Multi-step-ahead soil temperature forecasting at multiple-depth based on meteorological data: Integrating resampling algorithms and machine learning models

Khabat Khosravi (khabat.khosravi@gmail.com)
Ferdowsi University of Mashhad
Ali Golkarian
Ferdowsi University of Mashhad
Rahim Barzegar
McGill University

Research Article

Keywords: Soil temperature, Data mining, Forecasting, Resampling algorithm, ensemble algorithms

Posted Date: April 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1526396/v1

Abstract

Direct soil temperature (ST) measurement is time-consuming and costly; thus, the use of a simple and cost-effective machine learning (ML) tool is helpful. In this study, ML approaches, including KStar, instance-based K-nearest learner (IBK) and locally weighted learner (LWL) coupled with resampling algorithms of bagging (BA) and dagging (DA) were developed and tested for multi-step ahead (3, 6 and 9 days ahead) ST forecasting. In addition, a linear regression model (LR) was used as a benchmark to compare the results. A dataset with daily ST time-series (as models' output) along with meteorological data (mean (T_{Mean}), minimum (T_{Min}) and maximum (T_{Max}) air temperature, evaporation (*Eva*), sunshine hours (*SSH*) and solar radiation (*SR*); as models' input) were collected at Isfahan synoptic station (Iran), in a farmland, during 13 years (1992–2005) at 5 and 50 cm soil depths. Six different input combination scenarios were proposed to the models based on Pearson's correlation coefficients between inputs and outputs. For the model building, we used 70% of the data and the remaining 30% was considered for model evaluation through different visual and quantitative metrics. Our findings showed that variable T_{Mean} is the most effective input variable for ST forecasting in most of the developed algorithms, while in some cases the combination of several variables including T_{Mean}, T_{Max} and as well as the integration of T_{Mean}, T_{Max}, T_{Min}, *Eva* and *SSH* proved to be the best input combinations. Among the evaluated models, KStar showed more compatibility with the BA algorithm, while, in most cases and depending on soil depth, IBK and LWL obtained more accurate results when they were hybridized with DA. For soil depth of 5 cm, BA-KStar has superior performance (i.e. Nash-Sutcliffe Efficiency (NSE) = 0.90, 0.87 and 0.85 for 3, 6 and 9 months ahead forecasting, respectively) while for soil depth of 50 cm, DA-KStar outperforms other algorithms (i.e. NSE = 0.88, 0.89 and 0.89 for 3, 6 and 9 months ahead forecasting, respectively). Also, results confirmed that all hybrid models had higher prediction capability than the LR model.

1. Introduction

Soil temperature (ST) plays a key role in different ecosystems by affecting processes such as the hydrological response of the soil, accumulation and degradation of organic matter, plant growth, nutrient mineralization, carbon emissions, proper time of sowing and even micro-organisms activity (Brar et al., 1992, Peng et al., 2009; Hu et al., 2016; Citakoglu, 2017). ST variation can alter soil characteristics and accordingly has considerable environmental outcomes with the change in the carbon balance (Qian et al., 2011). It can be an important parameter in ecological, climatological and hydrological modeling (Tabari et al., 2015). Information about ST is therefore important in decision-making processes. ST also varies with depth, but this variation is much smaller in the deeper layers than near the soil surface, and thus, accurate soil temperature assessments have to be done at different depths. For instance, ST in the topsoil (less than 5 cm) affects seed sowing, while information about ST at a depth of 10–15 cm is required for tree grafting. Previous studies have tried to find relationships between ST at various depths and the most important factors that affect this parameter (Citakoglu, 2017).

Unfortunately, spatially distributed data of ST does not exist in many regions of the world (Hu et al., 2016), as accurate measurements are expensive and time-consuming (Napagoda and Tilakaratne, 2012). Moreover, many environmental factors influence ST, including meteorological variables (e.g. solar radiation, air humidity, pressure and temperature, precipitation, sunny hours and wind speed), topographic conditions and soil factors such as soil water content, texture and surface cover (Paul et al., 2004; Samadianfard, 2018; Zeynoddin, 2019). Despite the technological advances in sensors and devices, direct measurements of ST are still expensive and require continuous measurements at different soil depths (Plauborg, 2002). To overcome the challenge of ST quantification for large areas, researchers have been concentrated on ST modelling and prediction using various techniques (Sándor and Fodor, 2012). Accurate ST modelling reduces time, costs and instrument maintenance (Maryanaji et al., 2017). Because ST has a strong correlation with meteorological variables, several models have been developed based on these relationships such as linear models (Kang et al., 2000; Bond-Lamberty et al., 2005), analytical models –based on conduction heat transfer (Ozgener et al., 2013; Cleall et al., 2015; Badache et al., 2016)– and numerical models –which consider complex heat and mass transport of soil (Liu et al., 2005; Belghit and Benyaich, 2014; Gao et al., 2016)–. These three types of models have their limitations: Linear equations do not have reasonable prediction power due to their simple and linear structure, while analytical and numerical models are difficult to use because of their complexity and high data demand (Xing et al., 2018).

Over the past recent decades, machine learning (ML) models, as computational artificial intelligence-based (AI) models, have captured researchers' attention in distinct disciplines, especially in geosciences. ML tools are able to process large datasets efficiently. Moreover, non-linear models (e.g. AI models) have a high capability to simulate complex processes due to their non-linear and complex structures (Khosravi et al., 2018, 2019).

Previous studies applied ML models successfully for ST modeling. For example, Mihalakakou (2002) and Ozturk et al., (2011) used an ANN model with geographical and meteorological variables and concluded that ANN has good accuracy for predicting monthly mean ST. Araghi et al., (2017) showed that the wavelet artificial neural network (WANN) was an accurate approach for forecasting 1–7 days ST ahead at depths of 5–30 cm. Citakoglu (2017) compared ANN, adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) models in estimating ST. They indicated that ANFIS outperforms both ANN and MLR models. Kisi *et al.*, (2017) implemented ANN, ANFIS and gene expression programming (GEP) for ST prediction at the depths of 10, 50 and 100 cm by using climatic data. They concluded that GP outperforms other algorithms while developing models without climatic data obtained better performance for ANN than the ANFIS and GP. Sanikhani et al., (2018) used non-tuned data intelligent models including extreme learning machine (ELM), ANN and M5 Model Tree (M5 Tree) to predict ST with monthly meteorological information as inputs and found that the ELM model is a suitable tool for ST estimation at multiple soil depths. Samadianfard et al., (2018) developed two data-intelligent models including WANN and GEP for the short-term estimation of ST at different depths and found that WANN had the best performance in all considered depths. Xing et al., (2018) applied a support vector machine (SVM) to predict daily ST in a different climate of the USA and revealed that SVM has a good

performance in predicting ST. Zeynoddin et al., (2019) found that the multilayer perceptron neural network (MLP) model resulted in a good performance for daily ST at two weather stations in northwestern Iran. Feng et al., (2019) assessed the abilities of ELM, generalized regression neural networks (GRNN), backpropagation neural networks (BPNN) and random forests (RF) models in modeling half-hourly ST and found that all models had an acceptable performance at different depths, while ELM had slightly better performance. Zeynodin *et al.*, (2020) predicted ST using stochastic linear modeling and Holt-Winters AExpo-SARIMA model at Bandar Abbas and Kerman synoptic stations, Iran, at depths of 5, 10, 20, 30, 50 and 100 cm. They stated that the proposed method had a reasonable prediction power.

Although the above mentioned traditional models (e.g. ANN, ANFIS, SVM and ELM) were successfully applied for ST modeling, those models may have drawbacks, such as i) low generalization performance of ANN (Melesse *et al.*, 2011), ii) accurate determining the weights in the membership function of ANFIS (Bui *et al.*, 2016), iii) requirement of large training datasets of ELM, and iv) high sensitivity of the SVM model to the hyper-parameter selection (Waseem Ahmad *et al.*, 2018). We hypothesize that newer AI-based models and data mining algorithms can address these drawbacks.

Data mining algorithms have proved their usefulness in flood susceptibility mapping through Logistic Model Trees (LMT), Kernel Logistic Regression (KLR), Radial Basis Function Classifier (RBFC), Multinomial Naïve Bayes (NBM) (Pham et al., 2020), landslide susceptibility assessment (Luo et al., 2019); and groundwater potential mapping using LMT, logistic regression (LR) and RF (Razavi-Termeh *et al.*,2019). Recently, these types of algorithms have also been applied for time-series data of prediction, including bed-load transport rate prediction through the M5P, random tree (RT), RF and REPT and four types of hybrid algorithms trained with a Bagging (BA) (BA-M5P, BA-RF, BA-RT and BA-REPT) (Khosravi et al., 2020a); also spring discharge prediction using M5P, RF and SVR (Granata et al., 2018). Other environmental processes such as suspended sediment load (Khosravi et al., 2018; Salih et al., 2019), and fluoride concentration prediction in groundwater (Khosravi et al., 2020b), erosion processes (Pourgahasmi *et al.*, 2017, 2020) and soil respiration (Ebrahimi et al., 2019) have also employed these types of data-mining algorithms.

Most of the above-mentioned studies demonstrated that newer standalone data mining algorithms (e.g. M5P, RT) are promising alternatives to traditional ML methods (e.g. ANN, ANFIS, SVR) while hybrid algorithms can further improve predictive performance over standalone algorithms due to increasing model's flexibility (Khosravi et al., 2018; Pham *et al.*, 2019).

In this study, we aim to propose six novel hybrid resampling algorithms of BA and dagging (DA) with IBK, KStar and LWL, namely: i) BA-IBK, ii) BA-KStar, iii) BA-LWL, iv) DA-IBK, v) DA-KStar, and vi) DA-LWL; for 3, 6 and 9 days ahead daily soil temperature forecasting based on nearby meteorological data at two different soil depth levels (5 and 50 cm). To the best author's knowledge and literature review, these hybrid algorithms are novel not only in soil science but also generally in geoscience, and this is the first attempt to forecast ST using the proposed models. We tested these approaches in a largely arid and semi-arid area devoted to agriculture where soil and water resources are scarce and accurate ST

estimations are necessary to promote long-term sustainable agriculture. Thus, through these algorithms, ST may easily and accurately be predicted worldwide. Moreover, ST prediction through these models could reduce the time and resources for measuring ST.

2. Materials And Methods

2.1. Study area

The Isfahan province is located in central Iran. It is characterized by arid and semi-arid conditions and comprises 10% of Iranian deserts. Western parts of the province include mountains with a mild climate. The air temperature ranges between 40.6°C and 10.6°C while the mean annual temperature is about 16.7°C and the mean annual rainfall is only 117 mm. The Isfahan province lies between 30° 42' to 34° 30' N and 49° 36' to 55° E and covers an area of about 107,027 km² (Fig. 1) of which about 5,674,000 ha are suitability for the cultivation of crops such as wheat, saffron, onion, potato and melon. Besides, 77,351 ha is devoted to orchards with apples, cherries, apricots and other fruit trees.

<Fig. 1. Study area and location of Isfahan synoptic station >

2.2. Methodology

2.2.1. Dataset collection and processing

The dataset was measured and collected by the Isfahan regional water authority, at the Isfahan synoptic station. Specific datasets were prepared for the daily soil temperature (ST) at depths of 5, 10, 20, 30, 50 and 100 cm. ST was measured using mercury-in-glass thermometers. The available meteorological data [i.e. mean (T_{Mean}), minimum (T_{Min}) and maximum (T_{Max}) air temperature, evaporation (*Eva*), sunshine hours (*SSH*) and solar radiation (*SR*)] were collected from June 1992 to December 2005. These meteorological data were considered as input to forecast the ST at two soil depths (5 and 50 cm) and were also successfully used in previous studies (e.g. Sattari et al., 2020) for ST modeling. These two depths are representative of the conditions near the soil surface (i.e. topsoil) and below the arable layer (i.e. deep soil), respectively. At first, the data were divided randomly into two main sub-groups in a 70:30 ratio. Kisi *et al.*, (2019) stated that by increasing the length of the training dataset from 50–75%, the results were enhanced and the model's prediction power increased. It should be noted that there is no 'rule of thumb' for partitioning train/test sets. However, previous studies (Palani *et al.*, 2008; Barzegar *et al.*, 2016) generally agreed that the testing subset must employ data never used in the training phase and that data should represent approximately from 10–40% of the size of the training set. In this study, we used the 70:30 ratio which is the most common ratio in both spatially and time-series AI modeling. Data from 1992 to 2001 was used as a training dataset for the model building while the data from 2002 to 2005 was applied as a testing dataset for model validation. Descriptive statistics are presented in Table 1.

<**Table 1.** Descriptive statistics of the applied dataset>

2.2.2. Input variable combination scenario

Alongside data quality, the determination of the best input combination has a significant effect on the modeling outcome. In this study, the model's potential input variables comprised of T_{Mean}, T_{Max}, T_{Min}, *Eva*, *SR*, and *SSH*. Although there are several methods e.g. the gamma test, entropy theory, Akaike information criterion and Bayesian information criterion for input variable selection (Jaafar and Han, 2012), we used the simple and straightforward Pearson correlation to select the most relevant input variables to build a good model with high performance. This method has been applied to develop prediction/forecasting models successfully (Barzegar et al., 2017; Khosravi et al., 2020). According to Pearson's correlation coefficient (*r*) between the input variables and the ST as an output (Table 2), six different scenarios as six different input combinations were considered for modeling process. Finally, and according to the correlation coefficient between forecasted and measured values, the optimum/best input combination was considered for model development and further analysis. This procedure was successfully applied in different prediction modeling.

We assumed that the first variables with the highest correlation coefficients were considered as the first input to the model (i.e. T_{Mean}). The assumption is to examine whether variables with the highest *r* can predict ST accurately alone. Then, other variables with the next highest *r* were added one by one until the variable with the lowest *r* (i.e. *SR*) was added, and the input combination number 6 was constructed:

1. 1. ST = $f(T_{Mean})$
2. 2. ST = $f(T_{Mean}, T_{Max})$
3. 3. ST = $f(T_{Mean}, T_{Max}, T_{Min})$
4. 4. ST = $f(T_{Mean}, T_{Max}, T_{Min}, Eva)$
5. 5. ST = $f(T_{Mean}, T_{Max}, T_{Min}, Eva, SSH)$
6. 6. ST = $f(T_{Mean}, T_{Max}, T_{Min}, Eva, SSH, SR)$

<**Tabel 2.** Correlation coefficients (*r*) between input variables and output for the forecast of the STat different soil depths.>

2.2.3. Machine learning model

2.2.3.1. Instance-Based Learning (IBK)

The instance-based learning (IBK) is a lazy algorithm based on the K-nearest neighbor classifier using the Euclidean metric as a function for measuring the distance between the instances (Smusz et al., 2013). Over the years, several applications of the IBK algorithm can be found in literature, for example, classification of bioactive compounds (Smusz et al., 2013), imbalance handling methods using public binary imbalanced datasets (Zhang et al., 2019), classification of foot drop gait (Bidabadi et al., 2019), and prediction of nanoparticle in vitro toxicity (Furxhi et al., 2019). From a computational point of view, the IBK prediction and classification is achieved according to the relative node distance of instances from

each category, without prior knowledge of the appropriate K value; and the appropriate K value is determined using a cross validation procedure, and they save and use only the selected instance to generate classification predictions, and the approach is sensitive to the number of irrelevant attributes (Aha et al., 1991). The IBK algorithm is described according to three functions: (*i*) the similarity function, (*ii*) the classification function and (*iii*) the concept description updater (Aha et al., 1991).

2.2.3.2. Instance-based Learning with Entropy-Based Distance Function (KStar)

Instance-based learner with entropy-based distance function (KStar), also called *K** (Cleary and Trigg 1995), is a modified version for the *k*-nearest neighbour (KNN) classification approach, and belong into the category of lazy learning algorithms with the function of generalized distance (Cleary and Trigg 1995; Hall et al., 2009; Frank et al., 2016). The KStar for regression and classification problems has gained much popularity during the last few years and several applications can be found in published literature *(*Ravikumar et al., 2019; Joshuva and Sugumaran 2020; Gao et al., 2019). The KNN algorithm adopts the Euclidean distance as a measure to determine the k-nearest neighbours, while the KStar uses an entropic distance measure based on probability (Cleary and Trigg 1995). It was demonstrated that the used entropy as a distance measure leads to a high regression and classification accuracy and can help in handling symbolic and real-valued data (Gao et al., 2019; Joshuva and Sugumaran 2020*).* The KStar algorithm proceeds by successive summation of the probabilities of the new attribute to all the remaining members of the category and the final selection should be based upon the highest probability (Ravikumar et al., 2019).

2.2.3.3. Lazy Locally Weighted Learning (LWL)

Lazy locally weighted learning (LWL) is a nonparametric regression model belonging to the family of lazy learning and is mainly used for function approximation (Wei *et al.*, 2015). The nonlinear mapping of a dependent variable (*Y*) to a set of independent variables (*x*) can be formulated as follows:

(1)

Where *Y* is the predicted value, λ_i (x_i) is the *ith* attribute of the independent variable *x*; β_i is the coefficient for each λ_i (x_i). Building a nonlinear model using LWL is achieved by making a locally linear regression model, taking into account the similarity or the neighbouring information between the target data point and the past data stored in the dataset; instead of a global model covering the entire space (Atkeson et al., 1997; Wei *et al.*, 2015; Joshuva and Sugumaran 2020; Nakanishi *et al.*, 1997; Shigemori et al., 2011; Hazama and Kano 2015). A final averaging model is then calculated from the weighted average of the individual model as follows:

(2)

where w_k is the weight of each individual linear model, and Y_k is the output of the individual linear model for a specific region of the data space.

2.2.4. Resampling algorithm

2.2.4.1. Bootstrap Aggregating (Bagging)

Bootstrap aggregating (*Bagging*) (Breiman, 1996) belongs in the category of ensemble modeling approaches that are also called resampling techniques, which are mainly adopted as a robust solution for solving the overfitting problem that may occur during the training of machines learning models (Zounemat-Kermani et al., 2020). The idea behind the *Bagging* approach is that it builds several individual sub-models or several decision trees, by drawing several samples of the training subset through bootstrapping by replacement of the original dataset. Consequently, the final response of the model prediction is provided by voting or uniform average (Liu and Chen 2020; Yucalar et al., 2020). One of the major advantages of the *Bagging* method is its high capability of handling robust models with limited data size (Onan et al., 2016a). Bagging is suitable either for classification or regression problems and several applications can be found in the literature (e.g. Tama and Comuzzi, 2019; Abpeykar and Ghatee, 2019).

2.2.4.2. Disjoint Aggregating (Dagging)

The Disjoint Aggregating (Dagging), a well-known resampling integration technique, was introduced by Ting and Witten (1997). Using the Dagging method, the training dataset should be divided into several disjoint subsets rather than bootstrap samples used in the Bagging method (Chen *et al.*, 2016). The Dagging model uses the majority voting for building the final prediction (Tama and Comuzzi, 2019), and a robust model can be obtained by the weak learners being trained on a different subset of the training set (Onan et al., 2016b). From a computational point of view, with *N* patterns constituting the training dataset, the Dagging built *M* subset of data having each one *n* pattern without any common pattern, hence, a unique model is built for each dataset and the final model is selected based on voting strategy (Ting and Witten, 1997; Li and Huang, 2018).

2.2.5. Multiple linear regression (MLR)

In this study, the MLR was used as a benchmark model. The MLR measures the relationship between dependent and independent variables in terms of linear equations. In simple terms, it is a response-dependent variable (y) that is influenced by more than one independent variable $\left(x_i\right)$. An MLR can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_i x_i + \epsilon \quad (3)$$

The terms $\beta_0, \beta_1, \ldots, \beta_i$ are regression coefficients, and ϵ is an error term in the model, which represents effects other than those under experiment control, which were not considered in this study.

2.2.5. Model development

To evaluate the performance of the models, the best input combination was served to the models and modeling procedure was carried out for each algorithm in the Waikato Environment for Knowledge Analysis (WEKA 3.9; The University of Waikato, New Zealand) software. It is well known that the most important and significant step in the application of machine learning models is to find out and establish the best model structure that helps to adequately and efficiently achieve the objectives that are set out for us. In this study, the best model structure (i.e. model's parameter value) was selected by running a commonly used '*trial and error*' procedure (Sharafati et al., 2019), where for the best input-combination, different values for each parameter of each model were tried. Models were initially developed using default parameter values and, according to the objective function (i.e. RMSE between observed and forecasted ST) score in the training step, lower or higher values were arbitrarily explored for the model's parameters until the objective function was minimized and the optimum values were obtained.

2.2.6. Model Evaluation Criteria

The evaluation of the proposed models in the present study was done by calculating the following statistical measures: RMSE, mean absolute error (MAE), percent bias (PBIAS), the Nash-Sutcliffe efficiency (NSE) which depends on both the correlation and the bias, and the Pearson's correlation coefficient (R) (Table 3).

<**Table 3**. Statistical measures formula (where *N* is the number of data, ST_M, ST_F, $\overline{ST}_M$, $\overline{ST}_F$ are the measured, forecasted, mean measured, and mean forecasted soil temperature (°C), respectively).>

RMSE and MAE measure the error of the model and represent the difference between forecasted and measured/observed data. RMSE considers more weight to the extreme errors whereas MAE is calculated according to the average magnitude of the error without taking into account the direction of error. Therefore, RMSE and MAE provide a complimentary analysis of the model's error (Shiri and Kişi 2012). NSE is a normalized/dimensionless metric that computes the relative magnitude of the residual variance compared to the variance of the measured dataset. PBIAS calculates the mean tendency of the forecasted values with respect to the observed data, reflecting whether predictions are, on average, higher or lower than the true value. Except for the aforementioned quantitative metric, different visual criteria were applied for further analysis including scatter plot, box plot and Taylor diagram. These criteria allow making a fast and simple comparison. Some of these visual metrics have some advantages over quantitative metrics, such as investigating and comparing between extremes, median and first and third quartile of forecasted and measured values.

3. Results

3.1. Most effective input variables

It is observed that the correlation between the air temperature and ST (Table 2) is more important than other variables e.g. *Eva*, *SSH* and *SR*. However, the correlation between soil and air temperature decreased with increasing soil depth which is in agreement with previous studies (e.g. Salamene et al., 2010). The influence of *Eva*, *SSH* and *SR* also decreased in the deeper layers. The result of the correlation coefficient between the input variables and output revealed that the T_{Mean} is the most effective input variable to forecast ST, followed by T_{Max}, T_{Min}, *Eva*, *SSH*, and *SR* (Tables 4 and 5).

After running the models with different inputs, the *r*-metric was calculated for the training and testing steps between the measured and forecasted STs (Tables 4 and 5). For 3 and 6 days ahead at a depth of 5 cm, T_{Mean} was the best input variable for all developed models except BA-KStar. The BA-KStar model predicted (for 3 and 6 days ahead, at 5 cm soil depth) for scenario No. 2 (T_{Mean} and T_{Max}) and scenario No. 5 (T_{Mean}, T_{Max}, T_{Min}, *Eva, SSH*) the highest *r*, 0.962 and 0.948, respectively, in the test step. T_{Mean} was also the best input variable for BA-IBK- and BA-LWL-based 9 days ahead of ST at 5 cm forecasting while the BA-KStar gives the highest R (0.936) in the test step by considering the T_{Mean}, T_{Max}, T_{Min}, *Eva, SSH* as the inputs (No. 5).

For 3, 6 and 9 days ahead, 50 cm soil depth, T_{Mean} was the most suitable input variable for BA-IBK and BA-LWL models. While, scenario No. 6 (T_{Mean}, T_{Max}, T_{Min}, *Eva, SSH, SR*) gave the best performance for all BA-KStar. T_{Mean}, T_{Max} variables are the most powerful input variables to obtain the highest *r* for all DA-based algorithms for ST forecasting 3 and 6 days ahead at a depth of 50 cm, while the T_{Mean} (scenario No.1*)* obtained the highest *r* of 0.954, 0.956, 0.934 in testing step for ST at 50 cm forecasting for 9 days ahead, as the best input variable.

It can be stated that the considered assumption would be true in most of the cases and that the variables with the highest *r* values would be the most powerful variable in improving model performance and can predict ST accurately, especially for 5 cm soil depth. It is obvious from Tables 4 and 5 that by using T_{Mean} variable, ST can be predicted easily for a soil depth of 5 cm, while for deeper soil depth levels, due to more complexity, a combination of input variables is required.

< **Tabel 4.** Determination of the best input variable combination for 3, 6 and 9 days ahead forecasting of the soil temperature at 5 cm soil depth using R-value >

<**Tabel 5**. Determination of the best input variable combination for 3, 6 and 9 days ahead forecasting of the soil temperature at 50 cm soil depth using R-value >

3.2. Model Evaluation and Comparison

As models developed based on the training dataset, the results presented in this section only show how the developed algorithms fit with the dataset. Hence, the results of testing data that represent how the developed algorithms are suitable for forecasting ST, used for model evaluation and comparison purposes.

Table 6 lists the performance of the developed models for forecasting ST for 3, 6 and 9 days at a depth of 5 cm based on statistical criteria in the test step. It is observed that the BA-KStar model with an RMSE of 3.15°C, an MAE of 2.46°C, and an NSE of 0.9 outperforms the other models, followed by DA-IBK, BA-IBK, DA-KStar, DA-LWL and BALWL in ST (t + 3) forecasting. The performance of the developed models for forecasting soil temperature for 6 days at a depth of 5 cm is similar to the model ranking results for the 3-day ahead forecasting. The BA-KStar with an RMSE of 3.56°C, an MAE of 2.67°C and an NSE of 0.873 resulted in the best performance while the BA-LWL obtained the highest error with an RMSE of 4.59°C, an MAE of 3.51°C and lowest correlation with an NSE of 0.789. In terms of the 9 days forecasting ST models' performance, the DA-KStar (with RMSE = 3.78°C, MAE = 2.81°C and NSE = 0.858) outperforms the others, followed by the BA-KStar, DA-IBK, BA-IBK, BA-LWL, and DA-LWL. It is observed that the BA-IBK and BA-LWL have performed similarly in ST (t + 9) forecasting at a depth of 5 cm. The performance of the model for BPIAS is considered very good, good, satisfactory with $< \pm 10$, $\pm 10 \leq$ PBIAS $< \pm 15$, $\pm 15 \leq$ PBIAS $< \pm 25$, and PBIAS $> \pm 25$, respectively (Legates and Mccabe, 1999). Therefore, the performance of all developed models for ST forecasting at a depth of 5 cm is categorized as excellent. Moreover, the PBIAS values for all the models are negative, indicating the slight overestimation of ST values at a depth of 5 cm.

A scatter plot (Fig. 2) is used to show the linear relationship between the ST forecasted (y) and measured (x) values visually. The scatter plot illustrates that the results are in agreement with the calculated statistical metrics in Table 5. It is observed that the forecasted values through the BA-KStar model for ST (t + 3) at a depth of 5 cm show less scatter than the other models for this depth, and hence are much closer to the observed values. BA-LWL's ST forecasting at 5 cm shows a greater scatter, and therefore, is farther from measured ST (t + 3) values. For the 6 days ST forecasting, the forecasted values obtained using the BA-KStar model are also much closer to the measured ST (t + 6) values than for the other models, showing the highest R^2 (0.899) and a well-fitted linear regression equation. The measured and forecasted ST (t + 9) values are concentrated around the 1:1 ideal line ($y = x$) in the DA-KStar model while the points are mostly scattered in the BA-LWL model.

The visually models' performance evaluation using the Taylor diagram (Fig. 3), based on the correlation and standard deviation values, confirms that the BA-KStar and BA-LWL models result in the closest and farthest values to the measured soil temperature in 3 days ahead forecasting respectively. The Taylor diagram confirms that the BA-KStar model with a correlation of 0.94 and a standard deviation of 0.92 between forecasted and measured data shows better results compared to other developed models in ST (t + 6). Moreover, the forecasted BA-IBK values are close to the measured ST (t + 9) at 5 cm while the DA-LWL is not able to forecast the minimum and maximum ST (t + 9) levels properly.

The distributions of the measured and forecasted ST data at 5 cm depth are presented as box plots in Fig. 4. The box plot shows that the BA-IBK model can forecast the extreme values (minimum and maximum) ST values much better than other developed models. The BA- and DA-LWL models cannot capture the maximum and minimum measured soil temperature values for 3 days ahead of forecasting. Generally, the BA-KStar model forecasted ST (t + 6) most similar to the measured values based on

medians, quantiles, minimum and maximum parameters. Although the median and quartiles of the LWL-based models are relatively close to the measured ST, they cannot forecast the minimum and maximum ST values.

<**Table 6.** Model evaluation through quantitative metrics for 3, 6 and 9 days ahead soil temperature forecasting at a depth of 5 cm>

<**Fig. 2.** Scatter plot of 3 (i.e. red color), 6 (i.e. green color) and 9 (i.e. golden color) days ahead forecasted *VS*. measured soil temperature at a depth of 50 cm>

<**Fig. 3.** Model performance via Taylor diagram for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 5 cm>

<**Fig. 4.** Model performance using box-plot for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 5 cm>

According to error measurement criteria (RMSE and MAE) and NSE values, the DA-KStar model has the higher forecasting power with an RMSE of 2.29°C, MAE of 1.72°C and NSE of 0.887 while the BA-LWL shows the highest error (RMSE = 3.26 C° and MAE = 2.53°C) and lowest NSE (0.772) for ST (t + 3) at a depth of 50 cm (Table 7). These performances are also similar to the 6 and 9 days ahead ST forecasting at a depth of 50 cm. In other words, generally, the DA-Kstar model outperformed the other models for all time horizons (3, 6 and 9 days ahead) while the BA-LWL model had the worst performance. NSE is used to rank the efficiency of developed models from highest forecasting power to the lowest as DA-KStar > BA-KStar > DA-IBK > BA-IBK > DA-LWL > BA-LWL for 3 days ahead, as DA-KStar > DA-IBK > BA-IBK > BA-KStar > DA-LWL > BA-LWL for 6 days ahead and as DA-KStar > DA-IBK > BA-IBK = BA-KStar > DA-LWL > BA-LWL for ST for 9 days ahead ST forecasting at depth of 50 cm. NSE shows that all applied models for TS forecasting have an excellent forecasting power that has $0.75 < NSE \leq 1$ as recommended by Moriasi *et al.*, (2007).

Similar to the ST forecasting at depth 5 cm, the PBIAS metric indicates that all the developed models are suitable for soil temperature forecasting at a depth of 50 cm. The negative value for all calculated PBIAS suggests an overestimation of ST values at a depth of 50 cm. The ideal model has a PBIAS value closer to 0. Therefore, BA-IBK, DA-IBK and DA-LWL have the best performance based on this metric. Each model evaluation method has its own advantages and disadvantages. Thus, selecting the best model using different criteria is a challenging task (Khosravi et al., 2018). The current study considered all of these criteria in order to choose the best model.

It is observed from scatter plots (Fig. 5) that the forecasted ST values for the 3 days ahead modelling obtained using the DA-KStar model were much closer to the measured ST values than for the other models, showing the highest R^2 (0.916). For 6 days ahead, the point clouds of the scatter plots are less scattered compared to the 3 days ahead ST forecasting. For 9 days ahead of ST forecasting, the

measured *vs.* forecasted values are concentrated around the 1:1 line in the DA-KStar model (R^2 = 0.924), while the point clouds are mostly scattered in the BA-LWL model (R^2 = 0.838).

The Taylor diagrams (Fig. 6) illustrate the performance of the developed models for different time horizons (3, 6 and 9 days ahead) ST forecasting at a depth of 50 cm. The DA-KStar, BA-KStar, DA-IBK models correlate > 0.95, while the rest of the models correlate < 0.95. The DA-KStar model is much closer to the measured values indicating the superior performance of this algorithm. Both the BA-LWL and DA-LWL models are placed farther from the measured value. Therefore, these models show the worst performance.

Box plots of measured and forecasted ST 3, 6 and 9 days ahead using all the different models are shown in Fig. 7 (a, b, c). The median values of LWL-based models for 3 and 6 days ahead are further away from the median value for the measured ST. Moreover, they cannot forecast the minimum and maximum levels of the ST properly. Although the median values of the LWL-based models for the 9 days ahead are much closer than those in 3 and 6 days ahead, they still are not able to forecast the extreme (minimum and maximum) ST (t + 9) values. Generally, forecasted ST values of the KStar-based models are much close to the measured ST values, based on minimum, maximum, medians and quartiles compared to the IBK-based models.

<**Table 7.** Model evaluation through quantitative metrics for 3, 6 and 9 days ahead ST forecasting at a depth of 50 cm>

<**Fig. 5.** Scatter plots of 3 (i.e. red color), 6 (i.e. green color) and 9 (i.e. golden color) days ahead forecasted versus measured ST at a depth of 50 cm>

<**Fig. 6.** Model performance via Taylor diagram for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 5 cm>

<**Fig. 7.** Model performance using box-plot for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 50 cm >

4. Discussion

Timing for planting field crops is partially based on when soil temperatures are optimal for seed germination and seedling emergence (Lindstrom *et al.*, 1976). Therefore, the timing of spring crop planting is dependent on a number of variables. Apart from agronomic variables like seed availability, equipment readiness, field dryness, tillage possibility, another important factor is soil temperature. As direct soil measurement is time-consuming and empirical models are not accurate enough, the results of this study provide a basis for ST forecasting at different depths (e.g. 5 and 50 cm).

The different model characteristics may need diverse input variables, so a suitable selection of variables is a key factor for accurate ST forecasting. Inappropriate and irrelevant variables complicate the

development of models and decrease the performance of the model by creating a complicated relationship between the inputs and output. Although previous studies have used data with high correlation for developing machine learning models (e.g. Barzegar et al., 2017), this study confirms the findings of the survey by Khosravi et al., (2020b) in which they concluded that both high and low correlation variables should be considered for developing a high-performance predictive model. Besides, each developed data mining algorithm has a different input variable as the best input-combination as well as behaves differently with the same input-combination due to the different structure of each algorithm and different learning processes. This different structure causes different computing capabilities and then different prediction power.

Statistical performance of the developed models for ST forecasting at soil depths of 5 and 50 cm shows that they are capable of using as promising tools for not only ST forecasting but also for other catchment phenomena. The results showed that all developed models performed very well to forecast ST at multiple depths. In addition, the results of the study were compared with those of the MLR model as a benchmark (Table 8). The results show that all developed new ML models outperform MLR models. As a result of its simple and linear structure, the MLR model is not capable of capturing non-linearity between input variables and ST.

<Table 8. Model evaluation for 3, 6 and 9 days ahead ST forecasting at a depth of 50 cm using LR model>

The correlation analysis indicated that ST is correlated well with the air temperature, in particular with the T_{Mean} and T_{Max}. In general, most of the models' performance based on air temperature parameters (mainly T_{Mean} and T_{Max}) obtained the lowest error between the measured and forecasted ST. It indicates that the ST is mainly controlled by T_{Mean} followed by T_{Max}. So, it can be concluded that T_{Max} can have a more severe effect on the ST than T_{Min}. However, some models (e.g. BA-KStar for 6 and 9 days ahead forecasting) also performed very well by using the air temperature components and other variables like *SSH*, indicating the important role played by other meteorological variables. The significant effect of the air temperature on ST can be due to the energy balance at the ground surface (Zheng et al., 1993). It can be stated that ST has its own regional characteristics which make it difficult to directly compare the performance between prediction models constructed using different regions and their corresponding datasets. Therefore, an ST forecasting model should be provided for each region based on their specific meteorology.

Recently, Feng et al., (2019) indicated that with the increase of soil depth, the differences between the measured and estimated ST using the ELM, GRNN, BPNN and RF became larger. However, this study is not in agreement with their conclusion. This study shows that the models used in this study (BA- and DA-based KStar, LWL and IBK) even decrease the differences between the measured and estimated ST as the depth increased. In another research by Samadianfar *et al.*, (2018), ST was predicted at multiple depths (5, 10, 20, 50 and 100 cm) using hybrid MLP-FFA and standalone MLP models. They concluded that the prediction error was reduced by increasing soil depth, which is in accordance with the results of the

current study. Tabari et al., (2015) also carried out research to forecast ST one day ahead at different depths and their study outcome agrees with the results of the current study. The increase in predictability of ST with the depth is due to the stronger memory of the ST time series in the deeper soil layers. The ST at depth is less affected by variations of surface meteorological conditions; therefore, its fluctuation is minor and smooth during the day which leads to easier forecasting (Tabari et al., 2015).

Chen et al., (2019) recently indicated that the complexity of a hybrid model is higher than that of a single model. However, the learning performance can be remarkably enhanced to a certain extent. Therefore, the hybrid resampling-ML model cannot only decrease the fitting error, but could also ameliorate the generalization ability.

The main limitation of the applied models is their site-specific calibration, and thus, they do not have generalization power to predict/forecast the target phenomena in another study area, especially when the physiographic conditions differ significantly. As a result, ML models are sensitive to the learning section, and therefore, in order to build a model that is more robust, it is important to include a large and long dataset from multiple study domains. However, the main advantage of these models is their ability to forecast each target phenomenon at each site.

In this study, the capability of the developed models was tested using the available meteorological variables. However, it is suggested to develop the used models in this study or other suitable ML models e.g., deep learning models using other different meteorological variables (atmospheric pressure, precipitation, relative humidity, wind speed, etc.), which may provide additional information. In a study by Hu and Feng (2005), the effect of precipitation on soil temperatures was insignificant. Precipitation and other climate variables, however, may improve the accuracy of the ML models. The models also need to be further employed in different areas with different climatic characteristics and land uses (rangeland, arable land, forest, sub-urban areas) to validate the efficiency of the models in ST forecasting. Also, It is recommended to develop these types of algorithms in one study area and then use them to evaluate forecasted ST in another. As long as a model's predictive power is high, it can serve as an effective and simple tool for ST forecasting worldwide. **5. Conclusions**

This study successfully addressed the challenge of accurately forecasting ST through six novel hybrid algorithms of resampling methods (e.g. bagging (BA) and dagging (DA)) with machine learnings (MLs) (e.g. KStar, instance-based K-nearest learning (IBK) and locally weighted learner (LWL)) including BA-KStar, BA-IBK, BA-LWL, DA-KStar, DA-IBK and DA-LWL at two different soil depths, 5 and 50 cm, for the first time. The main goal was to develop algorithms with reasonable prediction power and propose them as simple and promising tools for multistep (up to nine days ahead) ST forecasting in the Isfahan area, Iran, as a case study.

The main findings of this research are as follows:

1. The modeling process based on correlation coefficient shows that mean air temperature (T_{Mean}) is the most effective input variable for ST forecast followed by maximum air temperature (T_{Max}),

minimum air temperature (T_{Min}), evaporation (*Eva*), sun-shine hours (*SSH*) and solar radiation (*SR*).

2. The input variables with the highest correlation coefficient with the output have the ability to forecast ST with high accuracy, particularly at a shallower depth (5 cm). However, in the deeper zone (50 cm) of the soil, the ST seems complex and a combination of several variables should be used to obtain accurate forecasts.
3. In terms of model performance, results illustrate that all developed hybrid models have very good performance based on quantitative metrics (i.e. Nash-Sutcliff efficiency (NSE) and PBIAS).
4. Among the different models, KStar is the best hybrid model combined with BA, while IBK and LWL have a higher prediction power with DA compared to BA in most cases.
5. The results for the soil depth of 5 cm showed that BA-KStar is the most effective algorithm (NSE = 0.90, 0.87 and 0.85 for 3, 6 and 9 days ahead forecasting), while for a soil depth of 50 cm, the DA-KStar algorithm outperformed the other algorithms (NSE = 0.88, 0.89 and 0.89 for 3, 6 and 9 days ahead forecasting).
6. ML models perform better in all cases than linear regression models.
7. To sum up, the BA-KStar and DA-KStar algorithms are cost-effective tools that can forecast ST reasonably using only the T_{Mean} as input. Therefore, costly and time-consuming direct measurements of ST can be replaced by low-cost and quick indirect proposed ML methods.

Declarations

Acknowledgements

The authors would like to thank Dr. Esmaeel Dodangeh for sharing the dataset. We also would like to extend our gratitude to the editor-in-chief and two anonymous reviewers for their help enhancing the quality of the paper during the review process.

Funding

Khabat Khosravi was partially supported by a grant from Ferdowsi University of Mashhad (No. FUM-…….).

Competing interests: The authors declare no competing interests.

References

1. Abpeykar S, & Ghatee M. 2019. Neural trees with peer-to-peer and server-to-client knowledge transferring models for high-dimensional data classification. Expert Syst. Appl. Expert Systems with Applications, **137**, 281–291. https://doi.org/10.1016/j.eswa.2019.07.003
2. Aha D W, Kibler D, & Albert M K. 1991. Instance-based learning algorithms. Mach. Learn., **6**(1), 37–66. https://doi.org/10.1007/BF00153759.

3. Araghi A, Mousavi-Baygi, M, Adamowski J, Martinez C. and van der Ploeg M. 2017. Forecasting soil temperature based on surface air temperature using a wavelet artificial neural network. Meteorol. App, 24(4), pp.603–611.
4. Atkeson CG, Moore AW, & Schaal S. 1997. Locally weighted learning for control, Lazy learning, Springer, Netherlands, pp.75–113.
5. Badache M, Eslami-Nejad P, Ouzzane M, Aidoun Z, and Lamarche L. 2016. A new modeling approach for improved ground temperature profile determination. Renew. Energy, 85, pp.436–444.
6. Barzegar R., Moghaddam A. A, Adamowski J, & Fijani E. 2017. Comparison of machine learning models for predicting fluoride contamination in groundwater. *Stochastic Stoch Environ Res Risk Assess*, *31*(10), 2705–2718.
7. Belghit A, and Benyaich M. 2014. Numerical Study of Heat Transfer and Contaminant Transport in an Unsaturated Porous Soil. *Water Resour. Prot.*, 6(13), p.1238.
8. Bidabadi S S, Murray I, Lee G Y F, Morris S, & Tan T. 2019. Classification of foot drop gait characteristic due to lumbar radiculopathy using machine learning algorithms. Gait Posture, *71*, 234–240. https://doi.org/10.1016/j.gaitpost.2019.05.010
9. Bond-Lamberty B, Wang C, Gower S T. 2005. Spatiotemporal measurement and modeling of stand-level boreal forest soil temperatures. Agric. For. Meteorol. 131: 27–40.
10. Brar GS, Steiner JL, Unger PW, and Prihar SS. 1992. Modeling sorghum seedling establishment from soil wetness and temperature of drying seed zones. Agron J: 905–910.
11. Breiman L. 1996. Bagging predictors. Mach. Learn, *24*(2), 123–140.https://doi.org/10.1007/BF00058655
12. Buckingham D, Skalka C. and Bongard J. 2015. Inductive machine learning for improved estimation of catchment-scale snow water equivalent. J Hydro, 524, pp.311–325.
13. Chen W, Hong H, Li S, Shahabi H, Wang Y, Wang X, & Ahmad B B. 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. J Hydro, 575, 864–873.
14. Chen W, Zhao X, Tsangaratos P, Shahabi H, Ilia I, Xue W, Wang X, Ahmad B B. 2020. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. J Hydro, *583*, 124602. https://doi.org/10.1016/j.jhydrol.2020.124602.
15. Citakoglu H. 2017. Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. Theor. Appl. Climatol., 130 (1–2):545–556.
16. Cleall PJ, Muñoz-Criollo JJ, and Rees SW. 2015. Analytical solutions for ground temperature profiles and stored energy using meteorological data. Transp. Porous Media, 106(1), pp.181–199.
17. Cleary JG, & Trigg LE. 1995. K*: An instance-based learner using an entropic distance measure. In *Machine Learning Proceedings 1995* (pp. 108–114). Morgan Kaufmann.http://www.cms.waikato.ac.nz/~ml/publications/1995/Cleary95-KStar.pdf.

18. Ebrahimi M, Sarikhani MR, Safari Sinegani AA, Ahmadi A, Keesstra S. 2019. Estimating the soil respiration under different land uses using artificial neural network and linear regression models. Catena 174, 371–382. https://doi.org/10.1016/j.catena.2018.11.035
19. Feng Y, Cui N, Hao W, Gao L, and Gong D. 2019. Estimation of soil temperature from meteorological data using different machine learning models. Geoderma, 338, pp.67–77.
20. Frank E, Hall, MA, & Witten IH. 2016. The WEKA Workbench. Online Appendix for" Data Mining: Practical Machine Learning Tools and Techniques".
21. Furxhi I, Murphy F, Mullins M, & Poland CA. 2019. Machine learning prediction of nanoparticle in vitro toxicity: A comparative study of classifiers and ensemble-classifiers using the Copeland Index. Toxicol. Lett, *312*, 157–166.https://doi.org/10.1016/j.toxlet.2019.05.016.
22. Gao W, Alsarraf J, Moayedi H, Shahsavar A, & Nguyen H. 2019. Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. Appl. Soft Comput., *84*, 105748.https://doi.org/10.1016/j.asoc.2019.105748
23. Gao Y, Fan R, Li H, Liu R, Lin X, Guo H. and Gao Y. 2016. Thermal performance improvement of a horizontal ground-coupled heat exchanger by rainwater harvest. Ener. Buil, 110, pp.302–313.
24. Granata F, Saroli m, De Marinis G, Gargano R. 2018. Machine Learning Models for Spring Discharge Forecasting. *Geofluid*, Article ID 8328167 | https://doi.org/10.1155/2018/8328167
25. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P. & Witten I H. 2009. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, *11*(1), 10–18.
26. Hazama K, & Kano M. 2015. Covariance-based locally weighted partial least squares for high-performance adaptive modeling. Chemometr Intell Lab Syst., *146*, 55–62.http://dx.doi.org/10.1016/j.chemolab.2015.05.007
27. Hu G, Lin Z, Wu X, Ren L, Wu T, Xie C, Qiao Y, Shi J, Cheng G. 2016. An analytical model for estimating soil temperature profiles on the Qinghai-Tibet plateau of China. J. Arid. Land 8 (2), 232–240.
28. Hussain D, Khan AA. 2020. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. Earth Sci. Inform. doi:10.1007/s12145-020-00450-z
29. Jaafar WW, & Han D. 2012. Variable selection using the gamma test forward and backward selections. J Hydra.Eng, *17*(1), 182–190.
30. Joshuva A, & Sugumaran V. 2020. A lazy learning approach for condition monitoring of wind turbine blade using vibration signals and histogram features. Measurement, *152*, 107295. https://doi.org/10.1016/j.measurement.2019.107295
31. Kang S, Kim S, Oh S, Lee D. 2000. Predicting spatial and temporal patterns of soil temperature based on topography, surface cover and air temperature. For. Ecol. Manage. 136: 173–184.
32. Khosravi K, Barzegar R, Miraki S, Adamowski J, Daggupati P, Alizadeh MR, Pham B, Taghi Alami M. 2020b. Stochastic Modeling of Groundwater Fluoride Contamination: Introducing Lazy Learners. Groundwater, 58(5):723–734.

33. Khosravi K, Cooper JR., Daggupati P, Pham B, Bui D. 2020a. Bedload transport rate prediction: Application of novel hybrid data mining techniques. J Hydro, 124774
34. Khosravi K, Daggupati P, Taghi Alami M, Muhammad Awadh S, Ismaeel Ghareb M, Panahi M, Pham B, Rezaie F, Qi C, Yaseen. Z. 2019. Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. Comput Electron Agric 167, 105041.
35. Khosravi K, Mao L, Kisi O, Yaseen ZM. and Shahid S. 2018. Quantifying hourly suspended sediment load using data mining models: case study of a glacierized Andean catchment in Chile. J Hydro, 567, pp.165–179.
36. Khosravi K, Barzegar R, Miraki S, Adamowski J, Daggupati P, Alizadeh MR, Pham BT. and Alami MT. 2020. Stochastic modeling of groundwater fluoride contamination: Introducing lazy learners. Groundwater, *58*(5), pp.723–734.
37. Li, J., & Huang,T.(2018).Predicting and analyzing early wake-up associated gene expressions by integrating GWAS and eQTL studies. Biochim Biophys Acta Mol Basis Dis, *1864*(6), 2241–2246.http://dx.doi.org/10.1016/j.bbadis.2017.10.036
38. Liu X. and Huang B. 2005. Root physiological factors involved in cool-season grass response to high soil temperature. Environ. Exp. Bot., 53, 3: 233–245.
39. Liu BC, LiuW. and Peng SW. 2005. Study of heat and moisture transfer in soil with a dry surface layer. Int. J. Heat Mass Transf., 48(21–22):4579–4589.
40. Liu H, & Chen SM. 2020. Heuristic creation of deep rule ensemble through iterative expansion of feature space. Inf. Sci.. https://doi.org/10.1016/j.ins.2020.02.001.
41. Luo X, Lin F, Chen Y, Zhu S, Xu Z, Huo Z, Y M, Peng J. 2019. Coupling logistic model tree and random subspace to predict the landslide susceptibility areas with considering the uncertainty of environmental features. *Sci Rep* : 9, Article number: 15369 (2019)
42. Maryanaji Z, Merrikhpour H. and Abbasi H. 2017. Predicting soil temperature by applying atmosphere general circulation data in west Iran. J Water Clim Chan, 8(2), pp.203–218.
43. Nakanishi J, Farrell J A, & Schaal S. 2005. Composite adaptive control with locally weighted statistical learning. Neural Netw, *18*(1), 71–90.http://dx.doi.org/10.1016/j.neunet.2004.08.009
44. Napagoda NA, and Tilakaratne CD. 2012. Artificial neural network approach for modeling of soil temperature: a case study for Bathalagoda area. Sri Lankan J Appl Stat, 13, pp.39–59.
45. Onan A, Korukoğlu S, & Bulut H. 2016a. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Syst. Appl., *62*, 1–16. http://dx.doi.org/10.1016/j.eswa.2016.06.005.
46. Onan A, Korukoğlu S, & Bulut H. 2016b. Ensemble of keyword extraction methods and classifiers in text classification. Expert Syst. Appl., *57*, 232–247.http://dx.doi.org/10.1016/j.eswa.2016.03.045.
47. Ozgener O, Ozgener L, and Tester, JW. 2013. A practical approach to predict soil temperature variations for geothermal (ground) heat exchangers applications. Int. J. Heat Mass Transf., 62, pp.473–480.

48. Ozturk M, Salman O, Koc M. 2011. Artificial neural network model for estimating the soil temperature. Can. J. Soil Sci. 91, 551–562.
49. Paul KI, Polglase PJ, Smethurst PJ, O'Connell AM, Carlyle CJ, and Khanna PK. 2004. Soil temperature under forests: a simple model for predicting soil temperature under a range of forest types. Agric. For. Meteorol., 121(3–4), pp.167–182.
50. Peng S, Piao S, Wang T, Sun J, Shen Z. 2009. Temperature sensitivity of soil respiration in different ecosystems in China. Soil Biol. Biochem. 41, 1008–1014. http://dx.doi.org/10.1016/j.soilbio.2008.10.023.
51. Pham B, et al., 2020. A Comparative Study of Kernel Logistic Regression, Radial Basis Function Classifier, Multinomial Naïve Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping. Water, 12, 239; doi:10.3390/w12010239.
52. Pourghasemi HR, Kariminejad N, Amiri M, Edalat M, Zarafshar,M, Blaschke T, Cerda A. 2020. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Sci. Rep. 10, 1–11. https://doi.org/10.1038/s41598-020-60191-3
53. Pourghasemi HR, Yousefi S, Kornejady A, Cerdà A. 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Sci. Total Environ. 609, 764–775. https://doi.org/10.1016/j.scitotenv.2017.07.198
54. Qian B, Gregorich EG, Gameda S, Hopkins DW, and Wang XL. 2011. Observed soil temperature trends associated with climate change in Canada. J. Geophys. Res. Atmos., 116(D2).
55. Ravikumar S, Kanagasabapathy H, & Muralidharan V. 2019. Fault diagnosis of self-aligning troughing rollers in belt conveyor system using k-star algorithm. Measurement, *133*, 341–349.https://doi.org/10.1016/j.measurement.2018.10.001.
56. Razavi-Termenh SV, Sadeghi-Niaraki A, Choi S. 2019. Groundwater Potential Mapping Using an Integrated Ensemble of Three Bivariate Statistical Models with Random Forest and Logistic Model Tree Models. Water, *11*(8), 1596; https://doi.org/10.3390/w11081596
57. Salamene S, Francelino,MR, Santana RM, Schaefer CEGR, & Setzer AW. 2010. Correlation between atmospheric physical factors and soil temperature of Keller Peninsula, King George Island, Antarctica. In Proceedings 19th World Congress of Soil Sciences, Brisbane, Australia: Soil Solutions for a Changing World (pp. 13–14).
58. Salih SQ, sharafati A, Khosravi K, Faris H, Kisi O, Tao H, Ali M, Yaseen Z. 2019. River suspended sediment load prediction based on river discharge information: application of newly developed data mining models, Hydrol Sci J., 65 (4): 624–637
59. Sattari MT, Avram A, Apaydin H, & Matei O. 2020. Soil Temperature Estimation with Meteorological Parameters by Using Tree-Based Hybrid Data Mining Models. Mathematics, *8*(9), 1407.
60. Samadianfard S, Ghorbani MA. and Mohammadi B. 2018. Forecasting soil temperature at multiple-depth with a hybrid artificial neural network model coupled-hybrid firefly optimizer algorithm. Inf. Process. Agric., 5(4), pp.465–476.

61. Sándor R, and Fodor N. 2012. Simulation of soil temperature dynamics with models using different concepts. *Sci. World* J., 590287 | https://doi.org/10.1100/2012/590287
62. Sanikhani H, Deo RC, Yaseen ZM., Eray O. and Kisi O. 2018. Non-tuned data intelligent model for soil temperature estimation: A new approach. Geoderma, 330, pp.52–64.
63. Sharafati A, Khosravi K, Khosravinia P, Ahmed K, Abdulridha Salman S, Yaseen Z, Shahid S. 2019. The potential of novel data mining models for global solar radiation prediction. Int J Sci Environ Technol 16 (11), 7147–7164
64. Shigemori H, Kano M, & Hasebe S. 2011. Optimum quality design system for steel products through locally weighted regression model. J.Process Control, *21*(2), 293–301. http://dx.doi.org/10.1016/j.jprocont.2010.06.022.
65. Smusz, S., Kurczab, R., & Bojarski, A.J. (2013). A multidimensional analysis of machine learning methods performance in the classification of bioactive compounds. Chemom. Intell. Lab.Syst., *128*, 89–100.http://dx.doi.org/10.1016/j.chemolab.2013.08.003.
66. Tabari H, Hosseinzadeh Talaee P, & Willems P. 2015. Shortterm forecasting of soil temperature using artificial neural network. Meteorol. Appl, *22*(3), 576–585.
67. Tama BA, & Comuzzi M. 2019. An empirical comparison of classification techniques for next event prediction using business process event logs. Expert Syst. Appl., *129*, 233–245.https://doi.org/10.1016/j.eswa.2019.04.016.
68. Ting KM, Witten IH. 1997. Stacking bagged and dagged models. (Working paper 97/09). University of Waikato, Department of Computer Science, Hamilton, New Zealand.
69. Wei C. 2015. Comparing lazy and eager learning models for water level forecasting in river-reservoir basins of inundation regions. Environ Model Softw, *63*, 137–155. http://dx.doi.org/10.1016/j.envsoft.2014.09.026.
70. Xing L, Li L, Gong J, Ren C, Liu J, Chen H. 2018. Daily Soil Temperatures Predictions for Various Climates in United States Using Data-Driven Model. Energy, 160: 430–440. https://www.sciencedirect.com/science/article/abs/pii/S0360544218312921
71. Yucalar F, Ozcift A, Borandag E, & Kilinc D. 2020. Multiple-classifiers in software quality engineering: Combining predictors to improve software fault prediction ability. Eng. Sci. Technol.https://doi.org/10.1016/j.jestch.2019.10.005.
72. Zeynoddin M, Bonakdari H, Ebtehaj I, Esmaeilbeiki F, Gharabaghi B. and Haghi DZ, 2019. A reliable linear stochastic daily soil temperature forecast model. Soil Tillage Res, 189, pp.73–87.
73. Zhang X, Li R, Zhang B, Yang Y, Guo J, & Ji X. 2019. An instance-based learning recommendation algorithm of imbalance handling methods. Appl. Math. Comput, *351*, 204–218. https://doi.org/10.1016/j.amc.2018.12.020.
74. Zheng D, Hunt Jr E R, & Running,SW. 1993. A daily soil temperature model based on air temperature and precipitation for continental applications. Clim. Res., 2(3), 183–191.
75. Zounemat-Kermani M, Stephan D, Barjenbruch M, & Hinkelmann R. 2020. Ensemble data mining modeling in corrosion of concrete sewer: A comparative study of network-based (MLPNN & RBFNN)

and tree-based (RF, CHAID, & CART) models. Adv. Eng. Inform, *43*, 101030. https://doi.org/10.1016/j.aei.2019.101030.

Figures

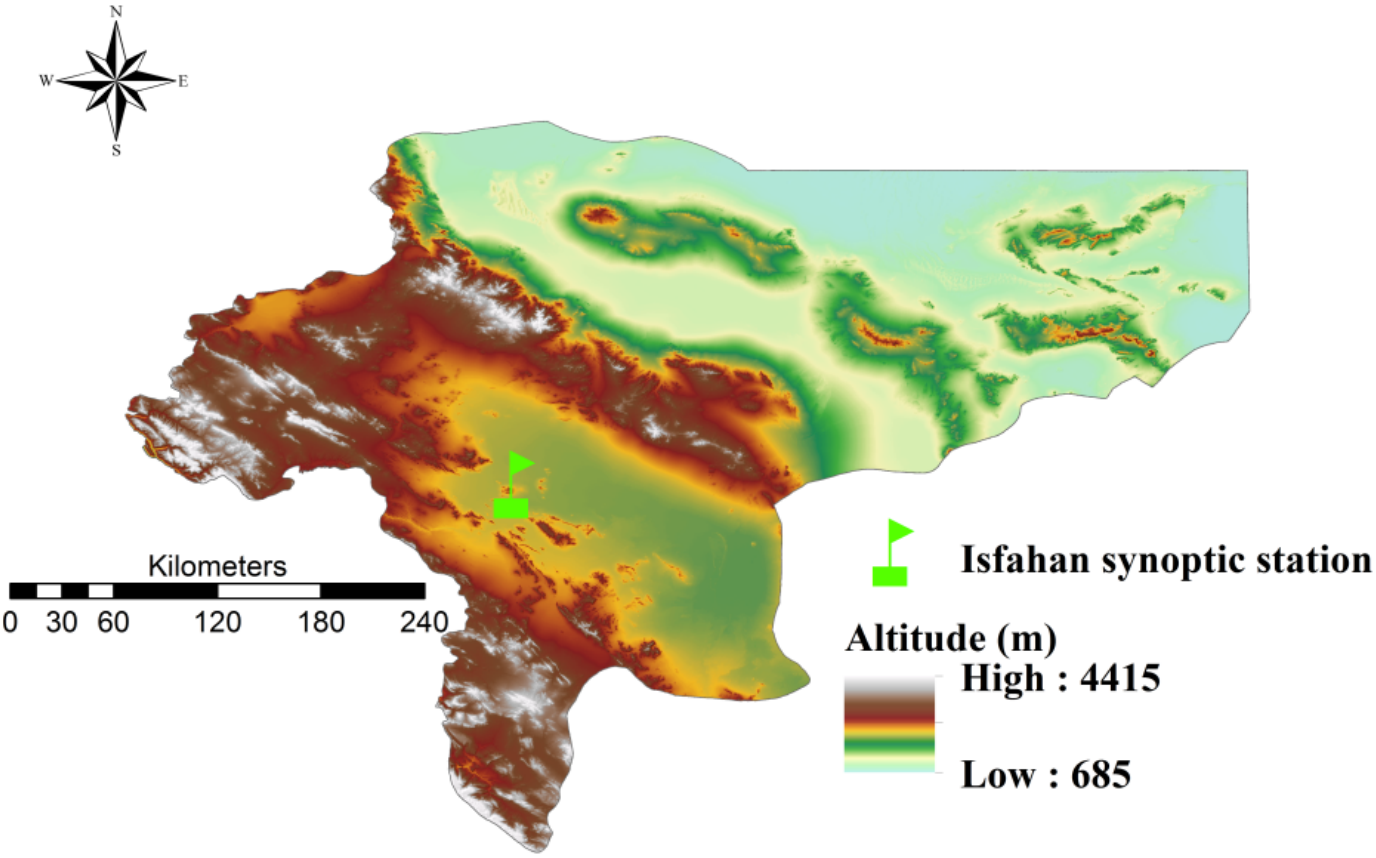

Figure 1

Study area and location of Isfahan synoptic station

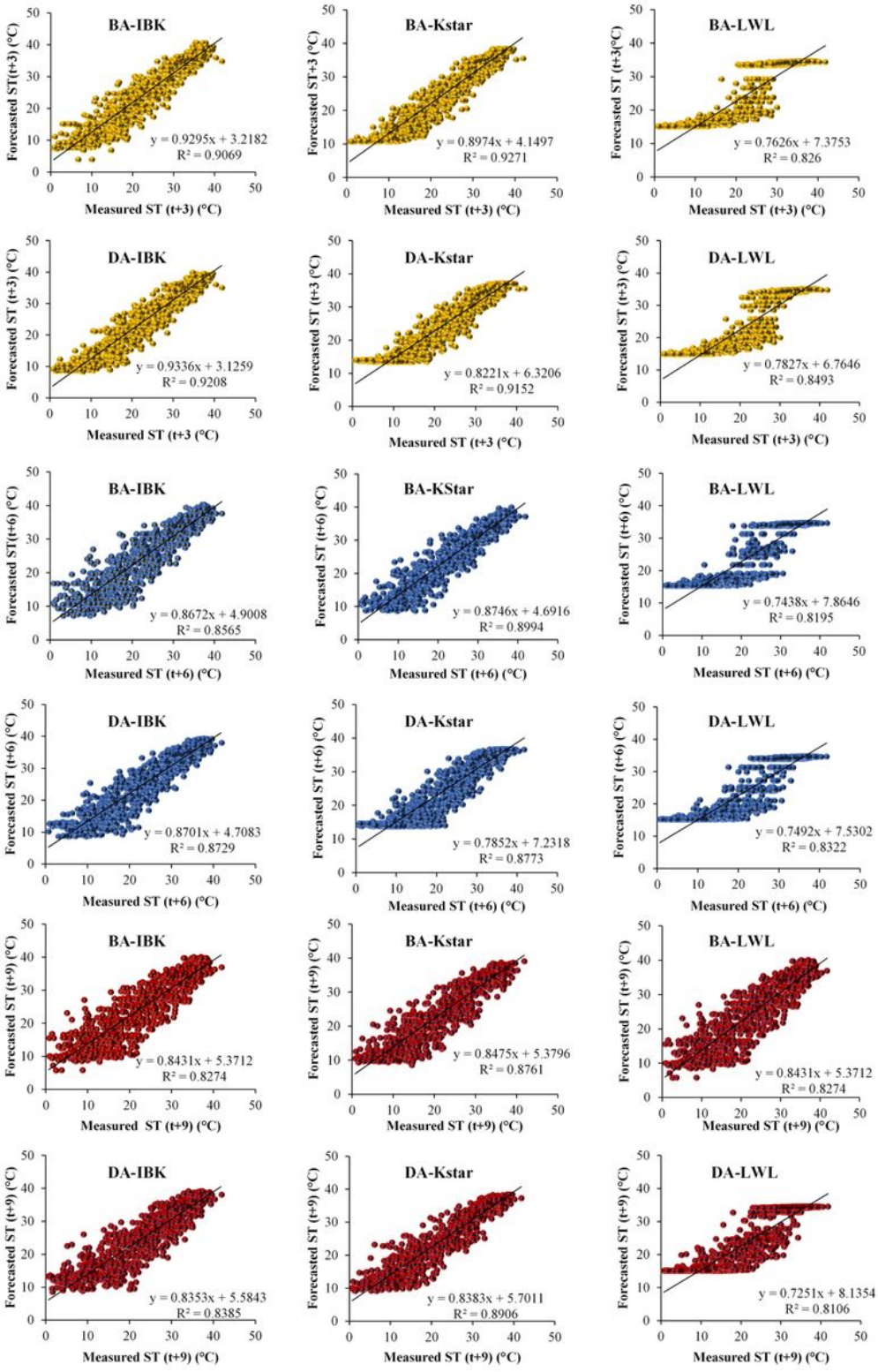

Figure 2

Scatter plot of 3 (i.e. golden color), 6 (i.e. blue color) and 9 (i.e. red color) days ahead forecasted *VS.* measured soil temperature at a depth of 5 cm

Figure 3

Model performance via Taylor diagram for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 5 cm

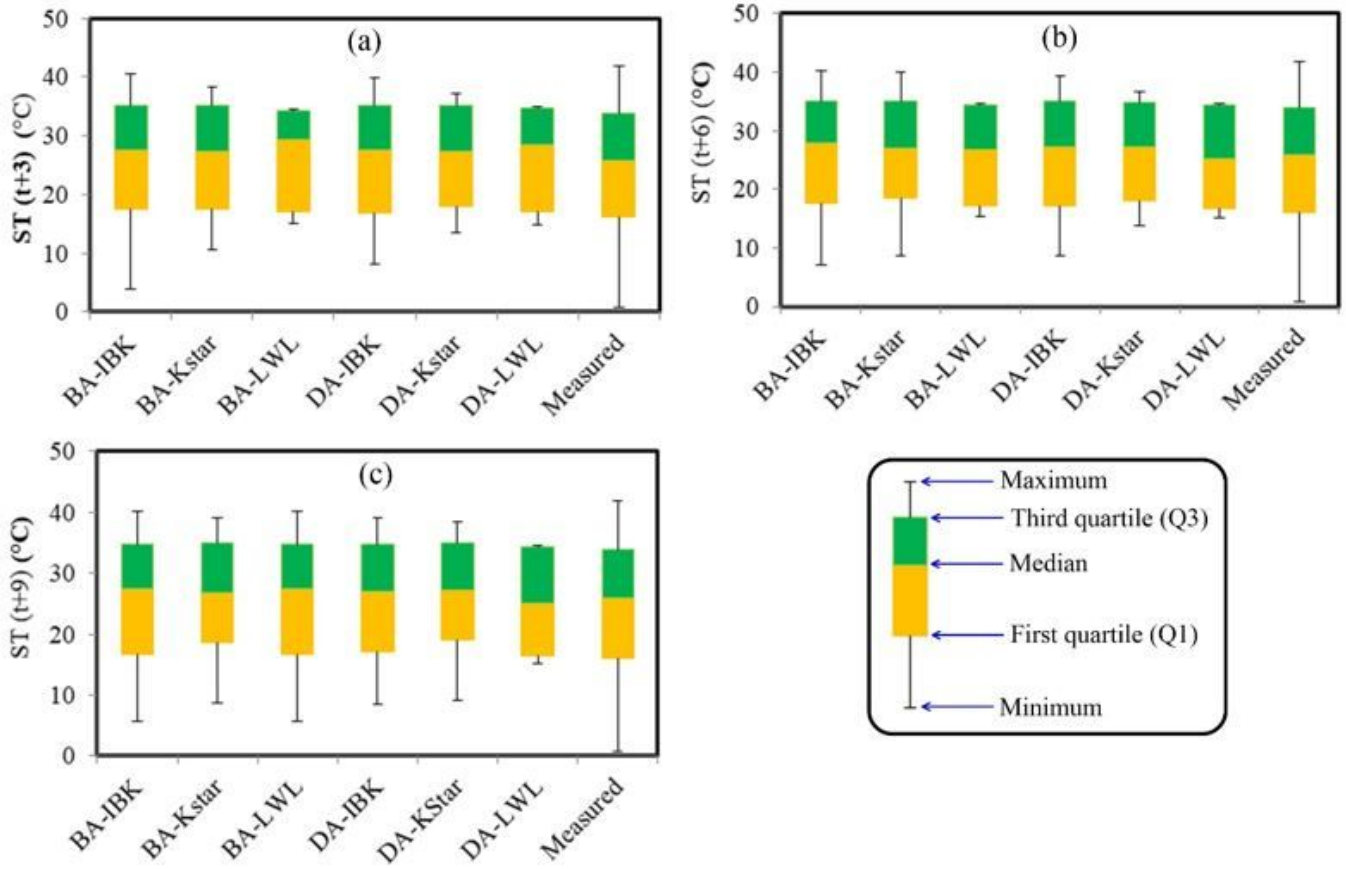

Figure 4

Model performance using box-plot for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 5 cm

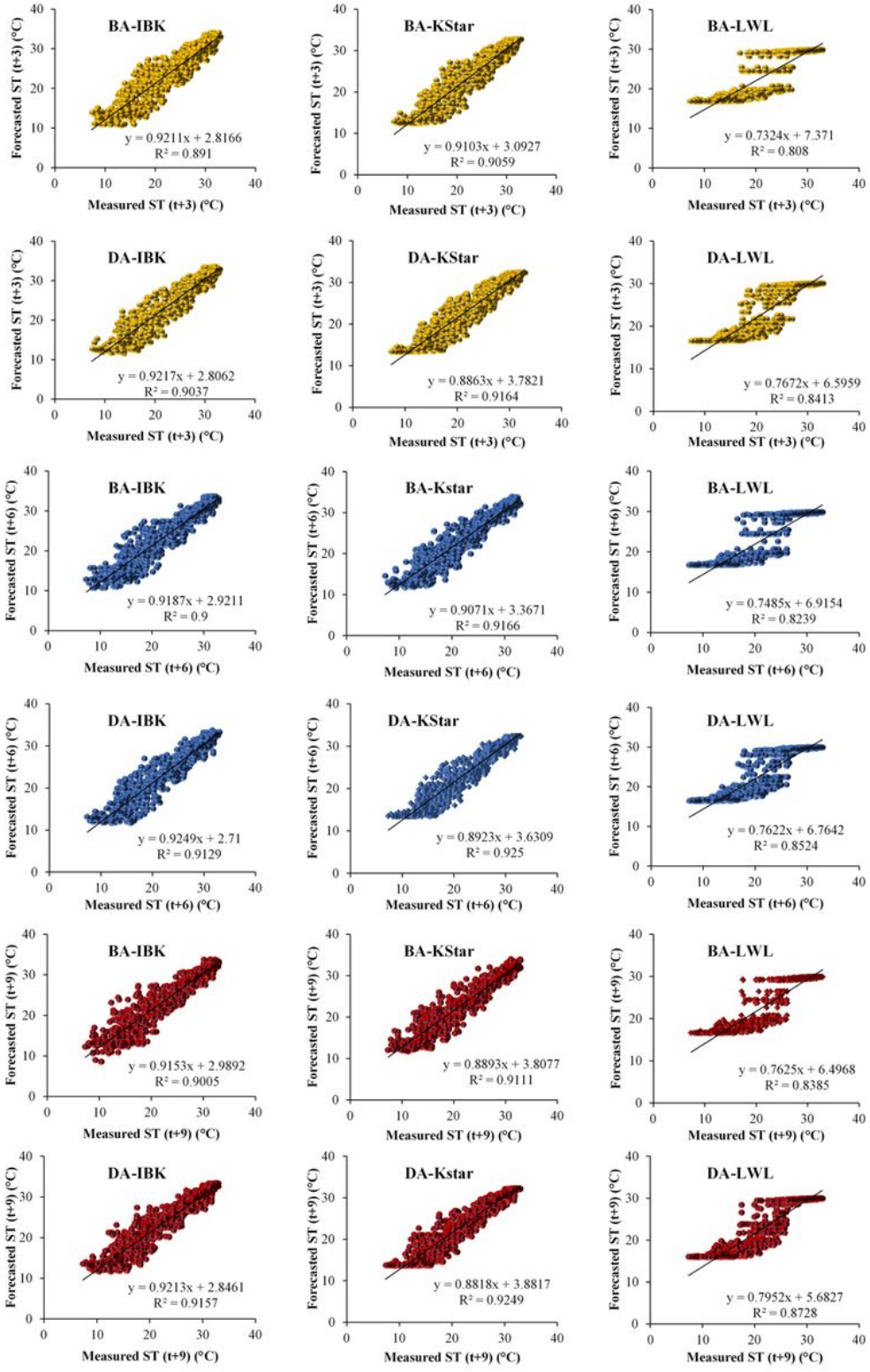

Figure 5

Scatter plot of 3 (i.e. golden color), 6 (i.e. blue color) and 9 (i.e. red color) days ahead forecasted *VS.* measured soil temperature at a depth of 50 cm

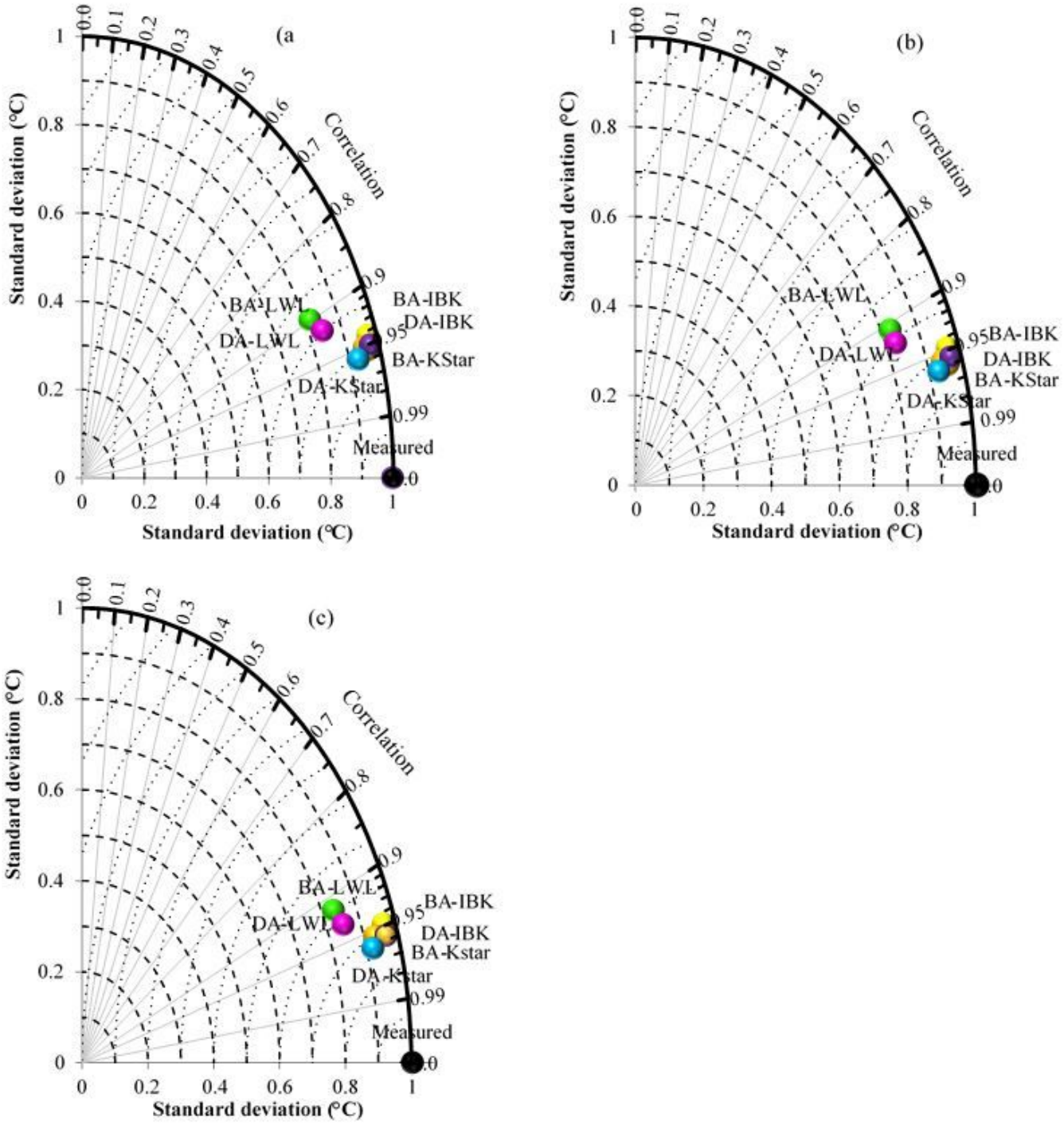

Figure 6

Model performance via Taylor diagram for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 50 cm

Figure 7

Model performance using box-plot for 3 (a), 6 (b) and 9 (c) days ahead ST forecasting at a depth of 50 cm