Data-driven Modeling of the Methane Adsorption Isotherm on Coal Using Supervised Learning Methods: A Comparative Study

Qihong Feng1,2, Jiaming Wang1,2, Jiyuan Zhang1,2,* and Xianmin Zhang1,2

1Key Laboratory of Unconventional Oil & Gas Development (China University of Petroleum (East China)), Ministry of Education, Qingdao, China
2School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, China

*Corresponding author email: zhjy221@126.com

Abstract. Methane adsorption isotherm on coals is key to the development of coalbed methane (CBM). Laboratory measurement of adsorption isotherm is time-consuming. This paper presents a comparative study on the accuracy and robustness of seven supervised learning (SL) methods in estimating the methane adsorption isotherm based on coal properties. The SL methods used include the Gaussian process regression (GPR), kernel ridge regression (KRR), classifier and regression tree (CART) and four ensemble decision tree methods (random forests (RF), Adaboost, gradient boosting decision tree (GBDT) and extreme boosting (XGBoost)). The results show that all these SL methods are capable of correlating methane adsorption amounts with the feature variables with reasonable accuracies in the training stage. However, the KRR, GBDT and XGBoost are demonstrated to outperform other SL techniques in terms of the robustness and generalization capability, which therefore are recommended for fast estimation of the methane adsorption isotherms on coals.

Keywords: Methane adsorption isotherm; Coal; Supervised learning; Gaussian process regression; Kernel ridge regression; Classifier and regression tree; Ensemble decision tree methods.

1. Introduction

The coalbed methane (CBM), which is also called coalmine methane (CMM) or coal seam methane (CSM), has become an important source of that adds to the geo-energy supply in China, Australia and India. The development and utilization of CBM is also beneficial to reducing the releases of methane into atmosphere as well as enhancing the mining safety by reducing the risks of gas outburst and explosion in underground coal mines.[1]

It is generally considered that the majority of CBM, with its primary component being methane, exists in coalbeds in a physically adsorbed state. The methane adsorption isotherm determines the level of CBM resources of a coal seam and affect significantly the production of CBM.[2-4]. Therefore, a sound understanding of the adsorption isotherm characteristics is key to the accurate prediction of CBM productions and the design of optimal development strategies. To date, the methane adsorption isotherm is commonly determined using laboratory measurement techniques such as [5]. These laboratory measurements are quite time-consuming because the adsorption equilibrium generally takes more than several hours for an individual pressure step. Besides, laboratory tests are associated with
varying sources of uncertainties that may result in discrepancies in the adsorption isotherm between different laboratories\cite{6,7}.

Compared with adsorption isotherm measurements, determinations of coal properties are faster, less expensive and have better inter-laboratory consistency. A large number of researches have confirmed that the adsorption capacity of methane in coal is controlled jointly by the properties of coal and the conditions of experiments (e.g., coal particle size and temperature). For example, it is well documented that methane adsorption capacity correlates negatively with the ash content because ash has no affinity to adsorbing methane\cite{5,8}. The maturity degree of coal, which can be represented with the fixed carbon content and/or vitrinite reflectance, is also demonstrated to significantly affect methane adsorption capacity\cite{9,10}. The methane adsorption capacity generally follows either a monotonic increasing or a U-shaped trend\cite{8} with increasing coal mature. It is also shown from most laboratory tests that the increase in temperature typically leads to decreasing methane adsorption capacity, with only several exception cases in\cite{11,12}. Having addressed these previous findings, it is reasonable to presume that the methane adsorption isotherm may be estimated based on the coal properties and experimental conditions.

To date, a variety of SL methods have been proposed for solving regression problems. Previous studies\cite{13-15} suggest that the performance of an SL method is problem-dependent and no rule-of-thumb exists regarding the choice of a “best” method for a specific regression problem. Therefore, it is should be of practical significance to compare the performance of various SL methods in the application of estimating the methane adsorption isotherm. This paper presents a comprehensive evaluation of different SL methods for estimating methane adsorption isotherms, which include the kernel ridge regression (KRR)\cite{16}, Gaussian process regression (GPR)\cite{17}, classifier and regression tree (CART)\cite{18} and four ensemble decision tree (EDT) methods (random forest (RF)\cite{19}, adaptive boosting decision tree (Adaboost)\cite{20}, gradient boosting decision tree (GBDT)\cite{21} and extreme gradient boosting (XGBoost)\cite{22}). The accuracy of robustness of these SL methods will be systematically investigated in order to select the one that produces the most reliable estimations. This paper is advancement over our previous study\cite{23} that involves only the GBDT method.

2. Materials and Methods

2.1. Experimental Data

The experiment data sets used for testing the performance of SL methods were collected from Zhang et al.\cite{23}. The database was developed based on a number of 165 coal samples that were collected from CBM wellbores in the Qinshui basin, China. Laboratory tests were conducted on these samples that include vitrinite reflectance determination, maceral group identification, proximate analysis and methane adsorption isotherm measurement. For each sample, the adsorption isotherm is constituted of adsorption amounts corresponding to eight equilibrium pressures. As such, the database composes of a total of 165×8=1320 data points. The experimental data for these coal samples are summarized in Table 1.

| Property                  | Maximum | Minimum | Average |
|---------------------------|---------|---------|---------|
| Inherent moisture (a.r.), %| 2.20    | 0.34    | 1.10    |
| Ash (a.d.), %             | 49.59   | 4.85    | 18.70   |
| Vitrinite (m.m.f), %      | 97.80   | 47.50   | 80.77   |
| Fixed carbon (d.a.f.), %  | 93.08   | 78.15   | 87.74   |
| Equilibrium moisture, %   | 33.90   | 6.00    | 14.22   |
| Vitrinite reflectance, %  | 3.18    | 1.67    | 2.39    |
| Equilibrium pressure MPa | 8       | 1       | 24.25   |
| Temperature, °C           | 45.0    | 24.0    | 33.71   |
| Adsorption amount, m³/t   | 2.90    | 1.52    | 2.03    |

Note: a.d. – air dry; a.r. - as received; m.m.f - mineral matter free; d.a.f – dry ash free.
2.2. **SL Methods**

2.2.1. **GPR.** The GPR is a Bayesian nonparametric approach that targets at mapping the distribution of a function through GP models. In a GPR process, it is assumed that the distribution over the mapping functions are jointly Gaussian, which can be defined by the prior mean and positive-definite kernel functions [24]. In this regard, the mapping (regression) problem of correlating the inputs with the outputs is transformed to an optimization problem of maximizing the log marginal likelihood [17]. The maximization of the log marginal likelihood is achieved by tuning its hyperparameters assisted by e.g., a gradient-based optimization algorithm.

2.2.2. **KRR.** The KRR combines the ridge regression with kernel trick to handle nonlinear function estimations [25]. The KRR is identical with the ordinary ridge regression for transformations of the regressors but is more computationally efficient due to the inclusion of an algebraic trick [26]. The basic philosophy of the KRR is transforming the nonlinear regression in the original space into a linear regression in the higher dimensional space through a nonlinear kernel function. For the implementation of KRR, a typical choice of kernel function is the radial basis function (RBF), which is used to construct the so-called Gram matrix in the transformed space [27].

2.2.3. **CART.** The basic philosophy behind a CART is to split a complex decision into several simpler decisions, which may simplify the interpretation of the results [28]. A CART consists of a root node, a set of internal nodes and a set of leaf nodes (Figure 1). The root node contains all the data sets that are subject to splitting into internal nodes according to binary decision rules; and the leaf nodes denote the final classes. The construction (or growth) of a CART is realized through iterative processes that involve i) optimizations of the splits that determine an internal node by maximizing the purity or similarity among the responses, and ii) assignments of the class to each leaf node. To reduce the risk of overfitting and improve model robustness, tree pruning is usually conducted following the growth process. The tree pruning process identifies branches that contribute least to the model accuracy rate and eliminates them [29].

2.2.4. **EDTs.** The goal of an EDT is to combine the predictions of several decision tree estimators in order to reduce the risk of overfitting and improve generalization capability over a single estimator. To date, there are two families of ensemble methods that have been frequently used, namely the bagging and boosting methods. The reasoning behind the bagging method is to construct a series of decision tree estimators independently and then to average the predictions given by all the estimators (Figure 1a). For the boosting method, a series of base decision tree estimators are built sequentially and each decision tree estimator is associated with a weight that is subject to tuning such that a powerful ensemble can be constructed (Figure 1b). In this study, four ensemble methods were evaluated, namely the RF, Adaboost, GBDT and XGBoost. The RF is a bagging-based while the remaining three are boosting-based ensemble methods.

![Figure 1. Illustration of the ensemble methods for decision tree: (a) bagging; (b) boosting.](image)

2.3. **Model Construction and Evaluation**

2.3.1. **Feature selection.** The input features are selected to be the ones that i) are easier, faster and less
expensive to determine than the measurement of adsorption isotherm and ii) may influence methane adsorption capacity based on empirical results from [6-12]. Following these two principles, the input features that are used to estimate the adsorption amount are assigned to be ash, equilibrium pressure, inherent moisture, fixed carbon, vitrinite reflectance, vitrinite, equilibrium moisture and temperature. It should be noted that previous studies [30,31] suggest to ascertain the dependencies of the response on input features in order to identify “redundant” features that should be eliminated from a regression model such that the model robustness can be improved. In this study, the effect of elimination such “redundant” features will be presented in the Discussion section.

2.3.2. Model selection using the K-fold cross validation. As stated previously in [32,33], the performance of a regression algorithm may be influenced by the settings of its hyperparameters. To construct and evaluate the estimation models based on different SL methods, the whole data set is randomly divided into two parts, namely the training and testing sets. The training set is used to train the SL model and to determine the optimal setting of the hyperparameters for each SL algorithm; the testing set is used to evaluate the model generalization capability. In this study, the 80% of the whole data set is randomly selected and assigned as the training set while the remaining 20% is used for testing. To select the optimal setting of each SL algorithm, the commonly used K-fold cross validation approach (Figure 2) was used, which is addressed as follows:

1) Specify the range of each of the hyperparameter that is to be tuned for each SL algorithm, which is shown in Table 2 in this study.
2) For each scenario of the hyperparameter settings, apply the K-fold cross validation approach on the training set and calculate the coefficient of determination ($R^2$) for each fold; the optimal hyperparameter setting for an SL algorithm is defined to be the one that produces the highest averaged $R^2$ on the K-split subsets.

![Figure 2. Model selection and evaluation using the K-fold cross validation (revised from [2]).](image)

**Table 2.** Ranges of the hyperparameters for each SL algorithm.

| Method     | Hyperparameter                  | Range          | Optimum |
|------------|---------------------------------|----------------|---------|
| GRP\(^1\)  | Length scale                    | $[10^{-3}, 10^{-1}, 1]$ | 1       |
| KRR        | Scale mixture                   | $[10^{-2}, 10^{-1}, 1]$ | 1       |
|            | Regularization parameter        | $[10^{-7}, 10^{-6}, 10^{-5}, 10^{-1}, 10^{-3}]$ | $10^{-4}$ |
|            | Kernel width                    | $[10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}]$ | $10^{-4}$ |
| CART       | Min. no. of samples to split a node | [5, 15, 25, 35, 45] | 15      |
| RF\(^2\)   | Min. no. of samples at a leaf node | [5, 15, 25, 35, 45] | 5       |
|            | Min. no. of samples to split a node | [2, 5, 15, 25, 35] | 5       |
|            | Min. no. of samples at a leaf node | [1, 5, 15, 25, 35] | 5       |
|            | Max. tree depth                 | [10, 30, 50, 70, 90] | 70      |
| Adaboost\(^2,3\) | Max. tree depth | [10, 30, 50, 70, 90] | 30      |
|            | Learning rate                   | [0.01, 0.05, 0.1, 0.3, 0.5] | 0.05   |
| GBDT/XGboost\(^2,3\) | Max. tree depth | [2, 3, 4, 5, 6] | 3       |
|            | Learning rate                   | [0.01, 0.05, 0.1, 0.3, 0.5] | 0.05   |

\(^1\) Kernel function: rational quadratic.
\(^2\) No. of estimators for the gradient based ensemble methods is 500.
\(^3\) Min. No. of samples to split a node: 10; Min. No. of samples to at a leaf node: 1.
2.3.3. Statistical matrices. The accuracy and performance of the estimation models were evaluated and compared using three statistical indicators, namely the mean relative error (MRE), root mean square error (RMSE) and coefficient of determination ($R^2$), which are written as

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - f_i}{y_i} \right|$$  \hspace{1cm} (1)

$$\text{RMSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - f_i)^2$$  \hspace{1cm} (2)

$$R^2 = \frac{\sum_{i=1}^{N} (y_i - f_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$  \hspace{1cm} (3)

where $y$ and $f$ are the measured and calculated adsorption amounts; $\bar{y}$ is the mean value of the measured adsorption amount; $N$ is the number of data points.

3. Results

3.1. Optimal Settings of the Hyperparameters

The optimal settings of the hyperparameters that were determined using the K-fold cross validation are summarized in Table 2 for each of the SL methods. It is noted that the maximum tree depths vary significantly for different ensemble methods although they are all based on decision tree estimators. The maximum depth for the RF (70) is the highest among the ensemble methods, which is followed by that of the Adaboost (30). The maximum tree depths for the GBDT and XGBoost are relatively small (3) compared with other ensemble methods. All the three boosting ensemble methods have a relatively small learning rates (0.05), which is consistent with Pedregosa et al. [31] that suggests the use of a learning rate of less than 0.1 in order to reduce the risk of overfitting.

3.2. Training and Testing Results

The regression and prediction results obtained from different SL methods are depicted in Figure 3. As shown, the data points in the training set are generally located on the 45-degree line for the GPR approach, indicating an extremely high training accuracy. However, a majority of the data points in the testing set deviate severely from the 45-degree lines. These results suggest the GPR approach is associated with the problem of overfitting and therefore may be not be suitable for accurate estimation of the adsorption amount. For the KRR approach, data points in both the training and testing sets are generally grouped around the 45-degree line (Figure 3b). Although the training data points show a more scattered pattern than that with the GPR approach, no significant deviations from the 45-degree line occur for the testing data points, indicating a relatively strong generalization capability of the KRR. Similar regression and prediction results are observed between the CART and RF methods (Figure 3c and 4d): most data points are tightly arranged on the line of 45-degree with several outliers occurring in both the training and testing sets. As shown in Figure 3e, f and g, for all the boosting ensemble methods (Adaboost, GBDT and XGboost), data points in the training set are generally located at the 45-degree line while testing sets are grouped closely around the 45-degree line.
3.3. Quantitative Evaluation

Figure 4 depicts the statistical matrices for different SL methods regarding the training set. As shown, the GPR produces the highest $R^2$ (~1.0) and lowest MRE ($2.5 \times 10^{-4}$%) and RMSE ($4.9 \times 10^{-4}$ m$^3$/t) among all the methods investigated for the training set. The Adaboost, GBDT and XGBoost have comparable matrices, with MRE in the range of 1.35-1.51 m$^3$/t, RMSE in the range of 0.24%-0.25%, and $R^2$ higher than 0.997. These matrices represent high training accuracies for these boosting ensemble methods. The bagging ensemble method RF performs worse than these boosting methods but better than the CART as suggested by the matrices. The KRR results in a lowest $R^2$ (0.962) and highest MRE (5.76%) and RMSE (0.95 m$^3$/t) among all the methods investigated.

Figure 5 depicts the statistical matrices regarding the testing set. As shown, the KRR approach produces the highest $R^2$ (0.951) and lowest MRE (6.06%) and RMSE (0.99 m$^3$/t) in the testing stage, although its regression accuracy ranks the lowest in the training stage. The statistical matrices for the GBDT and XGBoost are comparable with that for the KRR. The remaining three tree-based methods including the CART, RF and Adaboost result in noticeably lower $R^2$ and higher MAE and RMSE than...
the KRR/GBDT/XGBoost. The GPR produces the lowest $R^2$ (0.802) and highest MRE (11.42%) and RMSE (1.98 m³/t) among the SL methods investigated in this study; such noticeable errors, which are consistent with the significant deviations as shown in Figure 3a, suggests again that the GPR is associated with the issue of overfitting.

Figure 5. Statistical matrices for different SL methods regarding the testing set.

4. Discussion

4.1. Selection of Optimal Models

For assessing the quality of regression models, the robustness and generalization capability are generally the first notable criterion. A model that is capable of reproducing the training set but fails to predict the testing sets with reasonable accuracies (e.g., the GPR in this study) should not be considered for further applications. From this standpoint of view, the KRR approach is highly recommended for estimating methane adsorption isotherms considering its high accuracy in the testing stage (which indicates strong robustness and generalization capability).

For tree-based models, the ensemble methods outperform the CART for both the training and testing sets. This is because an assembly of trees are trained (grew) in the ensemble method, which significantly increases the versatility for tuning the model than using a single tree; meanwhile, the data splitting strategy used in the ensemble method (Figure 1) functions essentially as a constraint on tree growth, which is beneficial to reducing the risks of overfitting. The superiority of ensemble methods over the CART has been well confirmed in previous studies that involve distinctive regression problems [34,35]. Among the ensemble methods, GBDT and XGBoost have comparable accuracies, which is possibly due to the generally similar operation procedures of these two methods [22]. Statistical matrices indicate stronger robustness and generalization capability of the GBDT and XGBoost than the RF and Adaboost, which is in line with [36] but against [37]. Such contradiction highlights the necessity for thorough evaluations of different models for varying regression problems. Having addressed these findings, it is recommended that three SL methods, namely the KRR, GBDT and XGBoost, can be applied for estimating methane adsorption isotherms.

4.2. Effect of Input Features on the Prediction Accuracy

Figure 6 shows the features ranked based on the mutual information (MI) [30]. The MI is a quantitative measure of the nonlinear dependency of the response on features, and a higher MI indicates a stronger correlation between the response and the feature. As can be seen, equilibrium pressure has a highest MI, which is followed by fixed carbon, ash, temperature and vitrinite reflectance (Romax) in a descending order of MI. Inherent moisture, equilibrium moisture and vitrinite contents have an MI value of zero, which indicate insignificant effects on the adsorption amount. As such, inherent moisture, equilibrium moisture and vitrinite contents are identified to be the “redundant” features.
Figure 6. MI for the features evaluated.

Table 3 compares the regression results based on the features including and excluding the redundant features for the three recommended SL methods identified in the previous section. It is evident that all the three methods result in reduced accuracies when the “redundant” features are excluded, for both the training and testing sets. Beker et al. [38] argued that the key to attaining the high accuracy and strong generalization capacity for a supervised learning tool is the inclusion of features that have underlying links with the output. However, the derivation of the quantitative indicator of the dependency typically does not involve a “learning” process and therefore may not be capable of finding the underlying correlation between the response and features. In this regard, a feature that may potentially affect the response may be categorized to be a “redundant” one. As mentioned in the Introduction section, for the specific regression problem in this paper, all the features have been demonstrated previously [6-12] to exert potential influence on the adsorption amount to some extent, which therefore should be included in the regression model.

Table 3. Comparison of the statistical matrices with different input features.

| Input feature¹ | Method   | Training set | Testing set |
|---------------|----------|--------------|-------------|
|               |          | MRE (%)      | RMSE (m³/t) | R²  | MRE (%)      | RMSE (m³/t) | R²  |
| P, A, FC, Romax, T | KRR     | 6.626        | 1.172       | 0.943 | 7.505        | 1.294       | 0.916 |
| P, A, FC, Romax, T | GBDT    | 5.037        | 0.899       | 0.966 | 8.071        | 1.413       | 0.900 |
| P, A, FC, IM, V, Romax, T | XGBoost | 6.277        | 1.130       | 0.947 | 8.393        | 1.454       | 0.894 |
| P, A, FC, IM, V, Romax, T | GBDT    | 5.765        | 0.955       | 0.962 | 6.058        | 0.992       | 0.951 |
| P, A, FC, IM, V, Romax, T | XGBoost | 1.353        | 0.230       | 0.998 | 6.455        | 1.106       | 0.939 |

¹ P: pressure; A: ash; FC: fixed carbon; IM: inherent moisture; V: vitrinite; T: temperature; EM: equilibrium moisture.

5. Conclusions

In our research, the performance of seven most advanced SL methods are evaluated regarding the estimation of methane adsorption isotherm based on coal properties and experimental conditions. It is demonstrated that the KRR, GBDT and XGBoost have comparable accuracies and relatively strong robustness than other tree-based methods including the CART, RF and Adaboost. The GBP method is associated with the issue of overfitting and therefore may not be suitable for accurate estimation of methane adsorption isotherm. The results also show that the inclusion of “redundant” features identified with the MI in the regression models is favorable to improving the model accuracy and robustness.

Acknowledgements

This research was funded by China NSFC, grant numbers 51904319 and U1810105, and China Postdoctoral Science Foundation, grant number 2018M642727.
References

[1] Karacan, C.O., Ruiz F.A., Michael C., et al. 2011 Coal mine methane: A review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction. Int. J. of Coal G., 86(2-3):121-156.

[2] Moore, T. 2012 Coalbed methane: a review. Int. J. Coal Geol. 101, 36-81.

[3] Young, G.B.C.; Paul, G.W.; McElhiney, J.E.; McBane, R.A., 1992 A parametric analysis of Fruitland coalbed methane reservoir producibility. SPE Annual Technical Conference and Exhibition (Washington, D.C., United States) pp. 13.

[4] Chen, Z.; Liu, J.; Kabir, A.; Wang, J.; Pan, Z., 2013 Impact of various parameters on the production of coalbed methane. SPE J. 18 (5), 910-923.

[5] Busch, A.; Gensterblum, Y. 2011 CBM and CO2-ECBM related sorption processes in coal: a review. Int. J. Coal Geol. 87(2), 49-71.

[6] Gasparik, M.; Rexer, T. F.; Aplin, A.C.; Billemont, P.; De Weireld, G.; Gensterblum, Y.; Zhang, T. 2014 First international inter-laboratory comparison of high-pressure CH₄, CO₂ and C₂H₆ sorption isotherms on carbonaceous shales. Int. J. Coal Geol. 132, 131-146.

[7] Gensterblum, Y., Hemert, P.V., Billemont, P., Battistutta, E., Busch, A., Krooss, B. M. 2010 European inter-laboratory comparison of high pressure CO₂ sorption isotherms ii: natural coals. Int. J. Coal Geol. 84(2), 115-124.

[8] Feng, Q.; Zhang, J.; Zhang, X.; Shu, C.; Wen, S.; Wang, S.; Li, J., 2014 The use of alternating conditional expectation to predict methane sorption capacity on coal. Int. J. Coal Geol. 121, 137-147.

[9] Dutta, P.; Bhowmik, S.; Das, S., 2011 Methane and carbon dioxide sorption on a set of coals from India. Int. J. Coal Geol. 85 (3), 289-299.

[10] Weniger, P.; Kalkreuth, W.; Busch, A.; Krooss, B.M., 2010 High-pressure methane and carbon dioxide sorption on coal and shale samples from the Paraná Basin, Brazil. Int. J. Coal Geol. 84 (3), 190-205.

[11] Guan, C.; Liu, S.; Li, C.; Wang, Y.; Zhao, Y., 2018 The temperature effect on the methane and CO₂ adsorption capacities of Illinois coal. Fuel 211, 241-250.

[12] Crosdale, P.J.; Moore, T.A.; Mares, T.E., 2008 Influence of moisture content and temperature on methane adsorption isotherm analysis for coals from a low-rank, biogenically-sourced gas reservoir. Int. J. Coal Geol. 76 (1), 166-174.

[13] Elmaz, F.; Yücel, Ö.; Mutlu, A.Y., 2020 Predictive modeling of biomass gasification with machine learning-based regression methods. Energy 191, 116541.

[14] Cornejo-Bueno, L.; Casanova-Mateo, C.; Sanz-Justo, J.; Salcedo-Sanz, S., 2019 Machine learning regressors for solar radiation estimation from satellite data. Solar Energy 183, 768-775.

[15] Ao, Y.; Li, H.; Zhu, L.; Ali, S.; Yang, Z., 2019 The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. J. Petrol. Sci. Eng. 174, 776-789.

[16] Vovk V. Kernel ridge regression. 2013 In Empirical Inference; Schölkopf B., Luo Z., Vovk V. Eds.; (Springer, Berlin, Heidelberg) pp 105-116.

[17] Rasmussen, C.E.; Williams C.K.I. 2006 Gaussian processes for machine learning, MIT Press, (London, England) pp. 7-30.

[18] Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. 1984 Classification and regression trees. (Wadsworth, Belmont, CA)

[19] Breiman, L. Random forests. 2001 Mach. Learn. 45(1), 5-32.

[20] Freund, Y.; Schapire, R.E., 1997 A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. 55 (1), 119-139.

[21] Friedman, J.H. 2002 Stochastic gradient boosting. Comput. Stat. Data An. 38(4), 367-378.

[22] Chen, T.; Guestrin C. XGBoost: A Scalable Tree Boosting System. arXiv:1603.02754.

[23] Zhang, J.; Feng, Q.; Zhang, X.; Hu, Q.; Yang, J.; Wang, N. 2020 A Novel Data-Driven Method to Estimate Methane Adsorption Isotherm on Coals Using the Gradient Boosting Decision Tree: A Case Study in the Qinshui Basin, China. Energies 13, 5369.
[24] Ballabio, C.; Lugato, E.; Fernández-Ugalde, O.; Orgiazzi, A.; Jones, A.; Borrelli, P.; Montanarella, L.; Panagos, P. 2019 Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma* 355, 113912.

[25] Murphy, K.P. 2012 Machine learning: A probabilistic perspective. *The MIT Press* (London, England) pp. 488-495.

[26] Exterkate, P.; Groenen, P.J.F.; Heij, C.; van Dijk, D. 2016 Nonlinear forecasting with many predictors using kernel ridge regression. *Int. J. Forecasting* 32 (3), 736-753.

[27] Douak, F.; Melgani, F.; Benoudjit, N. 2013 Kernel ridge regression with active learning for wind speed prediction. *Appl. Energ.* 103, 328-340.

[28] Xu, M.; Watanachaturaporn, P.; Varshney, P.K.; Arora, M.K., 2005 Decision tree regression for soft classification of remote sensing data. *Remote Sens. of Environ.* 97 (3), 322-336.

[29] Rahmatian, M.; Chen, Y.C.; Palizban, A.; Moshref, A.; Dunford, W.G., 2017 Transient stability assessment via decision trees and multivariate adaptive regression splines. *Electr. Pow. Syst. Res.* 142, 320-328.

[30] Kraskov, A.; Stogbauer, H.; Grassberger, P. 2004 *Estimating mutual information*. *Phys. Rev. E* 69.

[31] 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.

[32] Zhang, L.; Mistry, K.; Lim, C.P.; Neoh, S.C., 2018 Feature selection using firefly optimization for classification and regression models. *Decis. Support Syst.* 106, 64-85.

[33] Kim, Y.; Kim, S.B., 2018 Collinear groupwise feature selection via discrete fusion group regression. *Pattern Recogn.* 83, 1-13.

[34] Ebrahimi, M.; Mohammadi-Dehcheshmeh, M.; Ebrahimie, E.; Petrovski, K.R., 2019 Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep learning and gradient-boosted trees outperform other models. *Comput. Biol. Med.* 114, 103456.

[35] Lan, T.; Hu, H.; Jiang, C.; Yang, G.; Zhao, Z., A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification. *Adv. Space Res.* (in press).

[36] Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y., 2018 Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agr. Forest Meteorol.* 263, 225-241.

[37] Zhang, C.; Liu, C.; Zhang, X.; Almpanidis, G., 2017 An up-to-date comparison of state-of-the-art classification algorithms. *Expert Syst. Appl.* 82, 128-150.

[38] Beker, W.; Gajewska, E.P.; Badowski, T.; Grzybowski, B.A. 2019 Prediction of major regio-, site-, and diastereoisomers in Diels–Alder reactions by using machine-learning: The importance of physically meaningful descriptors. *Angew. Chem. Int. Edit.* 58(14), 4515-4519.