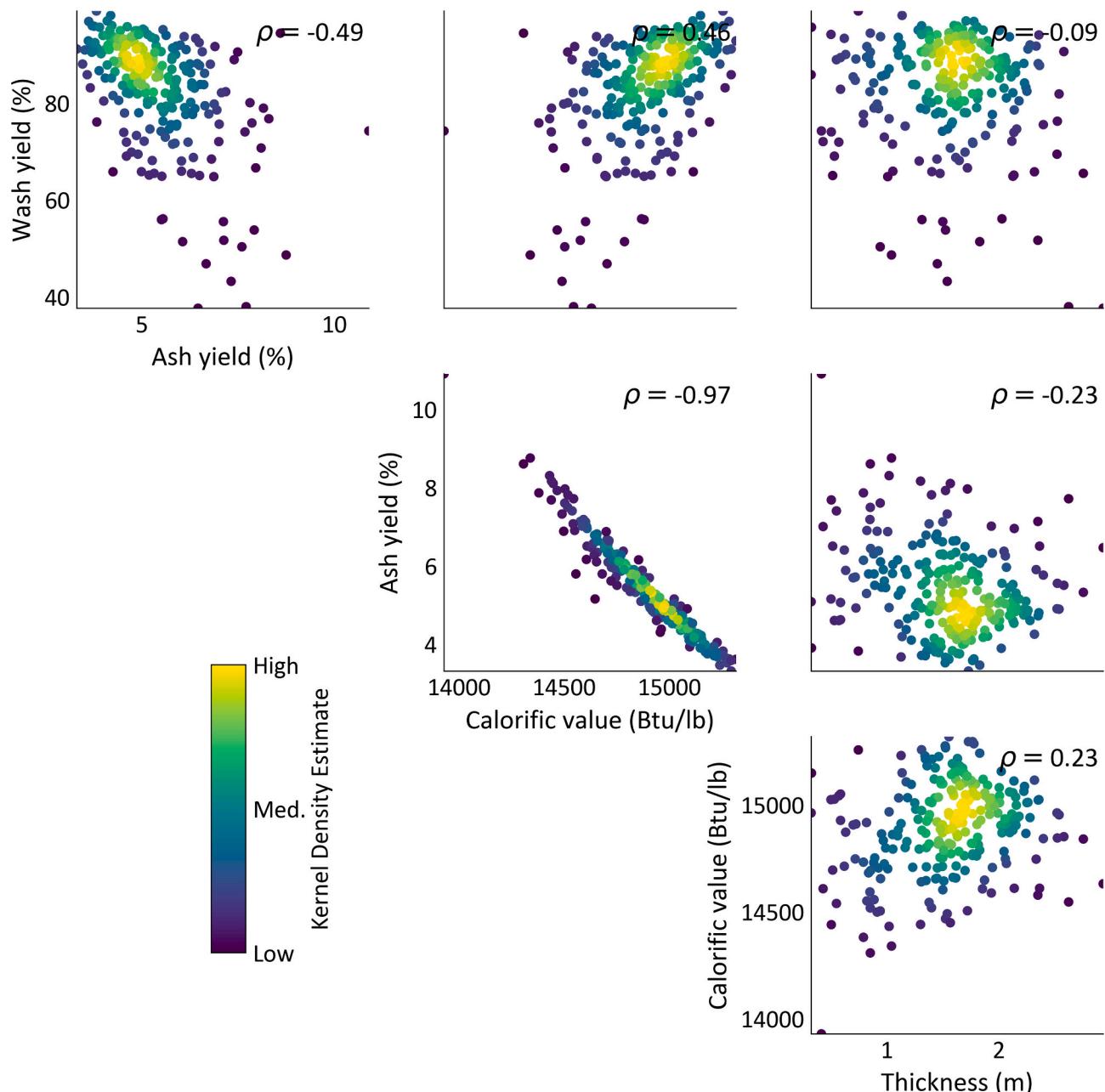


Fig. 3. Splitting the original data into training set (TS) and validation set (VS).

Table 1

The summary statistics of the original data, TS and VS (1 Btu/lb. = 0.556 kcal/kg).

	Original Data				Training Set				Validation Set			
	Wash Yield (%)	Ash Yield (%)	Calorific Value (Btu/lb)	Seam Thickness (m)	Wash Yield (%)	Ash Yield (%)	Calorific Value (Btu/lb)	Seam Thickness (m)	Wash Yield (%)	Ash Yield (%)	Calorific Value (Btu/lb)	Seam Thickness (m)
Count	324	324	324	324	243	243	243	243	81	81	81	81
Mean	83.08	5.37	14,896	1.56	83.01	5.39	14,891.2	1.55	83.29	5.31	14,910.5	1.59
Std.	11.87	1.18	218.6	0.48	11.13	1.14	215.05	0.49	13.84	1.28	228.32	0.45
Dev.												
Min	26.99	3.31	13,929	0.33	37.63	3.31	13,929.0	0.33	26.99	3.34	14,135.0	0.37
Q1	78.76	4.54	14,758.2	1.31	78.72	4.61	14,756.5	1.30	79.92	4.35	14,764.0	1.37
Median	86.48	5.15	14,921.0	1.58	85.83	5.15	14,915.0	1.58	87.40	5.10	14,949	1.58
Q3	91.35	6.06	15,054.2	1.85	90.97	6.03	15,041.0	1.83	92.98	6.10	15,075.0	1.86
Max	98.94	11.05	15,338	2.93	98.94	10.9	15,317.0	2.93	97.18	11.05	15,338.0	2.59

**Fig. 4.** Scatter plots of the training set (TS) variables (1 Btu/lb. = 0.556 kcal/kg).

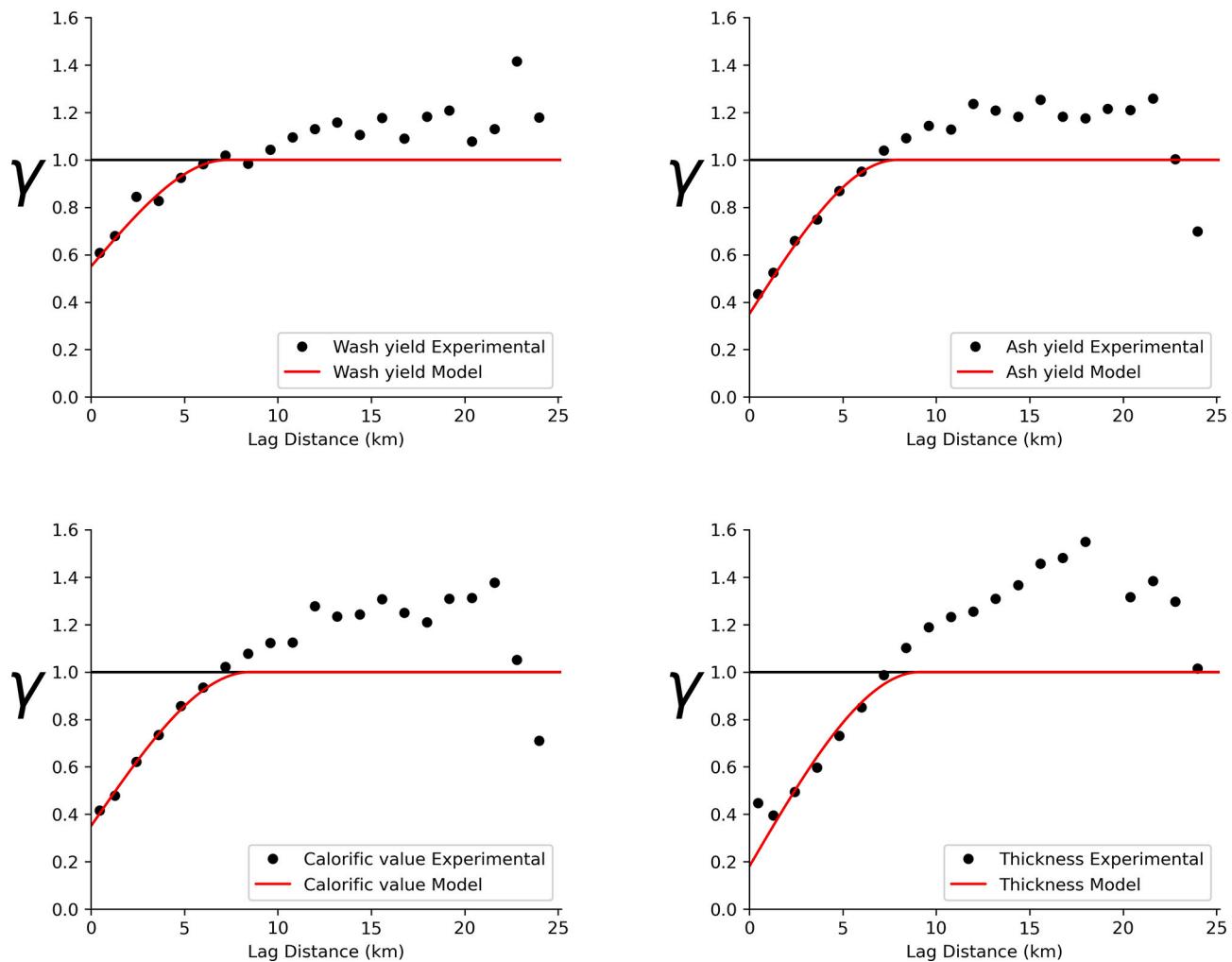


Fig. 5. Experimental omni-directional variograms of the training set (TS) variables and their fitted analytical models.

where \hat{z}_i is the i^{th} fitted value which is the point in the least squares regression line corresponding to z_i^* (Chatterjee and Hadi, 2006). $\hat{\beta}_0$ and $\hat{\beta}_1$ are intercept and slope estimated by the least squares method. The slope, $\hat{\beta}_1$ is estimated by:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (z_i - \bar{z}) \cdot (z_i^* - \bar{z}^*)}{\sum_{i=1}^n (z_i^* - \bar{z}^*)^2} \quad (34)$$

where z_i is the true value; \bar{z} is the mean of the true values; z_i^* is the estimated value and \bar{z}^* is the mean of the estimated values. The intercept, $\hat{\beta}_0$ is estimated by:

$$\hat{\beta}_0 = \bar{z} - \hat{\beta}_1 \cdot \bar{z}^* \quad (35)$$

The R^2 statistic is calculated by:

$$R^2 = 1 - \frac{\sum_{i=1}^n (z_i - \hat{z}_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad (36)$$

where the numerator denotes the sum of squared residuals (or errors), and denominator represents the sum of squared deviations in z from its mean \bar{z} .

Table 2 contains the values of the aforementioned criteria for all of the analytical models fitted.

It can be seen from **Table 2** that the slope of regression, $\hat{\beta}_1$ values are close to one for all variables, which confirms that there is no significant

Table 2

The results of the cross validation for the analytical models fitted to the experimental omni-directional variograms of the TS variables (1 Btu/lb. = 0.556 kcal/kg).

	Wash Yield (%)	Ash Yield (%)	Calorific Value (Btu/lb)	Seam Thickness (m)
R^2	0.222	0.384	0.419	0.563
$\hat{\beta}_1$	0.890	0.951	0.951	0.963
RMSE	9.842	0.899	164.1	0.325

conditional bias in the estimates. The R^2 values also indicate that the estimates generated using each variogram model appear to be close to their true values, except perhaps for wash yield where the variogram model uses a high nugget variance. The low RMSE values also indicate that the fitted variogram models appear to be plausible.

Each TS variable is then estimated through OK on 13,335 grid nodes ($0.3 \times 0.3 \text{ km}^2$). For each node estimation, up to 25 closest samples are retained, and no octants are used in the search neighborhoods. Finally, the maps of estimates and estimation variances for each variable are clipped to a given areal polygon (**Figs. 6 and 7**).

3.4. Modeling of the variables with MLAs

Considering the ESL approach (**Section 2.2.1**), the Easting (km) and Northing (km) coordinates of both TS and the estimation grid are first

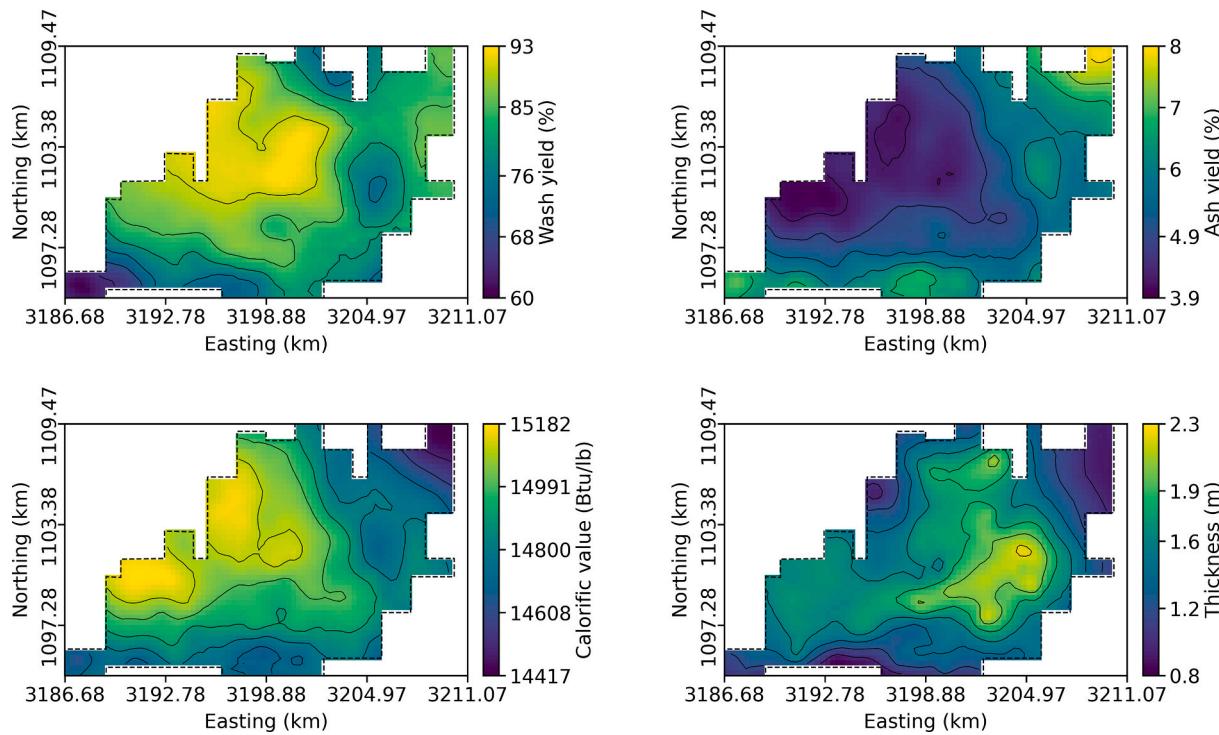


Fig. 6. Spatial maps of the ordinary kriging (OK) estimates of the training set (TS) variables (1 Btu/lb. = 0.556 kcal/kg).

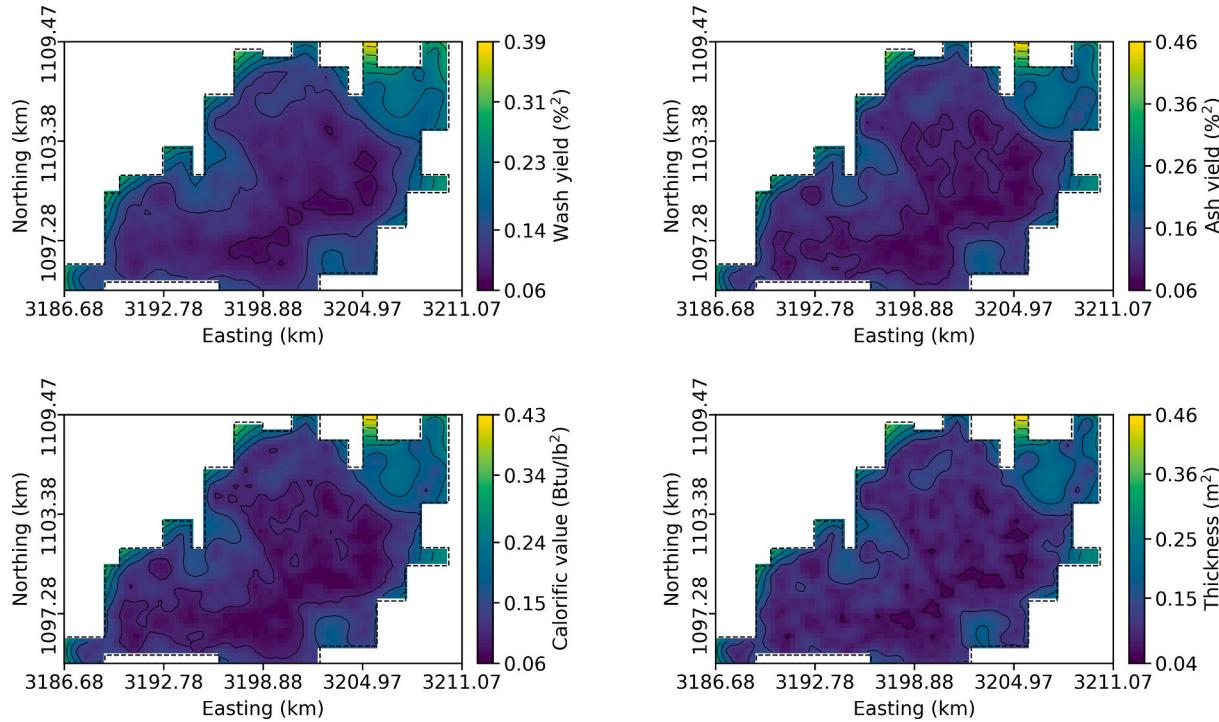


Fig. 7. Spatial maps of the local estimation variances of the estimated training set (TS) variables (1 Btu/lb. = 0.556 kcal/kg).

rotated clockwise, starting from $\alpha = 5^\circ$ to $\alpha = 90^\circ$ with an increment of 5° . Each of the resulting 18 TS with different coordinates is trained by individual SLMs (Section 2.2.1). The hyperparameters of MLAs in each SLM obtained from grid search with 5-fold cross-validation (Table 3). It should be noted here that the grid search cross-validation run time increases significantly for large datasets as this approach tries all combinations of parameters for a model. Therefore, the grid search algorithm

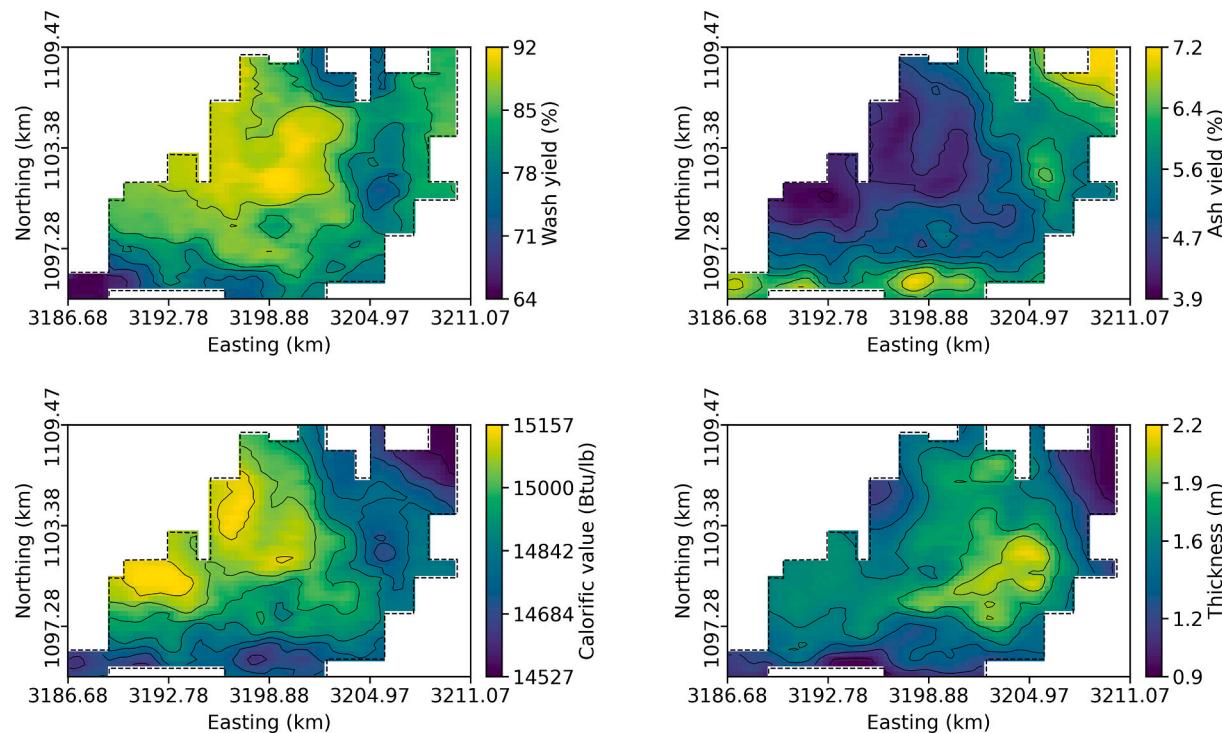
may not be feasible for SLM, which contains many MLAs for a large dataset. The estimates of each SLM on the corresponding rotated grid are then ensembled by mathematical averaging in order to yield the final spatial maps of the variables, which are shown in Fig. 8.

Considering the ERBFN approach, the initial learning rate is set to 0.1. TS is first split into 5-folds as training and test partitions, and a set of nodes for the hidden layer is then considered. For each number of nodes

Table 3

The optimized hyperparameters used for MLAs in each SLM (Pedregosa et al., 2011).

MLA	Hyperparameters	Wash Yield (%)	Ash Yield (%)	Calorific Value (Btu/lb)	Seam Thickness (m)
ANN	Batch size	10	10	10	10
	Epoch	100	100	100	100
	No. hidden layers	5	2	2	2
kNN	No. neurons	10	20	30	30
DecisionTree	No. neighbors	2	7	5	5
Bagging	Max depth	12	2	2	2
RandomForest	No. estimators	250	1000	500	750
ExtraTrees	No. estimators	250	500	750	500
AdaBoost	No. estimators	10	50	10	250
GradientBoosting	Learning rate	0.01	0.1	0.01	0.01
	Max depth	2	2	2	4
	No. estimators	50	50	400	250
SVR	C	1	1000	1	100
	Epsilon	0.01	1	0.01	1

**Fig. 8.** Spatial maps indicating the ensemble super learner (ESL) estimates of the training set (TS) variables (1 Btu/lb. = 0.556 kcal/kg).

in the set, individual networks are trained in the training partition, and predictions are made on the estimation grid. All of the estimates obtained from different number of nodes for each fold are then mathematically averaged to generate final ensembled estimates. The number of nodes in this study are set to 10% to 40% of the TS in each fold. Fig. 9 shows the spatial maps of the ERBFN estimates of the TS variables.

3.5. Modeling of the variables through hybrid spatial modeling approaches

The first hybrid modeling approach merges the model generated by ESL or ERBFN with the model generated by OK using the estimation variance-based weighting procedure (Section 2.3.1 and Eq. (27)). Considering the models generated by both ESL and ERBFN, the intercept, b_0 and slope, b_1 values (Eq. (29)) that yield the optimized value of the exponent b (Eq. (28)) are first determined by the sequential quadratic programming. Once the optimum b value is determined for both ESL and ERBFN models, the weights are assigned to ESL and OK

estimation models, and ERBFN and OK models according to Eq. (27). The final estimation models generated by combining ESL and OK, and ERBFN and OK are shown in Figs. 10 and 11, respectively.

The second hybrid modeling approach (Section 2.3.2) incorporates the model generated by ESL or ERBFN as exhaustive secondary information into the estimation of each variable of TS through OICCK (Section 2.1.2). The exhaustive secondary information (ESL or ERBFN estimates of each TS variable) as well as the samples of each TS variable are standardized. The variogram models (given in Fig. 5 and Eq. (31)) of the TS variables and the pairwise correlation coefficients between the samples of the TS variables and the collocated ESL or ERBFN models (Table 4) are the only required structural inputs to solve the OICCK system employing MM1 (Eq. (16)) for each grid node being estimated. The resulting estimation models generated by OICCK using ESL and ERBFN models as exhaustive secondary information are shown in Figs. 12 and 13, respectively.

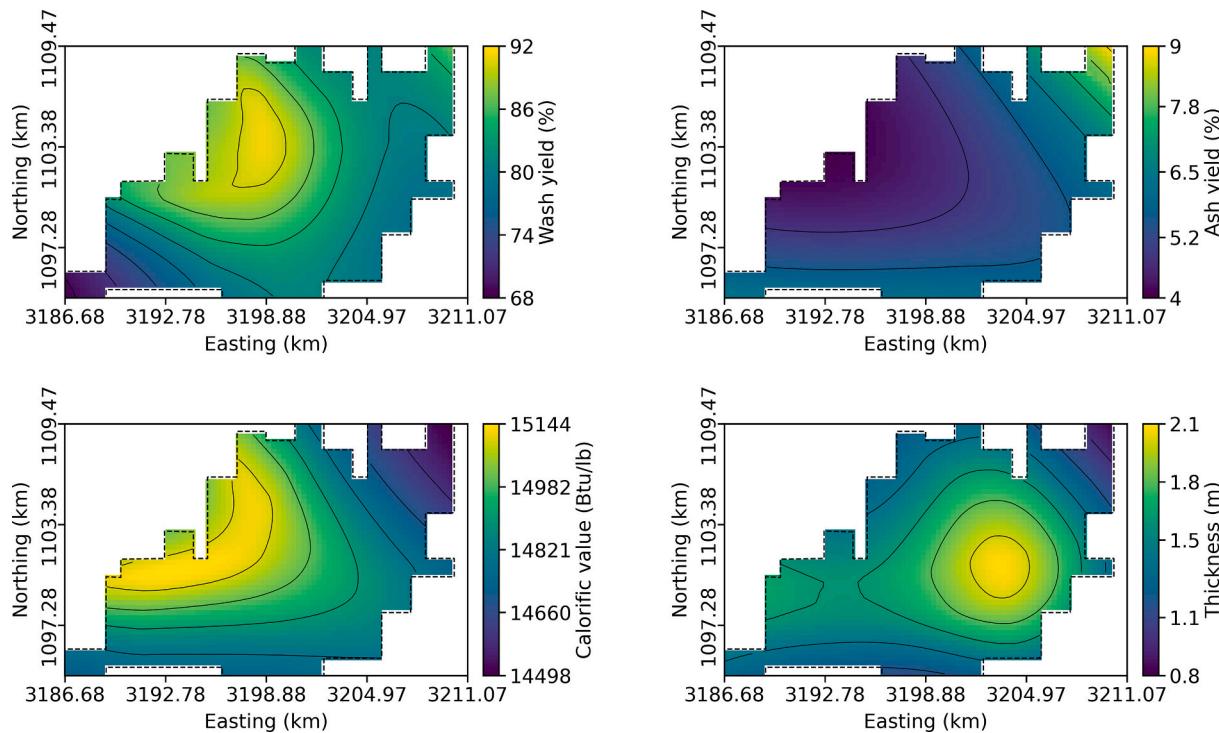


Fig. 9. Spatial maps indicating the elliptical radial basis neural network (ERBFN) estimates of the training set (TS) variables (1 Btu/lb. = 0.556 kcal/kg).

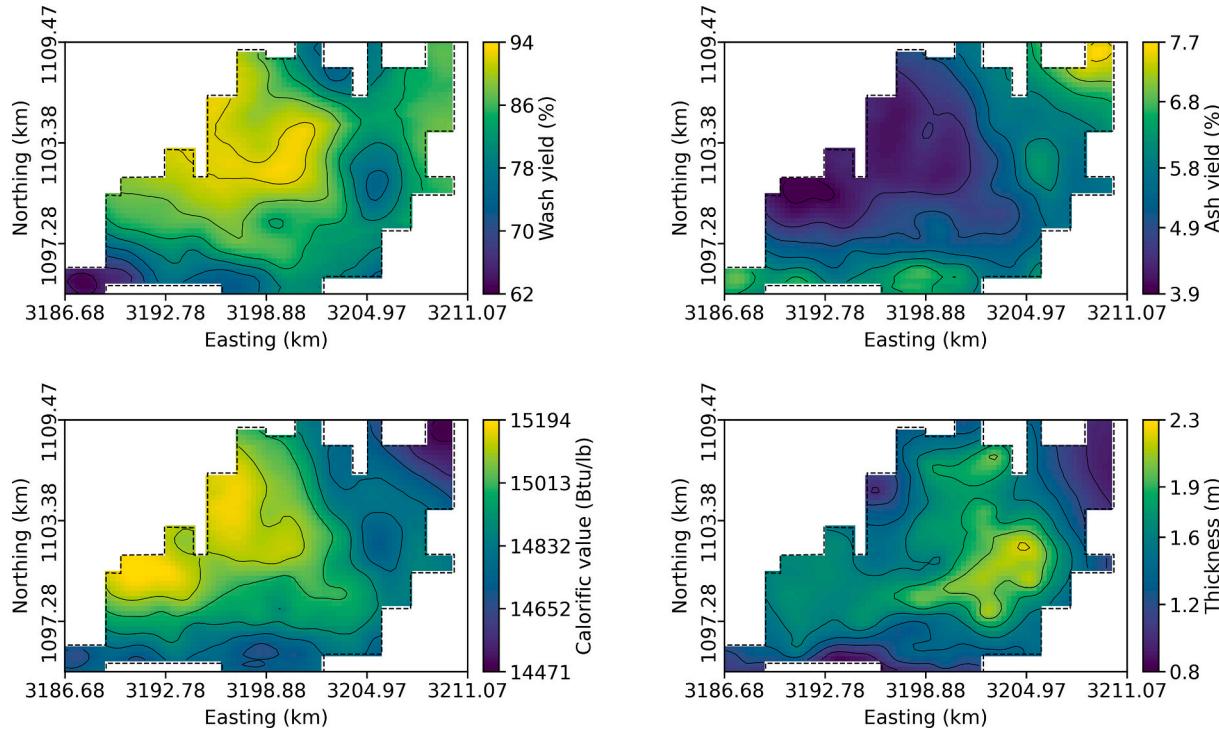


Fig. 10. Spatial maps of the estimates of the training set (TS) variables generated by merging ensemble super learner (ESL) and ordinary kriging (OK) (1 Btu/lb. = 0.556 kcal/kg).

4. Results and discussion

The performances of the estimation models generated by ESL, ERBFN, OK and the hybrid spatial modeling approaches are assessed in two ways: (i) visualization of each estimation model to detect artefacts, and (ii) comparison of the model predictions with the observed values of

the independent VS (Fig. 3). To evaluate the accuracy of each model, RMSE, which is a robust error measurement metric (Eq. (32)), is employed. Although the overall accuracy of the models is captured by RMSE, we also employ the slope of regression, $\hat{\beta}_1$ (Eq. (34)) and the correlation coefficient, r as alternative ways of assessing the quality of the estimation models.

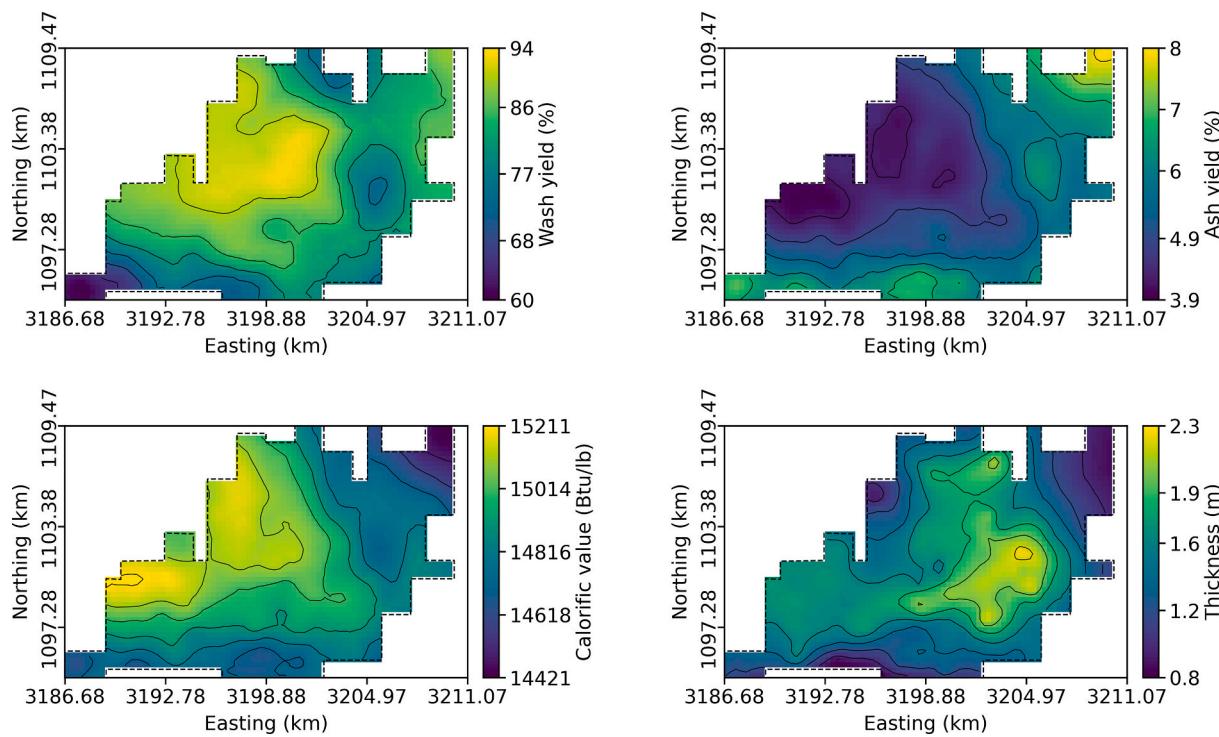


Fig. 11. Spatial maps of the estimates of the training set (TS) variables generated by merging elliptical radial basis neural network (ERBFN) and ordinary kriging (OK) (1 Btu/lb. = 0.556 kcal/kg).

Table 4

The correlation coefficients computed between each TS variable and the collocated values from the ESL and ERBFN models.

	Wash Yield (%)	Ash Yield (%)	Calorific Value (Btu/lb)	Seam Thickness (m)
ESL	0.69	0.81	0.82	0.86
ERBFN	0.52	0.67	0.71	0.78

Visually, the ESL estimation maps (Fig. 8) resemble those of OK (Fig. 6), but with some evident artefacts, which is mainly due to the fact that the spatial dependency between the samples is only explicitly accounted for in ESL. The ERBFN estimation maps (Fig. 9), on the other hand, show much less local variation than those of OK and ESL with no artefact is detected in the ERBFN maps. As for the hybrid modeling approaches, when ESL or ERBFN is merged with OK, the resulting estimation maps (Figs. 10 and 11) resemble rather closely those of OK, with no visible artefacts in the models. In the case where ESL estimates are used as secondary information in the OICCK system, the resulting maps (Fig. 12) are again comparable to those of OK, and the artefacts are much less obvious than those observed in the ESL estimation maps. Finally, the OICCK system using the ERBFN estimates as secondary information produces artefact-free estimation maps (Fig. 13) which are rather similar to the ones generated by OK.

Table 5 reports the RMSE values of all estimation models for VS, and Figs. 14 and 15 show the scatter plots of the estimates generated by each model versus the true values of VS. It can be seen from Table 5 that ESL and ERBFN produce estimation models with systematically higher RMSE values than those of OK for all variables, except for the ash model generated by ESL which appears to achieve almost the same accuracy as OK. The correlation coefficients between the OK estimates and the true values of VS also appear to be larger than those of ESL and ERBFN for all variables (Fig. 14). Considering the conditional bias in the models, ERBFN estimation models yield better slope of regression values than those generated by ESL and OK, but this is mainly because the ERBFN

produces relatively smooth estimation models, with less noise.

The estimation models generated by merging ESL and OK algorithms appear to yield consistently lower RMSE values than those of OK for all variables except the seam thickness, while the estimation models generated by merging ERBFN and OK algorithms result in higher RMSE values than those of OK for yield and ash variables and lower RMSE values for calorific value and seam thickness. The ash and yield variables tend to be more spatially varying in comparison to other variables. As the spatial variation in these variables cannot be captured adequately by the ERBFN estimation models, the accuracy of the final models for these variables are lower than those of OK.

Considering the case where the ESL estimates are incorporated into the estimation of primary variables through OICCK, the resulting estimation model appears to have systematically lower RMSE values for all variables except for seam thickness which exhibits greater continuity than the other variables. In addition, due to the high correlation coefficient between the values of the seam thickness in TS and the collocated values of the seam thickness from the ESL estimation model (Table 4), the resulting OICCK model for seam thickness tends to resemble the estimation model generated by ESL, which already has a high RMSE value. In the case where the ERBFN estimates are used as secondary information in the OICCK system, seam thickness gets better estimated than other variables. This is again because ERBFN estimates are smooth, and they perform better as secondary information for the estimation of seam thickness. Also, unlike the ESL case, the correlation coefficient between the seam thickness values from TS and the collocated seam thickness values from the ERBFN model is lower. Overall, the correlation coefficients between the estimation model values and the values of VS are higher for all hybrid models than individual models (Fig. 15).

According to the RMSE results (Table 5), there is no approach that consistently performed best for all of the coal variables. The case where the ESL estimation models are used as secondary information in OICCK appears to outperform OK and other hybrid spatial modeling approaches for all variables except seam thickness. However, the final estimation models may still exhibit some minor artefacts. On the other hand, the case where the ERBFN estimation models are used as secondary

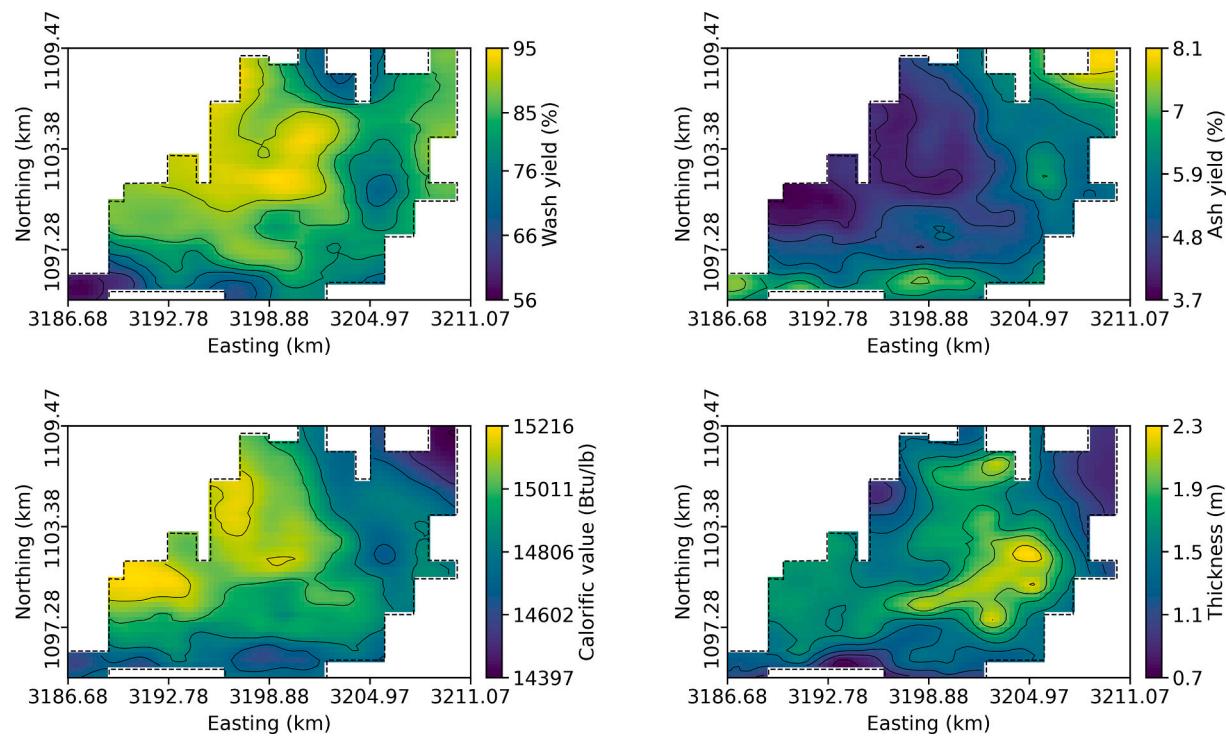


Fig. 12. Spatial maps of the estimates of the training set (TS) variables generated by ordinary intrinsic collocated cokriging (OICCK) where ensemble super learner (ESL) estimates are considered as exhaustive secondary information (1 Btu/lb. = 0.556 kcal/kg).

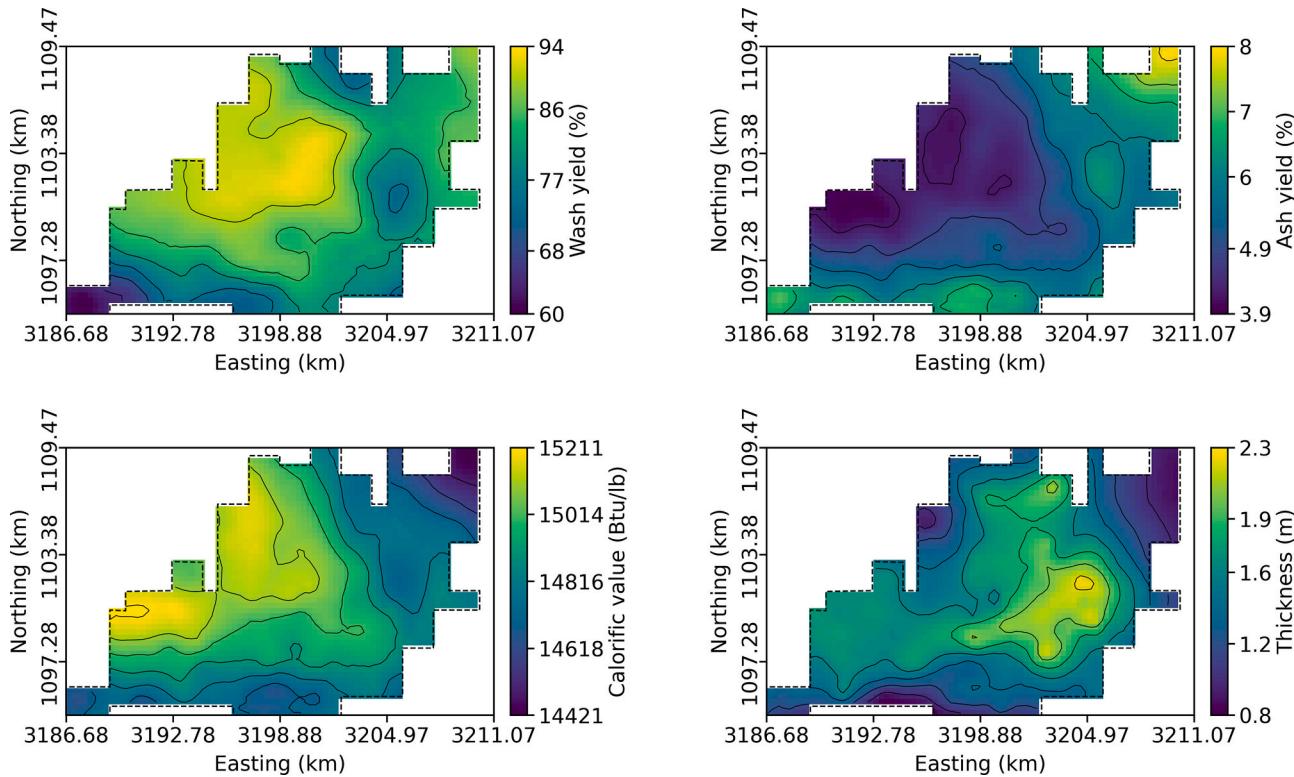


Fig. 13. Spatial maps of the estimates of the training set (TS) variables generated by ordinary intrinsic collocated cokriging (OICCK) where elliptical radial basis neural network (ERBFN) estimates are considered as exhaustive secondary information (1 Btu/lb. = 0.556 kcal/kg).

information in OICCK appears to generate estimation models with no visible artefacts.

We also checked the existence of the residual spatial autocorrelation in each estimation model by computing the experimental variograms of

the residuals calculated using the model and corresponding VS values. All experimental variograms of the residuals indicate no autocorrelation, which confirm that the spatial dependency between the samples is satisfactorily accounted for in all of the estimation models.

Table 5

The RMSE values of the estimation models computed with independent VS.

	Wash Yield (%)	Ash Yield (%)	Calorific Value (Btu/lb)	Seam Thickness (m)
OK	12.25	0.859	149.88	0.277
ESL	12.37	0.856	152.92	0.288
ERBFN	13.00	1.015	174.59	0.281
OK + ESL	12.20	0.854	148.85	0.278
OK + ERBFN	12.31	0.862	149.35	0.276
OICCK+ESL	12.04	0.799	142.46	0.290
OICCK+ERBFN	12.21	0.848	146.80	0.282

5. Conclusion

In this study, we evaluated the performance of estimation models generated by ESL, ERBFN, OK and hybrid spatial modeling approaches. The results suggest that machine learning, as a spatial estimation tool, suffers from two issues irrespective of the specified MLA used: (i) the spatial auto-correlation between the samples cannot easily be accounted for, and (ii) the estimates generated by MLAs do not reproduce the conditioning data at their locations. The former concern leads to estimation models with reduced accuracy and visible artefacts, while the latter concern generates model values not identical to sample values, which is important especially in geological modeling in three-dimensional space.

We have shown through the case study that when ESL or ERBFN is implemented alone, the resulting estimation models appear to be less

accurate than OK, and artefacts or extreme smoothness are evident in the estimation models. To address the above-mentioned shortcomings of MLA-based estimation workflows, we used two hybrid spatial modeling approaches: (i) the estimation variance-based weighting procedure ([Section 2.3.1](#)), and (ii) incorporation of MLAs through OICCK ([Section 2.3.2](#)). We demonstrated that the hybrid spatial modeling approaches can generally outperform OK with more accurate estimates and lower RMSE values, and that the data reproduction in the estimations and the spatial dependency among the samples can be considered through the combination of geostatistics and MLA.

We conclude that according to the RMSE results, the best performing method, which yielded the lowest RMSE value, appears to be the second hybrid approach where the ESL estimates were used as exhaustive secondary information in the OICCK system, but some minor artefacts were still evident in the final estimation models. The second best performing case would be the one where ERBFN estimates were used as exhaustive secondary information in the OICCK system, and the final estimation models do not exhibit any visible artefacts.

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.

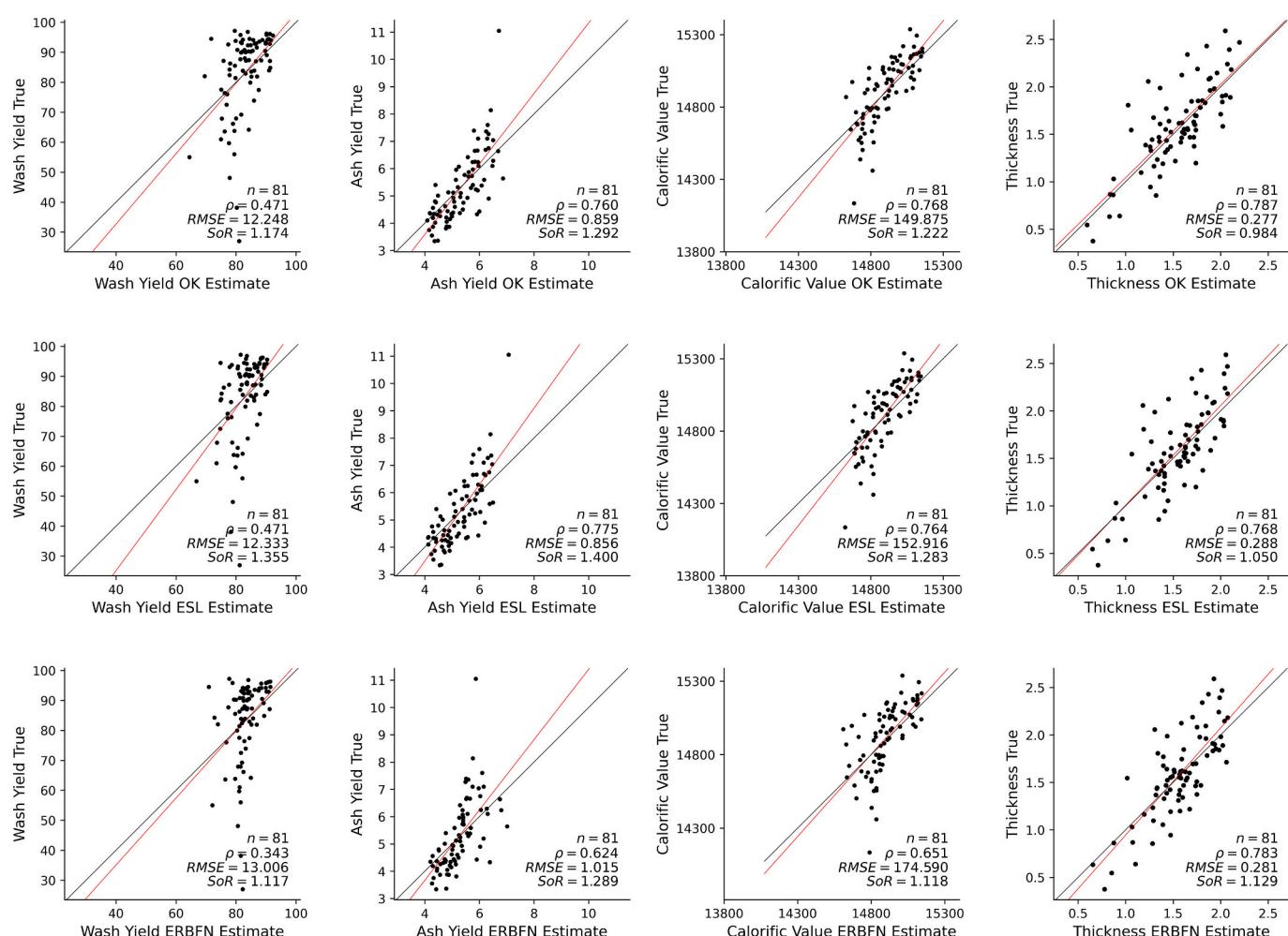


Fig. 14. Cross plots of the true values of the variables from validation sets (VS) versus the collocated estimated values generated by ordinary kriging (OK), ensemble super learner (ESL), and elliptical radial basis neural network (ERBFN).

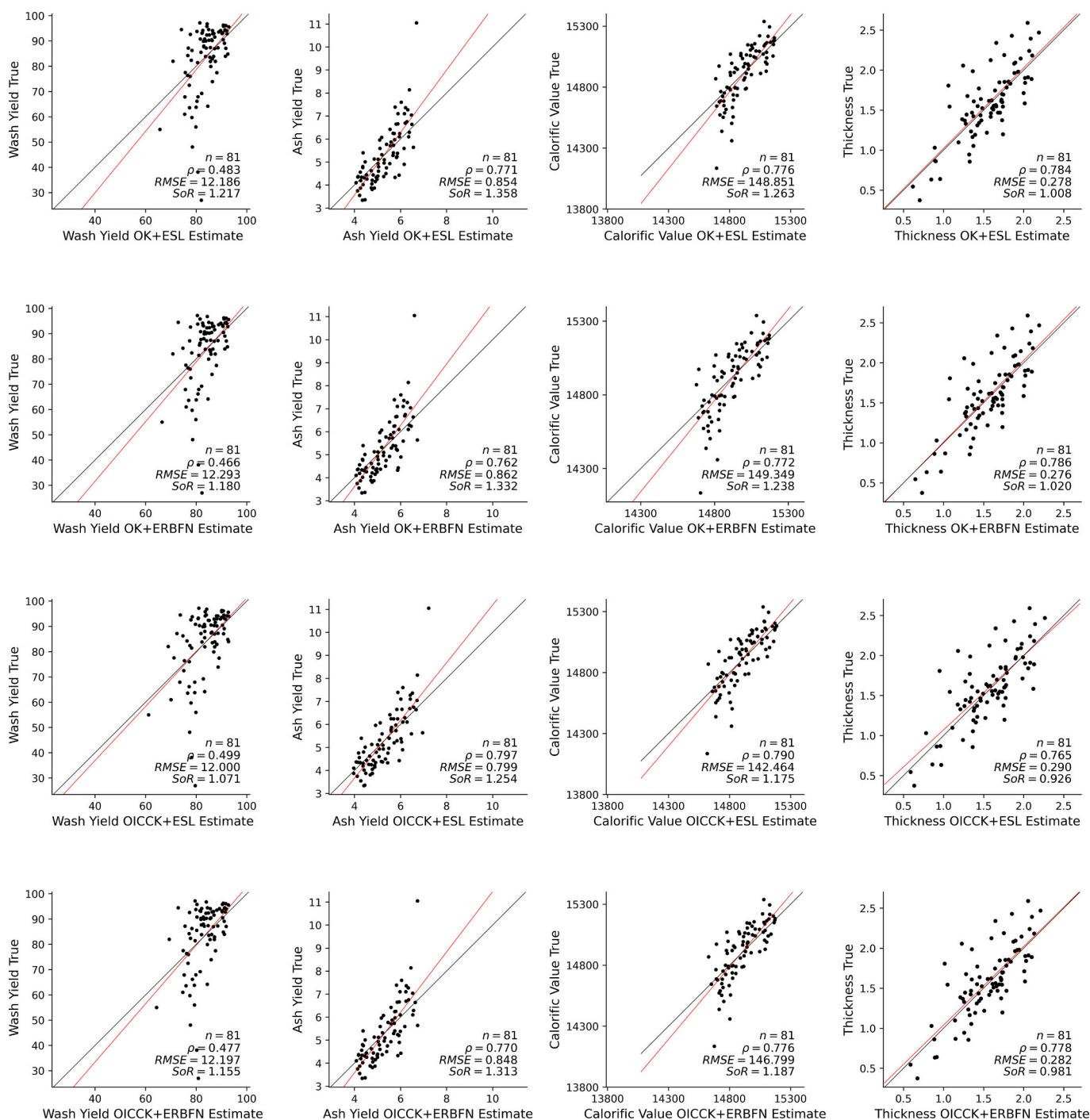


Fig. 15. Cross plots of the true values of the variables from validation sets (VS) sets versus the collocated estimated values generated by hybrid spatial modeling approaches.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgments

The study mine's management is gratefully acknowledged for granting permission to use their data. The authors who are affiliated with Centre for Computational Geostatistics (CCG) thank the industrial sponsors of CCG for providing the resources to prepare this paper. We thank Ashton Wiens (USGS) for reviewing earlier version of this paper.

We also thank Editor and the reviewers of the journal for their positive feedback and insightful comments that improved the paper. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Almeida, A.S., Journel, A.G., 1994. Joint simulation of multiple variables with a Markov-type coregionalization model. *Math. Geol.* 26, 565–588.
- Babak, O., Deutsch, C., 2009a. An intrinsic model of coregionalization that solves variance inflation in collocated cokriging. *Comput. Geosci.* 35, 603–614.
- Babak, O., Deutsch, C.V., 2009b. Collocated cokriging based on merged secondary attributes. *Math. Geosci.* 41, 921–926.

- Babak, O., Deutsch, C.V., 2009c. Improved spatial modeling by merging multiple secondary data for intrinsic collocated cokriging. *J. Pet. Sci. Eng.* 69, 93–99.
- Beheim, L., Zitouni, A., Belloir, F., de la Housse, C.D.M., 2004. New RBF neural network classifier with optimized hidden neurons number. *WSEAS Trans. Syst.* 2, 467–472.
- Behrens, T., Schmidt, K., Rossel, R.A.V., Gries, P., Scholten, T., MacMillan, R.A., 2018. Spatial modelling with Euclidean distance fields and machine learning. *Eur. J. Soil Sci.* 69, 757–770.
- Breiman, L., 1996. Stacked regressions. *Mach. Learn.* 24, 49–64.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
- Chatterjee, S., Hadi, A.S., 2006. Regression Analysis by Example. John Wiley & Sons.
- Chatterjee, S., Mastalerz, M., Drobniak, A., Karacan, C.Ö., 2022. Machine learning and data augmentation approach for identification of rare earth element potential in Indiana Coals, USA. *Int. J. Coal Geol.* 259, 104054 (ISSN 0166-5162).
- Chiles, J.P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. New York, second edn, John Wiley Sons.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27.
- da Silva, C.Z., Nisenzon, J., Boisvert, J., 2022. Grade control with ensembled machine learning: a comparative case study at the Carmen de Andacollo copper mine. *Nat. Resour. Res.* 31, 785–800.
- Deutsch, C.V., 2003. Geostatistics. In: Academic Press Encyclopedia of Physical Science and Technology, Third edition, pp. 697–707.
- Deutsch, C.V., Journel, A.G., 1998. GSLIB Geostatistical Software Library and User's Guide, second ed. Oxford University Press, New York.
- Erdogan Erten, G., Yavuz, M., Deutsch, C.V., 2021. Grade estimation by a machine learning model using coordinate rotations. *Appl. Earth Sci.* 130 (1), 57–66.
- Erdogan Erten, G., Yavuz, M., Deutsch, C.V., 2022. Combination of machine learning and kriging for spatial estimation of geological attributes. *Nat. Resour. Res.* 31 (1), 191–213.
- Ertuğç, G., Tercan, A.E., Hindistan, M.A., Ünver, B., Ünal, S., Atalay, F., Kllolu, S.Y., 2013. Geostatistical estimation of coal quality variables by using covariance matching constrained kriging. *Int. J. Coal Geol.* 112, 14–25.
- Fouedjio, F., 2020. Exact conditioning of regression Random forest for spatial prediction. *Artif. Intell. Geosci.* 1, 11–23.
- Fouedjio, F., 2021. Classification random forest with exact conditioning for spatial prediction of categorical variables. *Artif. Intell. Geosci.* 2, 82–95.
- Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm, vol. 96. Citeseer, pp. 148–156.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
- Gebovy, N.J., Olea, R.A., Engle, M.A., Martín-Fernández, J.A., 2013. Using simulated maps to interpret the geochemistry, formation and quality of the Blue Gem coal bed, Kentucky, USA. *Int. J. Coal Geol.* 112, 26–35.
- Georganos, S., Grippa, T., Gadiaga, A.N., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogriou, S., 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocart. Int.* 36, 121–136 (ISSN 1010-6049).
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach. Learn.* 63, 3–42.
- Ghosh, J., Nag, A., 2001. An overview of radial basis function networks. In: Radial Basis Function Networks, 2, pp. 1–36.
- Gilliland, E.S., Ripepi, N., Conrad, M., Miller, M.J., Karmis, M., 2013. Selection of monitoring techniques for a carbon storage and enhanced coalbed methane recovery pilot test in the Central Appalachian Basin. *Int. J. Coal Geol.* 118, 105–112.
- Goovaerts, P., 1997. Geostatistics for Natural Resource Evaluation. Oxford University Press.
- Grimm, R.P., Eriksson, K.A., Ripepi, N., Eble, C., Greb, S.F., 2012. Seal evaluation and confinement screening criteria for beneficial carbon dioxide storage with enhanced coal bed methane recovery in the Pocahontas Basin, Virginia. *Int. J. Coal Geol.* 90, 110–125.
- Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. Springer.
- Haykin, S., 2009. Neural Networks and Learning Machines. Pearson Education India.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Grler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6, e5518.
- Henika, W.S., Dishner, O.G., 1994. Internal Structure of the Coal-Bearing Portion of the Cumberland Overthrust Block in Southwestern Virginia and Adjoining Areas, Geology and Mineral Resources of the Southwest Virginia Coalfield. Commonwealth of Virginia Department of Conservation and Economic Development, Division of Mineral Resources, Charlottesville, Virginia, pp. 101–120.
- Heriawan, M.N., Koike, K., 2008. Uncertainty assessment of coal tonnage by spatial modeling of seam distribution and coal quality. *Int. J. Coal Geol.* 76 (3), 217–226.
- Heriawan, M.N., Koike, K., 2008b. Identifying spatial heterogeneity of coal resource quality in a multilayer coal deposit by multivariate geostatistics. *Int. J. Coal Geol.* 73 (3), 307–330.
- Ibrahim, A.F., 2022. Application of various machine learning techniques in predicting coal wettability for CO₂ sequestration purpose. *Int. J. Coal Geol.* 252, 103951 (ISSN 0166-5162).
- Ibrikci, T., Brandt, M.E., Wang, G., Acikkar, M., 2002. Mahalanobis distance with radial basis function network on protein secondary structures. In: Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society [Engineering in Medicine and Biology], vol. 3. IEEE, pp. 2184–2185.
- Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, vol. 112. Springer.
- Jeuken, R., Xu, C., Dowd, P., 2020. Improving coal quality estimations with geostatistics and geophysical logs. *Nat. Resour. Res.* 29 (4), 2529–2546.
- Journel, A.G., 1999. Markov models for cross-covariances. *Math. Geol.* 31, 955–964.
- Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic press.
- Kanevski, M., 2009. Machine Learning for Spatial Environmental Data: Theory, Applications, and Software. EPFL press.
- Kanevski, M., 2013. Advanced Mapping of Environmental Data. John Wiley & Sons.
- Karacan, C.Ö., Olea, R.A., 2018. Mapping of compositional properties of coal using isometric log-ratio transformation and sequential Gaussian simulation a comparative study for spatial ultimate analyses data. *J. Geochem. Explor.* 186, 36–49.
- Kingma, D.P., Ba, J., 2023. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Mak, M.-W., Li, C.K., 1999. Elliptical basis function networks and radial basis function networks for speaker verification: A comparative study. In: IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339), vol. 5. IEEE, pp. 3034–3039.
- Matheron, G., 2023. Les variables régionalisées et leur estimation [Regionalized variables and their estimation]. Editions Masson, Paris.
- Maxwell, K., Rajabi, M., Esterle, J., 2019. Automated classification of metamorphosed coal from geophysical log data using supervised machine learning techniques. *Int. J. Coal Geol.* 214, 103284.
- Maxwell, K., Rajabi, M., Esterle, J., 2021. Spatial interpolation of coal properties using geographic quantile regression forest. *Int. J. Coal Geol.* 248, 103869.
- Milici, R.C., Freeman, P.A., Bragg, L.J., 2000. A digital resource model of the Lower Pennsylvanian Pocahontas No. 3 coal bed, Pottsville Group, central Appalachian coal region. In: U.S. Geological Survey Professional Report 1625C. H1–H57.
- Nolde, J.E., 1994. Devonian to Pennsylvanian Stratigraphy and Coal Beds of the Appalachian Plateaus Province, Geology and Mineral Resources of the Southwest Virginia Coalfield, 131. Virginia Division of Mineral Resources Publication, pp. 1–85.
- Nolde, J.E., Spears, D., 1998. A preliminary assessment of in place coalbed methane resources in the Virginia portion of the central Appalachian Basin. *Int. J. Coal Geol.* 38 (1–2), 115–136 (ISSN 0166-5162).
- Olea, R.A., 1999. Geostatistics for Engineers and Earth Scientists. Springer Science & Business Media.
- Olea, R.A., Luppens, J.A., 2015. Mapping of coal quality using stochastic simulation and isometric logratio transformation with an application to a Texas lignite. *Int. J. Coal Geol.* 152, 80–93.
- Pardo-Igúzquiza, E., Dowd, P.A., Baltuille, J.M., Chica-Olmo, M., 2013. Geostatistical modelling of a coal seam for resource risk assessment. *Int. J. Coal Geol.* 112, 134–140.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pyrz, M., Deutsch, C., 2014. Geostatistical Reservoir Modelling, second edn. Oxford University Press.
- Quinlan, J.R., 1986. Induction of decision trees. *Mach. Learn.* 1, 81–106.
- Rossi, M.E., Deutsch, C.V., 2013. Mineral Resource Estimation. Springer Science & Business Media.
- Samson, M., Deutsch, C.V., 2022. A hybrid estimation technique using elliptical radial basis neural networks and cokriging. *Math. Geosci.* 54 (3), 573–591.
- Talebi, H., Peeters, L.J.M., Otto, A., Tolosana-Delgado, R., 2022. A Truly Spatial Random Forests Algorithm for Geoscience Data Analysis and Modelling. *Math. Geosci.* 54, 1–22.
- Tercan, A.E., Ünver, B., Hindistan, M.A., Ertuğç, G., Atalay, F., Ünal, S., Kllolu, Y., 2013. Seam modeling and resource estimation in the coalfields of western Anatolia. *Int. J. Coal Geol.* 112, 94–106.
- Tiwary, A.K., Ghosh, S., Singh, R., Mukherjee, D.P., Shankar, B.U., Dash, P.S., 2020. Automated coal petrography using random forest. *Int. J. Coal Geol.* 232, 103629.
- Van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2023. Super learner, Statistical applications in genetics and molecular biology, 6, p. 1.
- Wackernagel, H., 2003. Multivariate Geostatistics. Springer.
- Webber, T., Costa, J.F.C.L., Salvadoretti, P., 2013. Using borehole geophysical data as soft information in indicator kriging for coal quality estimation. *Int. J. Coal Geol.* 112, 67–75.
- Yang, X.-S., 2017. Chapter 23 - Optimization. In: Yang, X.-S. (Ed.), Engineering Mathematics with Examples and Applications. Academic Press, pp. 267–283.