Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for Soil Temperature Prediction: Novel Model

LIU PENGHUI[1], **AHMED A. EWEES**[2], **BESTE HAMIYE BEYAZTAS**[3], **CHONGCHONG QI**[4], **SINAN Q. SALIH**[5,6], **NADHIR AL-ANSARI**[7], **SURAJ KUMAR BHAGAT**[8], **ZAHER MUNDHER YASEEN**[9], **AND VIJAY P. SINGH**[10,11]

[1]Computer Science Department, Baoji University of Arts and Sciences, Baoji 271000, China
[2]Computer Department, Damietta University, Damietta 34511, Egypt
[3]Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey
[4]School of Resources and Safety Engineering, Central South University, Changsha 410083, China
[5]Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
[6]Computer Science Department, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq
[7]Civil, Environmental, and Natural Resources Engineering, Luleå University of Technology, 97187 Luleå, Sweden
[8]Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[9]Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[10]Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843-2117, USA
[11]Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-2117, USA

Corresponding author: Zaher Mundher Yaseen (yaseen@tdtu.edu.vn)

ABSTRACT An enhanced hybrid artificial intelligence model was developed for soil temperature (ST) prediction. Among several soil characteristics, soil temperature is one of the essential elements impacting the biological, physical and chemical processes of the terrestrial ecosystem. Reliable ST prediction is significant for multiple geo-science and agricultural applications. The proposed model is a hybridization of adaptive neuro-fuzzy inference system with optimization methods using mutation Salp Swarm Algorithm and Grasshopper Optimization Algorithm (ANFIS-mSG). Daily weather and soil temperature data for nine years (1 of January 2010 - 31 of December 2018) from five meteorological stations (i.e., Baker, Beach, Cando, Crary and Fingal) in North Dakota, USA, were used for modeling. For validation, the proposed ANFIS-mSG model was compared with seven models, including classical ANFIS, hybridized ANFIS model with grasshopper optimization algorithm (ANFIS-GOA), salp swarm algorithm (ANFIS-SSA), grey wolf optimizer (ANFIS-GWO), particle swarm optimization (ANFIS-PSO), genetic algorithm (ANFIS-GA), and Dragonfly Algorithm (ANFIS-DA). The ST prediction was conducted based on maximum, mean and minimum air temperature (AT). The modeling results evidenced the capability of optimization algorithms for building ANFIS models for simulating soil temperature. Based on the statistical evaluation; for instance, the root mean square error (RMSE) was reduced by 73%, 74.4%, 71.2%, 76.7% and 80.7% for Baker, Beach, Cando, Crary and Fingal meteorological stations, respectively, throughout the testing phase when ANFIS-mSG was used over the standalone ANFIS models. In conclusion, the ANFIS-mSG model was demonstrated as an effective and simple hybrid artificial intelligence model for predicting soil temperature based on univariate air temperature scenario.

INDEX TERMS Air temperature; soil temperature; hybrid intelligence model, metaheuristic, North Dakota region.

I. INTRODUCTION

Soil temperature (ST) controls most of the biological, chemical, and physical processes within the soil [1], such as

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno Garcia.

evapotranspiration, plant germination/growth, root development, and microbial activities is evident [2]. Also, ST is the most important variable that influences the freezing depth and affects agricultural activities, irrigation scheduling, and soil drainage [3]. Several factors, including topography, air temperature, solar radiation, precipitation, soil moisture, and

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE *Access*

thermal characteristics, govern ST [4], [5]. Attempts have focused on finding the relation between soil temperature and other variables for prediction of ST [6].

The relationship between AT and ST is significant to determine the soil's heat regime. In some regions, ST is not measured and therefore the relationship between AT and ST can be used to construct a predictive model. The prediction of ST for any future time horizon without the need for direct measurement that is labour intensive and logistically expensive has a significant advantage of soft computing models that can help predict soil temperature for soil health and agriculture. Based on the predictive performance of the proposed hybrid intelligence model and for the univariate air temperature scenario, decision-makers can rely on data inexpensive approaches to better understand soil temperature dynamics to guide their practices.

About 54% of the total energy from the sun is reflected by the earth while retaining the remaining 46% [7]. The reason for this phenomenon is the slow heat transport capability of the ground; it is also due to the high heat storage capacity and slow changes in temperature over long periods with respect to the measurement depth. This soil's low thermal conductivity allows the release of heat through the cooling period and heat retention during the heating period. The heat retained in the soil during the hot season is released back to the air during the cold season and this heat exchange occurs throughout the year, making the soil a good thermal reservoir [8]. The ST during the winter season is higher than the temperature of the ambient air but reverses during the summer. This implies that ST can serve as an important meteorological parameter for different applications, including solar energy applications, frost prediction, agricultural applications, and ground source heat pump applications [9].

Several studies have been carried out recently on short to mid-term ST predictions in two specific categories [6], [10]. The first category focused on the use of statistical techniques, such as numerical weather prediction methods, which assume that changes in the statistical properties of ST data series in the future would be similar to those in the past [11]. These models for long-term prediction require large amounts of data which usually is not be available [12]–[14]. The second category uses artificial intelligence (AI) models, like support vector machine (SVM), artificial neural network (ANN), genetic programming (GP), gene expression programming (GEP), adaptive neuro-fuzzy inference system (ANFIS), decision tree (DT), M5 Tree, etc [10], [15]–[22]. Many studies have modeled ST as a nonlinear physical phenomenon [19], [20], [23]–[26].

One of the earliest studies used linear regression (LR) and ANN models for soil temperature prediction using several hydrometeorological variables including AT, atmospheric pressure, solar radiation, wind speed, relative humidity, and sunshine at Adana, Turkey [7]. Results of the ANN model showed better predictive performance. Xing *et al.* (2018) predicted daily ST for different seasons in different climatic zones at United States using the SVM model with AT and solar radiation as predictors. The results evidenced the capacity of the SVM model.

Samadianfard *et al.* (2018) integrated wavelet transformation with ANN and GEP models for ST prediction at different depths in Tabriz, Iran. The predictors were sunshine, radiation, and AT. Sanikhani *et al.* (2018) used the extreme learning machine (ELM) model as non-tuned predictive model for simulating ST at different soil depths at Mersin and Adana, Turkey. The ELM model was validated against ANN and M5 Tree models. Different climate variables, including AT, solar radiation, relative humidity, and wind speed, were used as predictors.

Soil temperature is a stochastic variable of great importance in various morphological and engineering areas [29]. Different AI models have been developed irrespective of internal parameters tuning [15]. Selection of the best model parameters using an optimization tool will ensure better model prediction [30]. Hybridized AI models have become one of the most successful nonlinear time series prediction methods in the field of signals and time-series analysis [31], [32]. Many studies have applied hybrid AI models but are limited on the measurement of ST at both surface and underground levels. The major aim of this study is to develop a univariate machine learning (ML) model for ST prediction using correlated independent variables.

Among several ML models, ANFIS model have demonstrated an excellent predictive model for diverse engineering applications [33]–[39]. Based on the literature, the ANFIS model has been found to be a reliable intelligent model for simulating soil temperature, owing to its ability to account for the uncertainty of data [40]. However, this model is associated with the problem of internal membership function parameter optimization [41]. Hence, the ANFIS model is hybridized with new integrative bio-inspired optimization algorithm called Grasshopper Optimization Algorithm (GOA) and Salp Swarm Algorithm (SSA). Several standard hybrid models are developed for the proposed model ANFIS-mSG model. In this context, the SSA algorithm showed promising results in solving a variety of optimization problems [42]–[45], because it has many advantages such as low computational cost, easy to implement, and few parameters to be optimized. However, it has some limitations like other metaheuristic methods, including slow convergence and low exploitation ability as well as it can get trapped in local optima. Therefore, the proposed method helps the SSA by adding a mutation strategy from getting trap in a local minimum, evolution stagnation, and premature convergence. The mutation strategy can assist metaheuristic algorithms to guide their populations toward the global optimum rather than getting trapped in local optima as well as increase the diversity of the search domain to speed up the convergence rate. Hence, the first phase of the proposed method is to improve the original SSA, then use this phase as a local search for the GOA algorithm.

However, GOA, a recent metaheuristic algorithm, has drawbacks, including high computation time and premature convergence [46], [47]. Therefore, the second phase of the

IEEE *Access*

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

proposed method is to improve the exploration and exploitation of the original GOA using the first phase (i.e. mutation SSA phase). These two phases are mentioned in this manuscript as mutation-SSA-GOA (mSG).

The final phase of the proposed method is to apply the mSG to train the ANFIS model in order to improve the predictiveness of the original ANFIS. The entire proposed method is called ANFIS-mSG. Thus, this study is the hybridization of the ANFIS model with a nature optimization algorithm (i.e., ANFIS-mSG) to develop a hybrid ST prediction model.

The literature shows that hybrid models are receiving attention as evidenced by the number of studies integrating nature optimization algorithms with standalone ML models for solving hyperparameter problems. Until now, limited attention has been given to the use of metaheuristic optimization algorithms with ANFIS model for ST modeling. Therefore, this study aims to implement a suite of standalone and hybrid ANFIS models for ST modeling at different meteorological stations located in North Dakota, United States. This study used a univariate modeling scheme as only AT variability was used to model the ST. Indeed, establishing such a kind modeling process based on univariate weather information is highly essential for the regions where climatological or geoscience information availability are limited. In addition, establishing such intelligent predictive with less input variables to predict soil temperature is significantly important for multiple geoscience related implications that contributes to basic knowledge of various perspectives of soil engineering. The proposed ANFIS-mSG was validated against ANFIS, ANFIS hybridized with grasshopper optimization algorithm (ANFIS-GOA), salp swarm algorithm (ANFIS-SSA), grey wolf optimizer (ANFIS-GWO), particle swarm optimization (ANFIS-PSO), genetic algorithm (ANFIS-GA), and Dragonfly Algorithm (ANFIS-DA). The main contribution of the current research is the newly developed hybrid ANFIS model for ST prediction which is important for soil health, agriculture, and the ecosystem.

II. CASE STUDY AND DATA EXPLANATION

North Dakota (ND), located in the center of North America, has a typical climate with cold winters and hot summers. Its climate is characterized by large temperature variations, which cause different weather conditions for each of the four seasons. The eastern and western parts of ND have different climate conditions. According to the Köppen–Geiger climate classification system, the climate of the eastern part of ND has a humid continental climate, while the western part has a semi-arid climate [48]. As noted by U.S. Global Change Research Program (2000), the average temperature in ND has increased about 3 °C. Figure 1 presents the locations of the stations. Figure 2 illustrates variations of soil temperature over 2010-2018 for all five working stations (Baker, Beach, Cando, Crary, and Fingal) at 10 cm soil depth along statistical analysis of all metrological station (see Table 1). For Crary station, the variations of soil temperature throughout the consecutive nine years were persistent as well as extreme than at

TABLE 1. Statistical analysis of ST dataset at meteorological station.

	Minimum	Maximum	Average	Skewness	SD
Baker	31.0	82379.0	47433.2	0.32	17770.8
Beach	53.0	82713.0	48494.0	0.29	17117.5
Cando	29.0	83643.0	46328.0	0.30	17006.5
Crary	8513.0	86404.0	46465.1	0.20	19870.3
Fingal	30.0	87334.0	48349.8	0.20	19019.0

other stations. The ST for most of the years appeared stable for both Baker and Fingal stations. In contrast, Cando station had the most variability in ST among the nine consecutive years. The lowest ST was in 2014 and slowly inched upward till 2018, and 2010-2012 showed equivalent as well as peak ST among all years. For Beach station, the ST variations appeared broader and sharpener for each of the years equally mostly except in 2012 and 2015 where it was the lowest throughout the period.

In 2012, ST was the least for three meteorological stations (Beach, Cando and Crary), where for Baker and Fingal the lease ST was in 2017 and 2015, respectively. A couple of lower values of ST occurred throughout all nine years for Baker, Beach and Cando, where three times of lower values of ST occurred for Fingal and in case of Crary it was more than three times.

III. METHODOLOGICAL OVERVIEW
A. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS)

ANFIS, developed by Jang [49], is a combination of fuzzy logic and neural networks, drawing on their advantages. ANFIS applies a Takagi–Sugeno inference method that generates nonlinear mapping, by the fuzzy IF–THEN rules, from input to output domains. It uses five layers to address its tasks. Figure 3a summarizes these layers and the following steps explain the working sequence of these layers.

Layer 1 passes the inputs x and y to the nodes to calculate the output of this layer using the generalized Gaussian membership μ as in the following equations [50]:

$$O_{1i} = \mu_{A_i}(x), \quad i = 1, 2, \tag{1}$$
$$O_{1i} = \mu_{B_{i-2}}(y), \quad i = 3, 4$$
$$\mu(x) = e^{-(x - \rho_i / \sigma_i)^2} \tag{2}$$

where B_i and A_i denote the values of the membership μ; and σ_i and ρ_i denote the parameters set.

Layer 2 applies Eq. 3 to calculate the output of each node (the firing strength of a rule). After that, the results are normalized in Layer 3 by Eq. 4:

$$O_{2i} = \mu_{A_i}(x) \times \mu_{B_{i-2}}(y) \tag{3}$$
$$O_{3i} = \bar{w}_i = \frac{\omega_i}{\sum_{i=1}^{2} \omega_i}, \tag{4}$$

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction
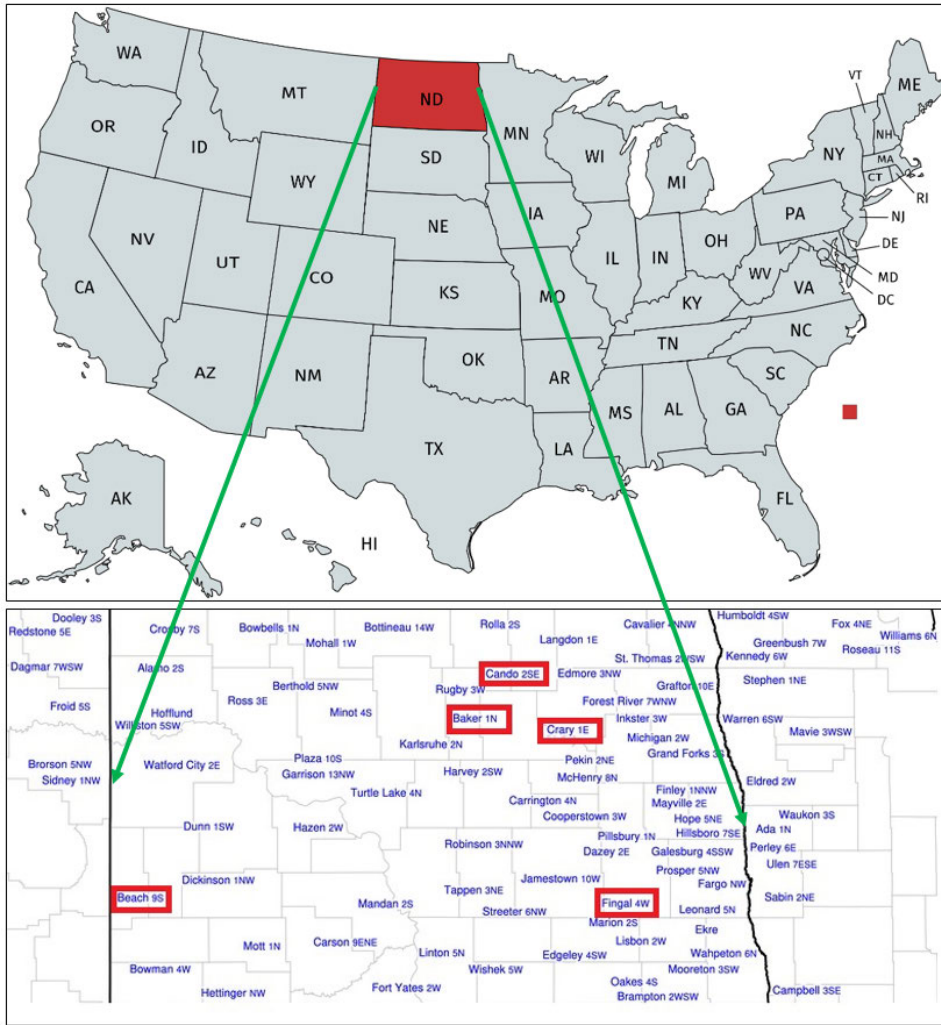
IEEE *Access*

FIGURE 1. Locations of meteorological stations in North Dakota.

In Layer 4, the adaptive nodes are computed by Eq. 5:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \tag{5}$$

where r, q, and p denote the parameters of the i-th node.

Layer 5 computes the output results using the following equation:

$$O_5 = \sum_i \bar{w}_i f_i \tag{6}$$

In several cases the search domain of the ANFIS model becomes wider but the convergence becomes slower, therefore it can get trapped in local optima [51]. Consequently, training the weights of the ANFIS model is a valuable step to overcome such a problem.

B. SALP SWARM ALGORITHM (SSA)

Salp swarm algorithm (SSA), proposed by [42], is an optimization technique mimicking the behaviour of the salp chains in nature. This behaviour is considered as a swarm behaviour; the target of this swarm is a food source. Salps use this behaviour to forage and move with fast harmonious diversity [52].

The SSA is formed mathematical to be used in the computational process. The initial population of SSA, after generation, is divided into two groups. The front one is called salp leader, whereas the reset is called salp follower. The search space of a given problem is represented in n-dimensions where n is the number variables. The salp leader positions are frequently updated using the following equation:

$$x_j^i = \begin{cases} F_j + c_1((ub_j - lb_j) \times c_2 + lb_j), & c_3 \leq 0 \\ F_j - c_1((ub_j - lb_j) \times c_2 + lb_j), & c_3 > 0 \end{cases} \tag{7}$$

where x_j^1, ub_j, and lb_j denote the position, upper and lower bounds in j-th dimension, respectively; F_j defines the food source; c_2 and c_3 are generated randomly in the range $[0, 1]$; and coefficient c_1 balances the exploration and the exploitation stages, whereas the following equation is used

IEEE *Access*

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction
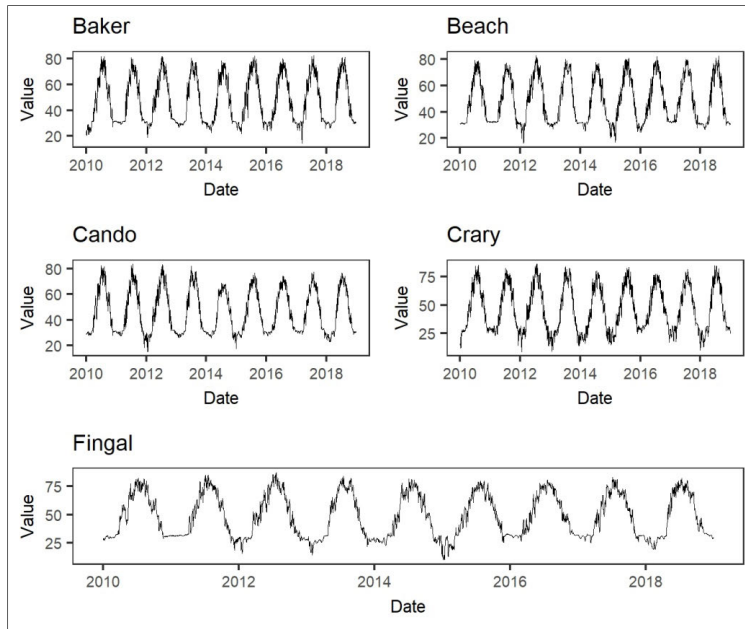

FIGURE 2. Trend of actual bare soil temperature at the five studied meteorological stations at 10 cm depth.

to compute its value:

$$c_1 = 2e^{-(\frac{4p}{P})^2} \tag{8}$$

where P and p denote the maximum number of loops and the current loop, respectively.

The followers' positions are also updated by the following equation:

$$x_j^i = \frac{1}{2}(x_j^i + x_j^{i-1}) \tag{9}$$

where $i > 1$ and x_j^i denotes the i-th follower position.

C. GRASSHOPPER OPTIMIZATION ALGORITHM (GOA)

GOA is an optimization technique and works as the grasshopper insects in nature (Saremi *et al.* 2017). The grasshopper is a kind of pest effects on agriculture and crop production. Its life cycle includes three stages: egg, nymph, and adulthood [54]. In the nymph stage, it moves slowly in rolling cylinders and jumps by small steps and eats vegetation in its path. In the adulthood stage, it uses swarm behaviour to migrate a long distance with long- range and abrupt movement.

This behaviour can be mathematically expressed by the following equations [47]:

$$x_i = S_i + G_i + A_i, \quad i = 1, 2, \dots, N \tag{10}$$

where x_i is the position of the i-th grasshopper, and S_i is its social interaction that can be defined as follows:

$$S_i = \sum_{\substack{j=1 \\ i \neq j}}^{N} s(d_{ij})\hat{d}_{ij}, \quad d_{ij} = |x_i - x_j|, \quad \hat{d}_{ij} = \frac{x_i - x_j}{d_{ij}} \tag{11}$$

where d_{ij} denotes the space between the j-th and i-th grasshoppers. $\hat{d}_{ij}$ indicates a unit vector from the i-th to the j-th grasshoppers, and s denotes the social forces' strength that can be defined as:

$$s(y) = fe^{\frac{-y}{l}} - e^{-y} \tag{12}$$

where l and f denote the attractive length scale and the attraction's intensity, respectively.

The wind direction is a main factor in the movement of nymph grasshoppers because they have no wings.

In Eq. (13), A_i and G_i denote the wind advection and the gravity force for the i-th grasshopper, respectively:

$$A_i = u\hat{e}_w, \quad G_i = -g\hat{e}_g \tag{13}$$

where u and g denote a constant drift and the gravitational constant, respectively; and $\hat{e}_w$ and $\hat{e}_g$ denote the unity vectors of the direction of wind and towards earth's center, respectively.

Taking these factors into account Eq. (13) is rewritten in Eq. (14) to be more suitable for searching for the solution of a given problem.

$$x_i^d = c\left(\sum_{\substack{j=1 \\ i \neq j}}^{N} c\frac{u_d - l_d}{2}s(|x_j^d - x_i^d|)\frac{x_j - x_i}{d_{ij}} \right) + \hat{T}_d, \tag{14}$$

where u and l denote the upper and the lower bounds of the search domain, respectively; $\hat{T}_d$ denotes the current best solution; N is the population number; D denotes the dimension of the problem; and c denotes a decreasing coefficient to provide

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction
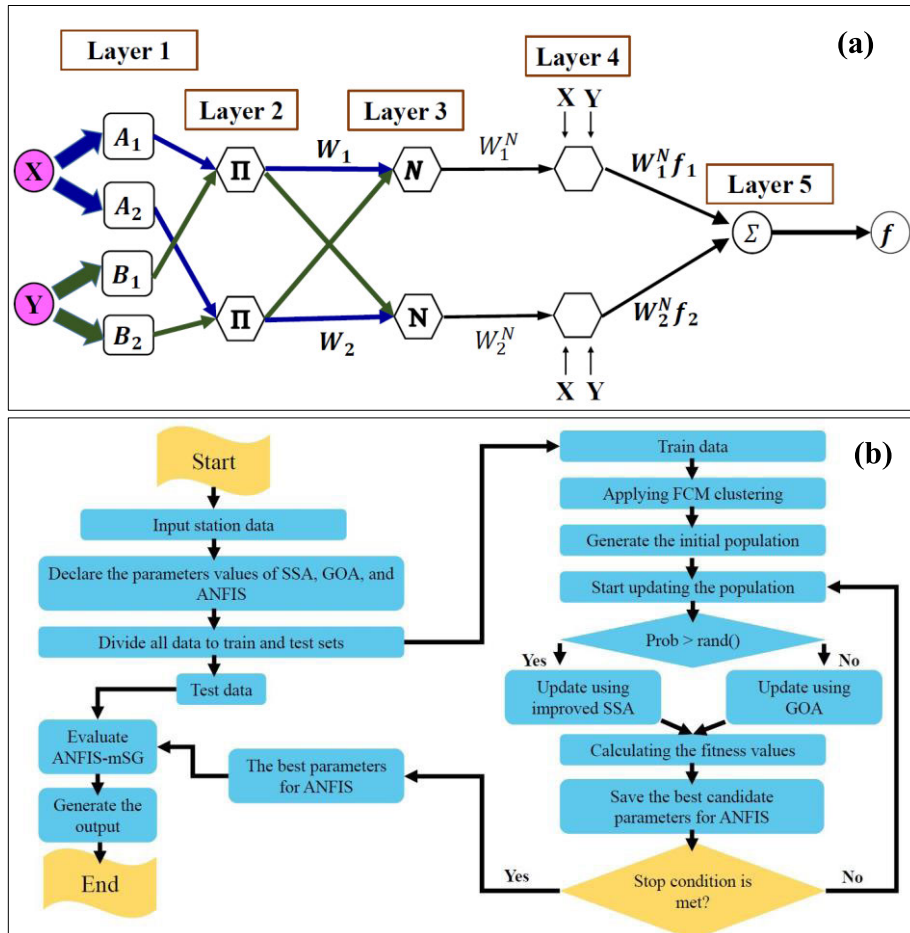
IEEE*Access*

FIGURE 3. a) Basic ANFIS model structure, b) Flowchart of the proposed method.

good balancing between exploration and exploitation phases and can be calculated by the following equation:

$$c = c_{max} - t\frac{c_{max} - c_{min}}{t_{max}} \tag{15}$$

where c_{max} and c_{min} indicate the maximum and the minimum values, respectively, namely 1 for c_{max} and 0.0001 for c_{min}; t denotes to the number of the current iteration; and t_{max} is the maximum number of iterations.

D. PROPOSED METHOD (ANFIS-mSG)
The proposed method contains two improvement stages. The first stage is to improve the basic SSA using mutation phase in order to enhance its exploration phase. After that, the improved SSA algorithm is used as a local search for GOA algorithm; this stage produces an mSG algorithm. The second stage uses the mSG algorithm to improve and train the basic ANFIS model.

These stages are explained in the following subsections.

1) FIRST STAGE (mSG)
In this stage the basic SSA is improved by adding a mutation phase in its structure. This phase generates a mutation vector

x_{mu} as in the following equation:

$$x_{mu,i} = x_{p1} + \delta \times (x_{p2} - x_{p3}) \tag{16}$$

where, $i = 1, 2, 3, \ldots, N$; whereas x_q, x_w, and x_r are random populations; and δ is a constant $\in [0, 2]$ and works to control the differential variation.

Then this vector ($x_{mu,i}$) is evaluated by the fitness function; if its fitness value is better than the fitness value of the original vector (x_i), then it will be used, else the old fitness value will be retained. Consequently, the improved SSA is used as a local search for the basic GOA to help in exploring more domains in the search space.

2) SECOND STAGE (ANFIS-mSG)
In this stage the mSG algorithm is used to adjust the ANFIS's parameters by passing the optimal weights between layers 4 and 5. The ANFIS-mS starts by receiving the predictors and splits them into train and test sets as well as setting up all parameter values. The membership function is determined by using the fuzzy c-mean method. The next step is to apply the mSG algorithm to optimize and adapt the ANFIS's weight values, where the mSG searches for the best weights that can give the best solution by exploring various domains. In this

stage, the mSG starts searching for the best parameters for the ANFIS. These parameters are fed to train the ANFIS, then the fitness values are evaluated by Eq. (17) to check if the candidate parameters are good and better than the old ones or the algorithm should search for another one.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (a_i - p_i)^2 \qquad (17)$$

where a denotes the actual values; p denotes the output values; and n is the size of inputs. MSE is the well-established metric used as determination for the fitness matrix, as reported over the literature [55], [56]. MSE helps to know how close the predicted results to regression line, since the smaller value of MSE indicates the good results. It can effectively work with large error whereas, it gives a relatively high weight to them. MSE is a simple and common metric for measuring the performance of the regression models besides, it does not affect the computational cost.

Therefore, the best parameters will be considered according to the smallest error between actual and predicted values. The ANFIS-mS works until reaching the maximum number of iterations which will be considered as a stop criterion. Then, the selected parameters are fed to the ANFIS to start preparing the predicted results in the test phase. The entire process of the proposed ANFIS-mSG model is illustrated in Figure 3b. It is worth to report the magnitudes of the internal tuning parameters of each model and those values reported in Table 2.

E. MODEL DEVELOPMENT
This study aims to investigate the ability of the ANFIS-mSG algorithm for predicting the soil temperature. For this reason,

TABLE 2. The magnitudes of the tuning parameters of the developed models.

Models	Parameter settings
ANFIS	Error_Goal = 0, initial step = 0.01, Max epochs = 100, Error goal = 0, Initial step = 0.01, Decrease rate = 0.9
ANFIS-GOA	$c_{max} = 1$, $c_{min} = 0.00004$
ANFIS-SSA	C_2 in the range [0, 1], C_3 in the range [0, 1]
ANFIS-GA	pc=0.8, γ=0.2, pm=0.3, mu=0.02, β=8
ANFIS-PSO	w=1, wDamp = 0.99, C_1=1, C_2=2
ANFIS-GWO	Convergence constant (a) in the range [2, 0]
ANFIS-DA	β in the range [0.9, 0.4], e in the range [0.1, -0.1]
ANFIS-mSG	Same as ANFIS, GOA, and SSA

the experimental dataset is divided into two parts; the first part is used for training the proposed method, whereas the second part is used for the testing phase.

In the training phase, the ANFIS-mSG begins by generating a random population x; each $x_i, (i = 1, 2, 3, \ldots, N)$ represents a solution.

This population is updated by the improved SSA and GOA using a probability (pb) to switch between them as in Equation 18:

$$pb_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \qquad (18)$$

where f_i is the current fitness value. If $pb_i > rand()$, the improved SSA will update the solution, else the GOA will be used.

The reason for using this probability is to overcome the limitation of the original GOA, such as high computation time, premature convergence, and getting trapped in a local minimum. After this step, each solution is evaluated using the fitness function and consequently, the prediction error is calculated using Equation (17). The result of this equation is used to check if the current solution is better than the best solution or not. If true, the current solution will be saved for comparison in the next iteration. This sequence is repeated till meeting the stop condition which is the maximum number of iterations. Then, the best candidate parameter is passed to update the ANFIS model to start the testing phase.

In the testing phase, the ANFIS-mSG method receives the test data to evaluate the candidate parameter of ANFIS. The output of this phase is evaluated using seven performance measures as shown below.

F. PERFORMANCE METRICS
Seven performance measures are presented which were used to evaluate the proposed predictive models [57].

➤ Root mean square error (RMSE): It is computed by the following equation:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (a_i - p_i)^2} \qquad (19)$$

where a denotes the output values, p denotes the real values, and n is the total number of items.

➤ Standard deviation (STD): It is computed by the following equation:

$$STD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (RMSE_i - \mu)^2} \qquad (20)$$

where μ is the mean value of the RMSE.

➤ Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |a_i - p_i| \qquad (21)$$

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE *Access*

➢ Root Mean Squared Relative Error (RMSRE)

$$RMSRE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{a_i - p_i}{a_i}\right|\right)^2} \tag{22}$$

➢ Average Absolute Percent Relative Error (AAPRE)

$$AAPRE = \frac{100}{n}\sum_{i=1}^{n}\left(\left|\frac{a_i - p_i}{p_i}\right|\right) \tag{23}$$

➢ Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum (a_i - p_i)^2}{\sum (a_i - \mu_a)^2} \tag{24}$$

➢ Nash–Sutcliffe efficiency index (NES):

$$NES = 1 - \frac{\sum (a_i - p_i)^2}{\sum (a_i - \mu_a)^2} \tag{25}$$

where, μ_a is the mean value of the a values.

IV. RESULTS AND ANALYSIS

The main motivation of the current research was to investigate the feasibility of new hybrid intelligent models based on the integration the viability of nature inspired algorithms for optimizing the internal parameters of the ANFIS model and modeling soil temperature at five different meteorological station located in North Dakota, USA. The models were trained using a univariate modeling procedure based on AT data, including maximum, mean and minimum values. The proposed hybrid predictive model and the competing models were assessed and evaluated using statistical performance indicators.

Figure 4 reveals two statistical performance metrics, including MAE and RMSE, for the applied predictive models (i.e., ANFIS-mSG, ANFIS, ANFIS-DA, ANFIS-GA, ANFIS-GO, ANFIS-GWO, ANFIS-PSO and ANFIS-SSA) for all five meteorological stations (Baker, Beach, Cando, Crary and Fingal). With respect to MAE and RMSE statistics, the ANFIS-mSG model had the lowest magnitudes, showing the best performance for all stations except Fingal station where both proposed ANFIS-mSG and ANFIS-GA models were at the top with 81% and 82% enhancement in the case of RMSE and MAE, respectively, in comparison with the ANFIS. The highest performance enhancement was observed using the proposed ANFIS-mSG by (82% of RMSE and 79% of MAE) at Cando station in comparison with the least performed model (i.e., ANFIS-GOA). The lowest MAE and RMSE values of ANFIS-mSG model appeared for Crary, almost equivalent for both Baker and Fingal stations, while slightly higher and simultaneously highest values of the same matrices for Beach and Cando stations. In contrast, the highest MAE and RMSE values appeared for ANFIS model alone for all stations except Cando where ANFIS-GOA ranked the highest for the same performance matrices (See Table 3-7). This can be explained by the most unstable soil and air temperatures throughout 2010-2018, among
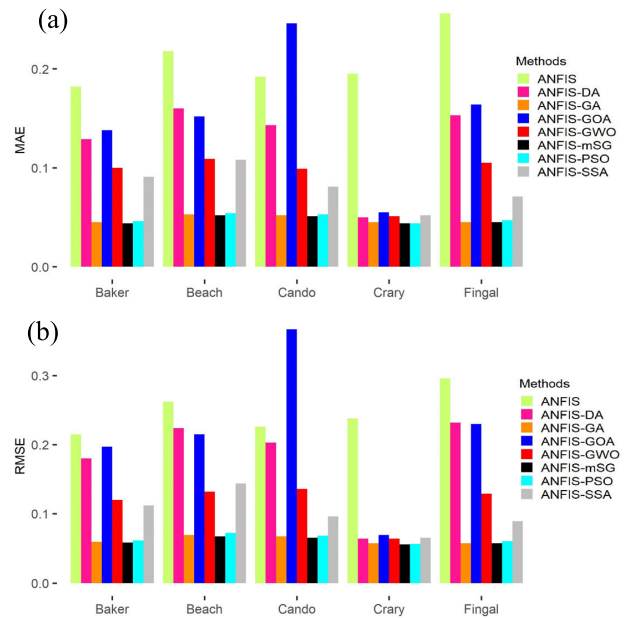

FIGURE 4. The magnitudes of (a) MAE and (b) RMSE of predictive models in the testing phase for all meteorological stations.

all stations (Fig. 2), also this model supported the unstable variability. The second-best performing model was the ANFIS-GA model of all stations except Crary station where ANFIS-PSO scored higher than ANFIS-GA as per the same values of performance matrices. This can be clarified due to the capacity of the genetic algorithm and the particle swarm optimization as a robust nature inspired optimization algorithms for tuning the internal parameters of AI models [50], [58]. In addition, this could be due to the extreme ST traced within a year among all meteorological stations and AT showed more observations of low temperature throughout the years.

The second least performing predictive model was ANFIS-GOA, followed by ANFIS-DA model as per the magnitude of MAE and RMSE metrics. The third best-performing model came up with higher values of MAE and RMSE (i.e., ANFIS-SSA) which can be a promising competitor with ANFIS-GWO model, except for Beach station where the RMSE value was higher for the ANFIS-SSA and nearly similar with the MAE value. The variation of the results was apparently owing to different geographical locations of Beach among all meteorological stations and may cause the variety in various metrological parameters (Fig. 1); in the case of AT it exhibited the low variability with a year.

Figure 5a shows the scatter plots of actual and predicted values of training and testing phases of all models for Baker station. The ANFIS-mSG and ANFIS-GA performances were at the top ($R^2 = 0.977$) for Baker station and as well as for all other stations where those two predictive models were followed by ANFIS-PSO models. On the contrary, ANFIS-GOA performed with least statistical matrices score and was followed by ANFIS-DA and ANFIS models, whereas

IEEE Access

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

TABLE 3. Statistical performance of predictive models in the testing phase at Baker meteorological station.

Models	RMSE	MAE	MRE	NES	RMSRE	AAPRE	R^2	STD	Time, sec
ANFIS-mSG	0.058	0.044	0.289	**0.955**	0.483	13.435	0.977	0.0014	29.528
ANFIS-GOA	0.197	0.138	-10.067	0.818	2.511	46.651	0.734	0.0602	39.243
ANFIS-SSA	0.112	0.091	-4.134	0.818	0.487	25.901	0.923	0.0165	7.278
ANFIS-GWO	0.120	0.100	-3.804	0.781	0.495	26.505	0.913	0.0255	7.965
ANFIS-PSO	0.061	0.046	0.687	0.948	0.595	14.149	0.975	0.0008	6.544
ANFIS-GA	0.059	0.045	0.326	0.953	0.552	13.819	0.977	0.0013	7.240
ANFIS-DA	0.180	0.129	-10.687	0.530	0.925	38.297	0.795	0.0965	11.667
ANFIS	0.215	0.182	-32.661	0.292	2.063	52.013	0.819	0.1123	2.978

TABLE 4. Statistical performance of predictive models in the testing phase at Beach meteorological station.

Models	RMSE	MAE	MRE	NES	RMSRE	AAPRE	R^2	STD	Time, sec
ANFIS-mSG	0.067	0.052	2.261	**0.939**	0.034	12.407	0.967	0.0022	29.912
ANFIS-GOA	0.215	0.152	3.383	0.302	0.703	42.883	0.749	0.0614	39.694
ANFIS-SSA	0.144	0.108	-2.643	0.724	0.280	30.246	0.834	0.0299	7.247
ANFIS-GWO	0.132	0.109	-4.194	0.746	0.175	28.667	0.873	0.0327	8.188
ANFIS-PSO	0.072	0.054	2.825	0.923	0.045	13.288	0.963	0.0025	6.581
ANFIS-GA	0.069	0.053	1.952	0.927	0.036	12.813	0.966	0.0030	7.384
ANFIS-DA	0.224	0.160	0.581	0.196	0.729	45.171	0.722	0.0824	11.374
ANFIS	0.262	0.218	-22.335	0.006	0.402	42.731	0.771	0.2094	2.88

TABLE 5. Statistical performance of predictive models in the testing phase at Cando meteorological station.

Models	RMSE	MAE	MRE	NES	RMSRE	AAPRE	R^2	STD	Time, sec
ANFIS-mSG	0.067	0.052	2.261	**0.931**	0.034	12.407	0.967	0.0022	29.912
ANFIS-GOA	0.215	0.152	3.383	-1.595	0.703	42.883	0.749	0.0614	39.694
ANFIS-SSA	0.144	0.108	-2.643	0.860	0.280	30.246	0.834	0.0299	7.247
ANFIS-GWO	0.132	0.109	-4.194	0.820	0.175	28.667	0.873	0.0327	8.188
ANFIS-PSO	0.072	0.054	2.825	0.918	0.045	13.288	0.963	0.0025	6.581
ANFIS-GA	0.069	0.053	1.952	0.919	0.036	12.813	0.966	0.0030	7.384
ANFIS-DA	0.224	0.160	0.581	0.334	0.729	45.171	0.722	0.0824	11.374
ANFIS	0.262	0.218	-22.335	-0.378	0.402	42.731	0.771	0.2094	2.88

ANFIS-SSA and ANFIS-GWO remained at the average level. It can be noted that, ANFIS-GOA consumed the highest convergence time among all models, followed by ANFIS-mSG. The least time was scored by ANFIS model alone, whereas the remaining models took on an average form.

Figure 5b shows the agreement between actual and predicted values of ST for the training and testing phases of the eight models at Beach station. The best performance was by ANFIS-mSG ($R^2 = 0.967$), followed by ANFIS-GA and ANFIS-PSO, despite having high variance in AT and a smaller number of the negative value of ST in each year. ANFIS-SSA and ANFIS-GWO models

had typical prediction performance. However, ANFIS-GOA and ANFIS models performed with lower predictability; yet, were better than ANFIS-DA whose performance was the least among all, with determination coefficient value $R^2 = 0.722$. The highest time consumption was observed by ANFIS-GOA (T = 39.694 sec), followed by ANFIS-mSG (T = 29.912 sec), while ANFIS-PSO (T = 6.581 sec) was he least time-consuming model.

Figure 5c illustrates the training and testing phases of the proposed and competing predictive models for Cando station. The highest agreement between observed and predicted values was by ANFIS-mSG ($R^2 = 0.967$) which was

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE*Access*

TABLE 6. Statistical performance of predictive models in the testing phase at Crary meteorological station.

Models	RMSE	MAE	MRE	NES	RMSRE	AAPRE	R^2	STD	Time, sec
ANFIS-mSG	0.0556	0.044	-0.881	**0.951**	0.070	13.415	0.976	0.0003	33.999
ANFIS-GOA	0.0690	0.055	0.865	0.934	0.082	16.121	0.963	0.0066	39.515
ANFIS-SSA	0.0651	0.052	1.147	0.935	0.070	15.073	0.968	0.0040	7.225
ANFIS-GWO	0.0638	0.051	0.322	0.935	0.070	14.853	0.968	0.0032	7.689
ANFIS-PSO	0.0560	0.044	-0.692	0.949	0.072	13.421	0.976	0.0008	6.590
ANFIS-GA	0.0571	0.045	-0.831	0.948	0.074	13.634	0.975	0.0023	7.519
ANFIS-DA	0.0638	0.050	1.308	0.936	0.079	14.917	0.969	0.0030	11.442
ANFIS	0.2378	0.195	-19.818	0.144	0.330	39.324	0.953	0.1428	2.952

TABLE 7. Statistical performance of predictive models in the testing phase at Fingal meteorological station.

	RMSE	MAE	MRE	NES	RMSRE	AAPRE	R^2	STD	Time, sec
ANFIS-mSG	0.057	0.045	1.210	**0.946**	0.027	11.274	0.973	0.0011	28.117
ANFIS-GOA	0.230	0.164	2.809	0.614	0.782	47.253	0.802	0.0565	37.433
ANFIS-SSA	0.089	0.071	0.021	0.856	0.083	19.020	0.928	0.0048	6.718
ANFIS-GWO	0.129	0.105	-3.577	0.736	0.181	26.562	0.826	0.0545	7.454
ANFIS-PSO	0.060	0.047	1.702	0.936	0.033	11.945	0.970	0.0017	6.128
ANFIS-GA	0.057	0.045	0.820	0.943	0.027	11.288	0.972	0.0013	6.914
ANFIS-DA	0.232	0.153	-5.289	0.141	0.924	44.319	0.787	0.0912	10.922
ANFIS	0.296	0.256	-42.010	-0.343	0.420	48.780	0.900	0.1732	2.824

almost similar to ANFIS-GA ($R^2 = 0.966$) and ANFIS-PSO ($R^2 = 0.965$). These models performed better even for negative values of AT that appeared more times and most of the value of AT and ST were near to the mean value. The distinguished least performance was by ANIFS-GOA model in respect of $R^2 = 0.537$ as well as T = 40.131 sec. The remaining models performed average (ANIFS-SSA, ANFIS, ANIFS-GWO and ANIFS-DA). In the case of time consumed by each of the models, the least time spent was by ANFIS model where ANIFS-GOA used up the highest, followed by ANFIS-mSG.

Figure 5d shows the scatter plots of training and testing phases of each model for Crary station. In this case, ANFIS-mSG performed the best ($R^2 = 0.976$), likewise for other stations. Another observation was that all models performed well with the range of $R^2 = 0.975$ (ANIFS-GA) to $R^2 = 0.953$ (ANFIS). The ST and AT were observed almost stable with its peak value as well throughout the years. The ANFIS model had spent the least time T = 2.952 sec. On the other hand, ANIFS-GOA expended for the largest time (T = 39.515 sec), followed by ANFIS-mSG (T = 33.999 sec). The remaining models did not have a significant difference, the average time range was 6.5-11.4 sec.

Figure 5e presents the scatter diagram of training and testing phases of the models to determine the relationship between the actual and predicted values for Fingal station. The ANFIS-mSG model showed the superior predictability. Likewise, at Fingal station, the least performing model

was ANIFS-DA. ANFIS had an average performance in accordance with ANIFS-SSA, ANIFS-GWO and ANIFS-GOA, unlike other stations. Interestingly the time difference among all models was noticeable. Highest times were noticed for the ANIFS-GOA model, likewise other station, followed by ANFIS-mSG. By contrast, the least time (2.824 sec) was used by ANFIS and the remaining models had the average range of time (6-10 sec) consumed.

Figure 6 presents the distribution of relative error of all models for the testing phase for all stations. The demonstration laid out the equivalent interquartile range (IQR) of ANFIS-mSG. In contrast, ANIFS-DA and ANFIS had unstable relative errors in the testing phase with noteworthy number of outliers. For instance, at Baker station, ANFIS has its median value tending towards the 1$^{\text{st}}$ quartile (25 percentile of relative error) with the highest peak value in the case of minimum as well as maximum, while in the case of ANIFS-DA, it was opposite i.e. the median value was towards the 3$^{\text{rd}}$ quartile (75 percentile of relative error). A distinguished position of the boxplot was unveiled by all models which were slightly on the upper side in comparison with other station.

The relative error at Beach station demonstrated no significant IQR difference between ANFIS-mSG and ANIFS-GOA. In contrast, ANIFS-DA, ANIFS-SSA, and ANIFS-GWO displayed a gap with an important number of outliers, along with the least IQR in the case of ANIFS-DA model. The ANFIS model stood with the extreme point of relative error along with the piece of IQR being higher than value +1 among all models, likewise in the case of Crary and
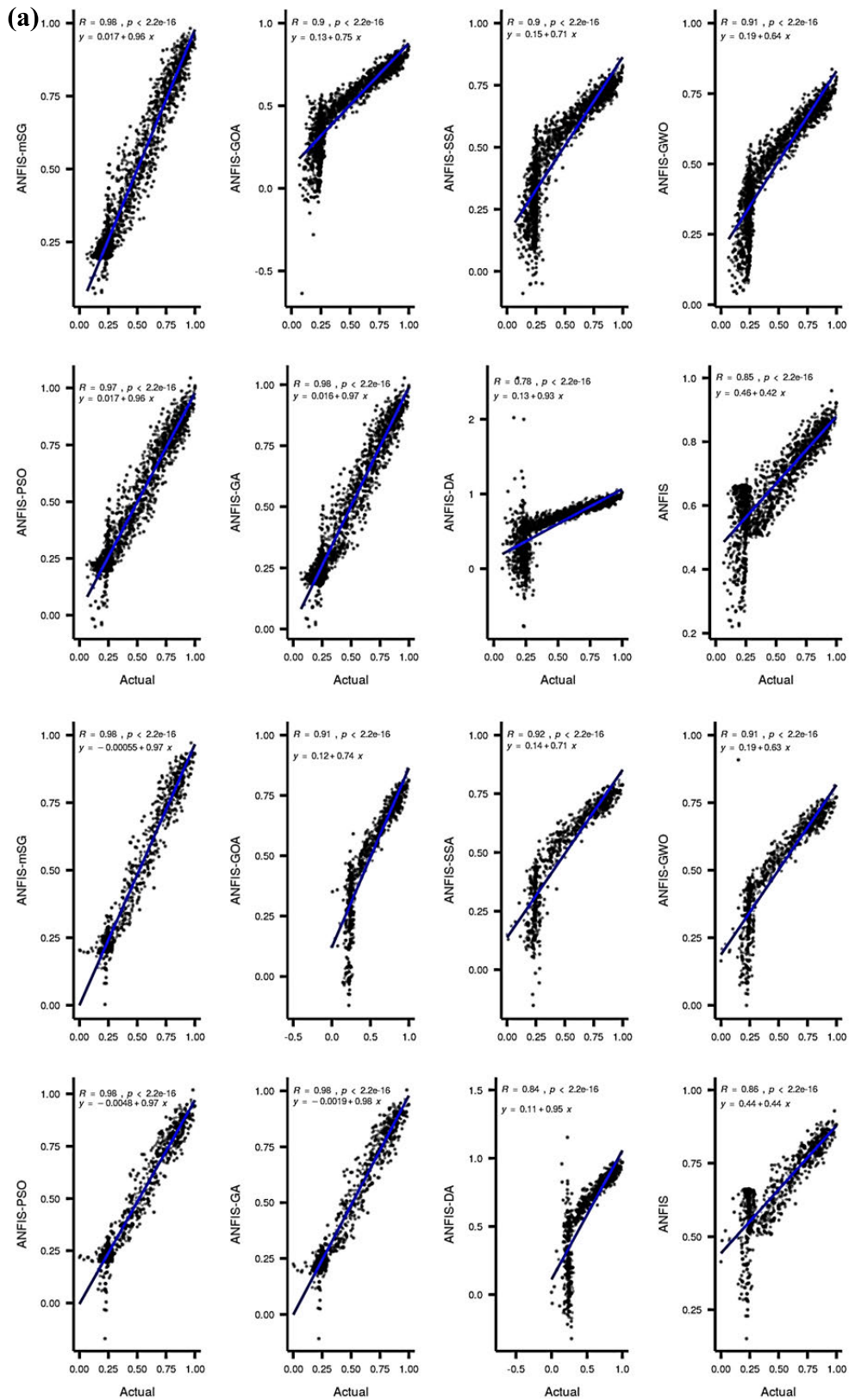
IEEE *Access*

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

(a)

FIGURE 5. Scatter plots for training and testing phases for predictive models at a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.
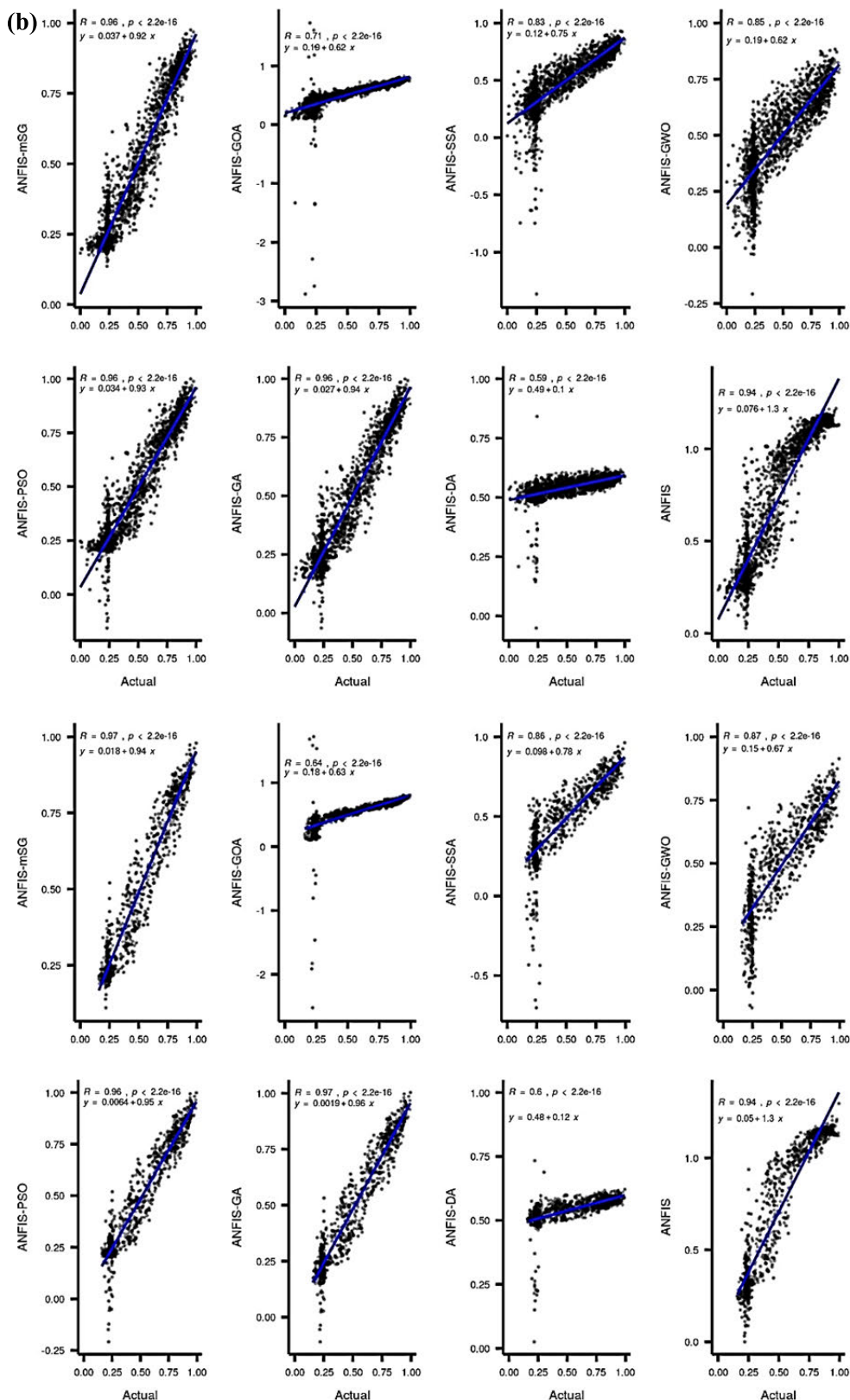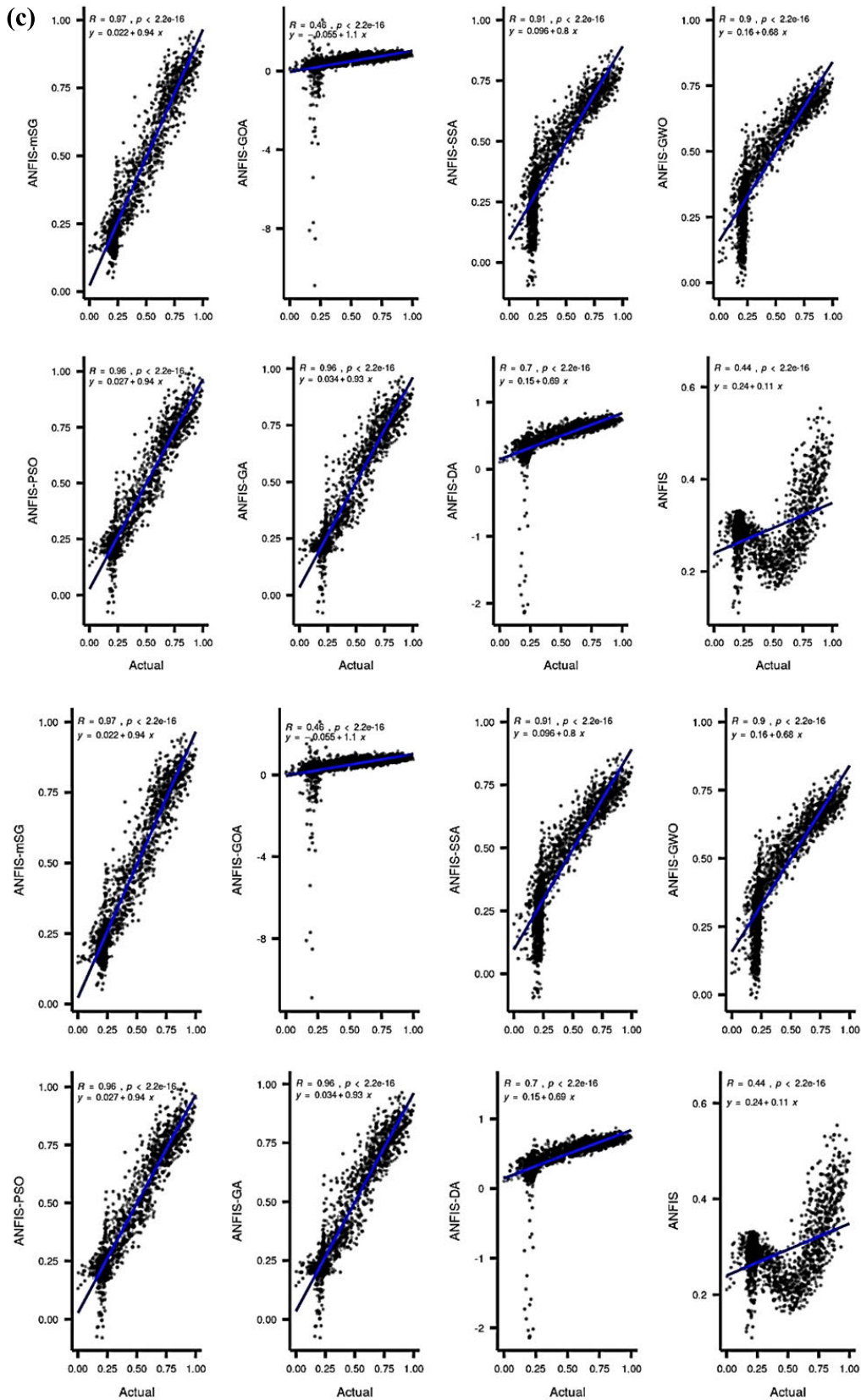
L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE *Access*

FIGURE 5. *(Continued)* Scatter plots for training and testing phases for predictive models at a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.
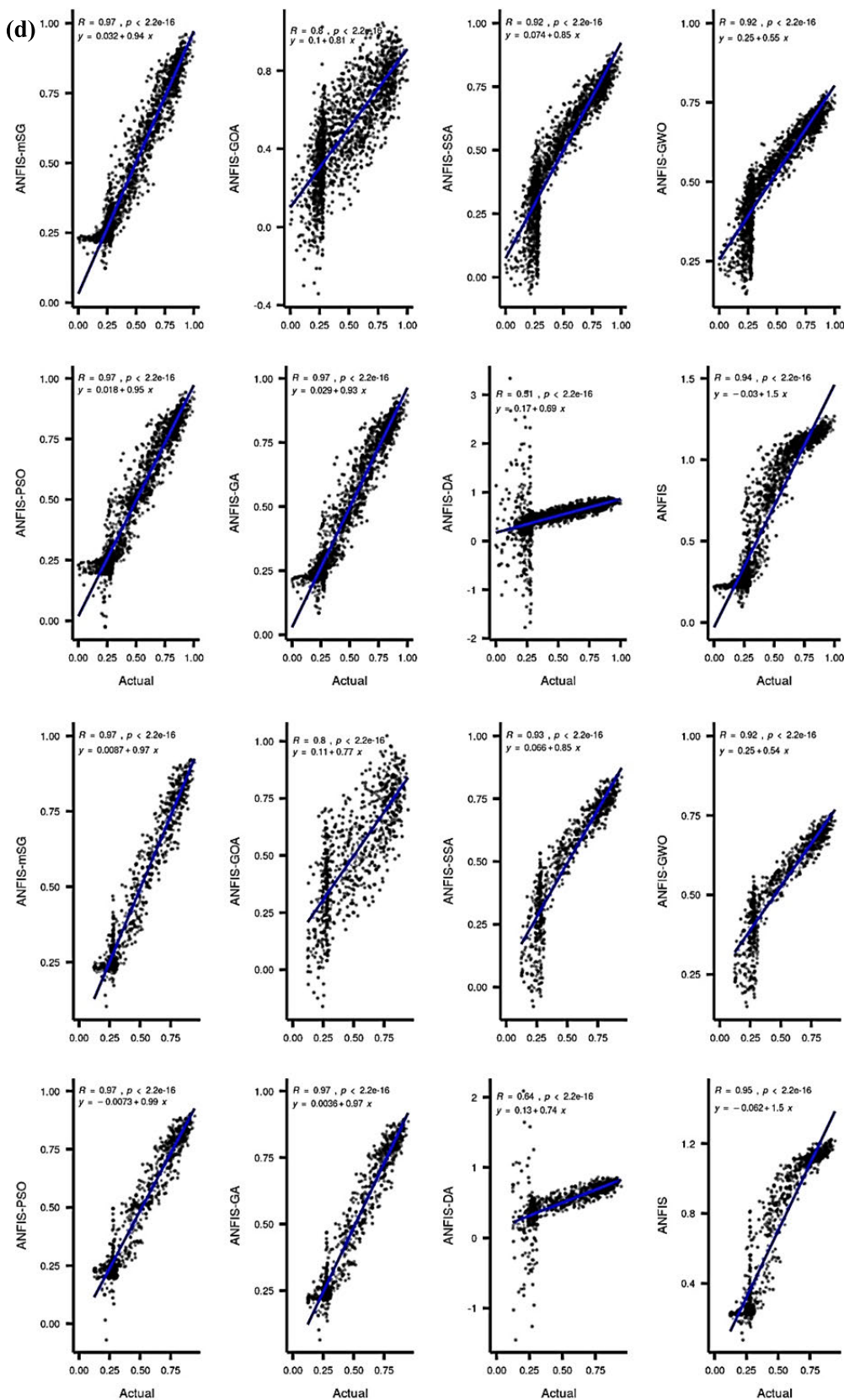
IEEE *Access*

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

FIGURE 5. *(Continued)* Scatter plots for training and testing phases for predictive models at a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.
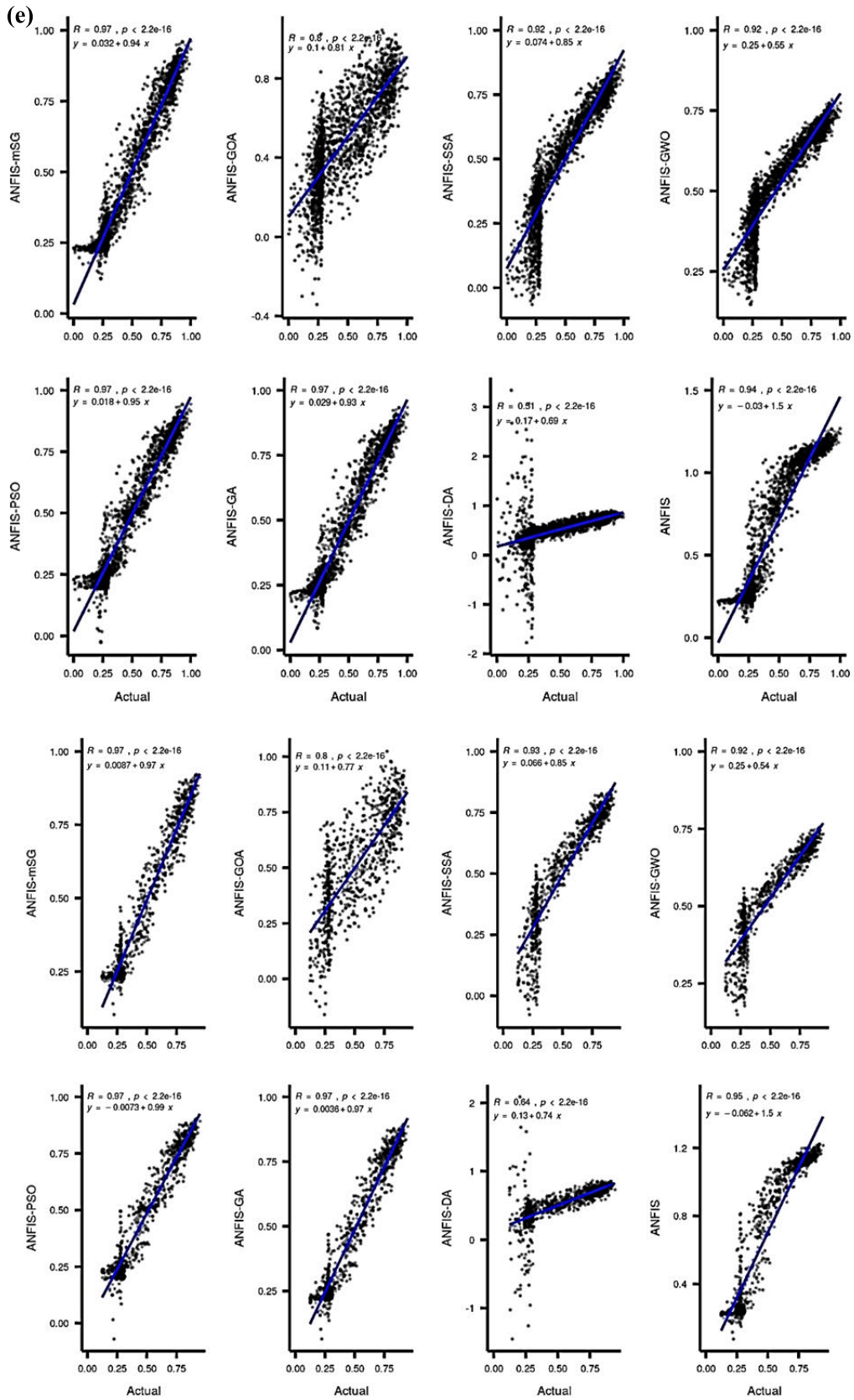
L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE *Access*

FIGURE 5. *(Continued)* Scatter plots for training and testing phases for predictive models at a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.
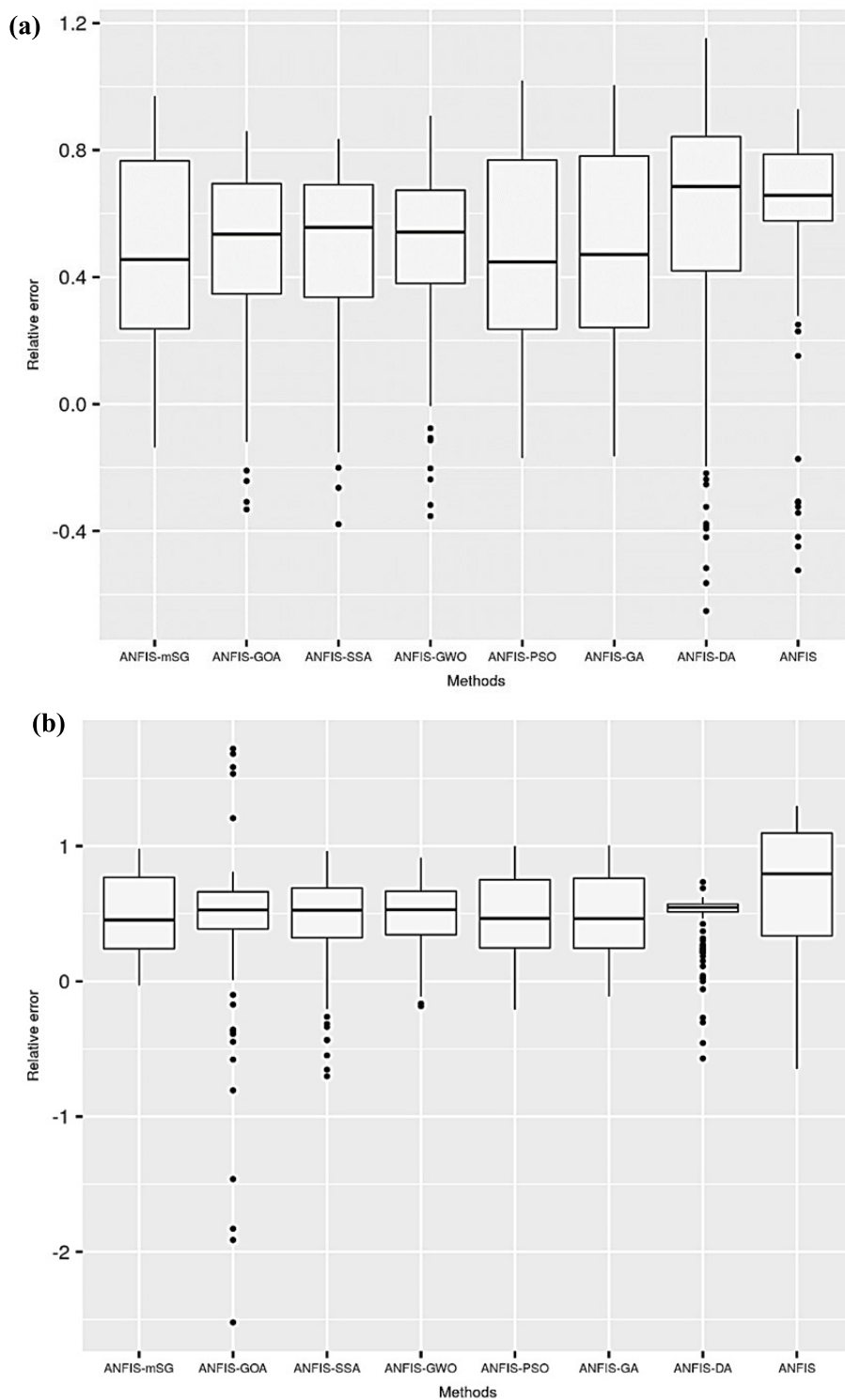
IEEE *Access*

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

(e)

FIGURE 5. *(Continued)* Scatter plots for training and testing phases for predictive models at a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE *Access*

FIGURE 6. Relative error for testing phase for predictive models at all meteorological stations: a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.

Fingal stations. There was no weighty gap appearing in line to IQR between ANIFS-PSO and ANIFS-GA except for higher minimum value in the case of ANIFS-PSO.

Based on the relative error in the testing phase for Cando station, the box plot exhibited no significant IQR difference among ANFIS-mSG, ANIFS-GO, ANIFS-PSO and ANIFS-GA, and the median value was slightly bending towards 1st quartile (Figure 6c). However, for the case of ANIFS-SSA, ANIFS-GWO and ANIFS-DA the median values were found towards the 3rd quartile with outliers towards the minimum
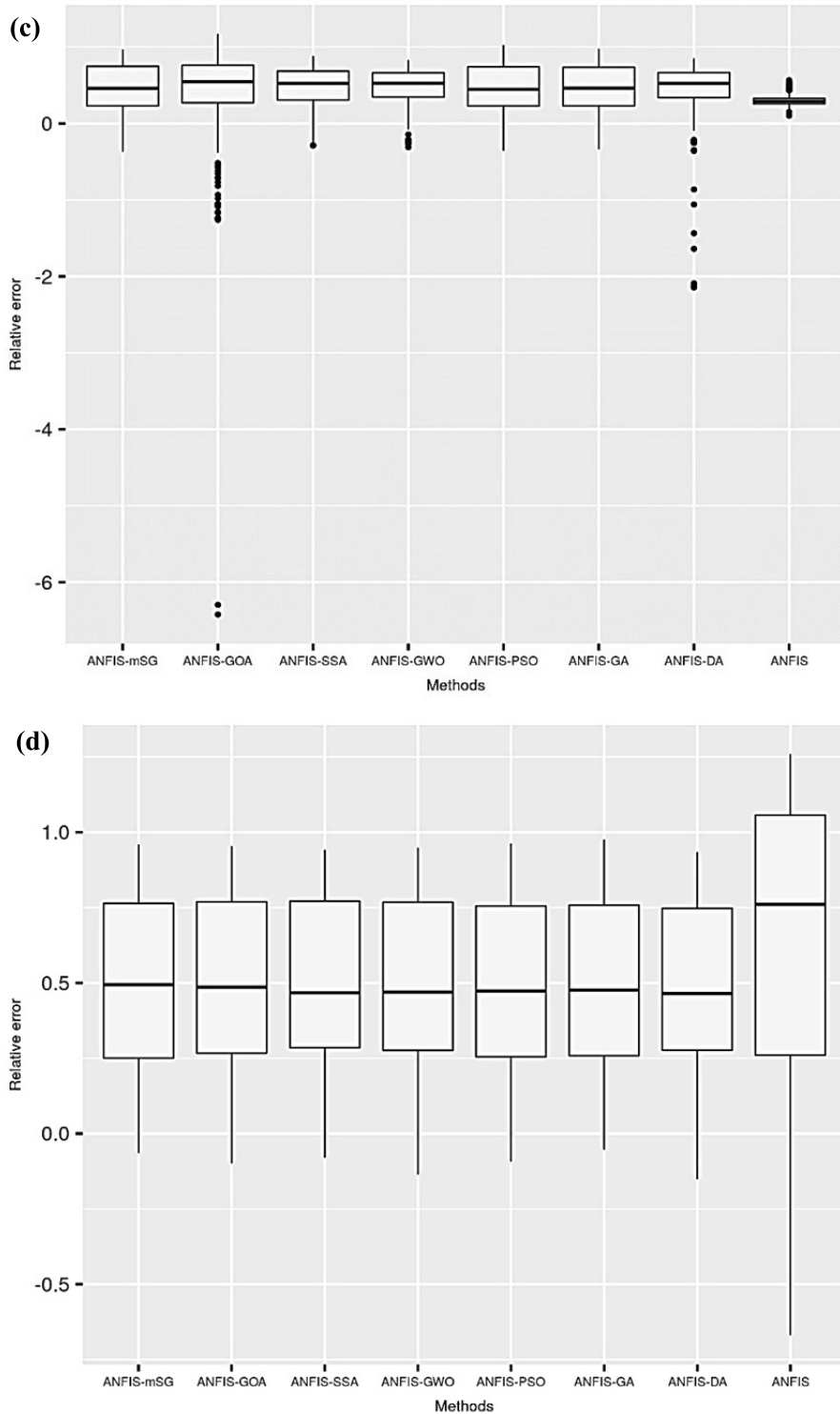
IEEE Access

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction
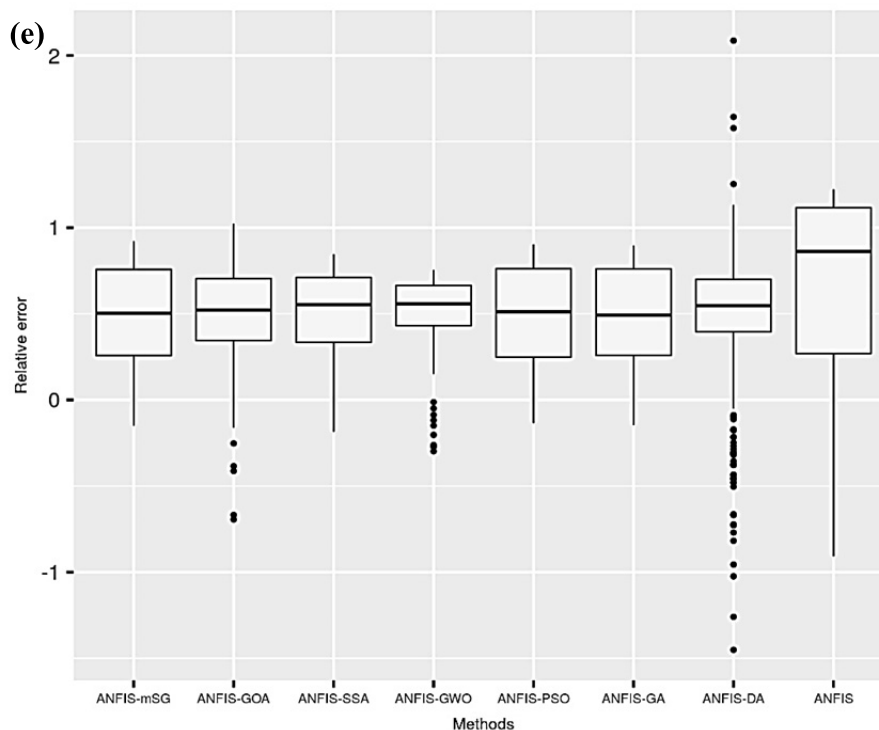
FIGURE 6. *(Continued)* Relative error for testing phase for predictive models at all meteorological stations: a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.

error value. ANFIS and ANIFS-DA displayed significant statistical differences i.e., least IQR value of ANFIS with maximum outliers inside of the maximum error value, where the opposite position of maximum outlies was found in the case of the ANIFS-DA model.

Figure 6d displays the boxplot of the relative error for the testing phase at Crary station. There was a noticeable-position of median observed in the case of ANFIS-mSG and ANFIS-GOA which was exactly between 1st and 3rd quartile. There was no major important gap of IQR among

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE *Access*

FIGURE 6. *(Continued)* Relative error for testing phase for predictive models at all meteorological stations: a) Baker, b) Beach, c) Cando, d) Crary and e) Fingal stations.

all models except ANFIS which had the largest value of IQR and the median bending towards 75 percentile relative error value. However, the performance of ANIFS-SSA, ANIFS-GWO, ANIFS-PSO, ANIFS-GA and ANIFS-DA exhibited the median value inclined to 25 percentile of relative error value. Figure 6e shows a noteworthy gap for IQR boxplots of ANIFS-GWO, ANIFS-DA, and ANFIS, whereas ANIFS-DA was found with a maximum number of outliers in the case of minimum as well as maximum, followed by ANIFS-GWO in line with outliers but limited to the minimum error value only. The median value of ANFIS and ANIFS-SSA slightly tended towards 3rd quartile, while ANFIS had he largest value of IQR among all models with the least relative error. There was no substantial break between the IQR value of ANFIS-mSG, ANIFS-GOA, ANIFS-PSO and ANIFS-GA; and the evident point was the median value appearing exactly between 1st and 3rd quartiles.

It is worth highlighting that soil physical properties, such as bulk density, moisture content, organic matter, and mineral type, could affect the specific heat capacity and thermal conductivity of the soil and in turn affect the diffusion coefficient of soil temperature. With the increase in depth, the coefficient of thermal diffusivity decreases in soil. Furthermore, the surface temperature transfers with a smaller lag time to deeper layers. Thus, with increasing depth, the heat flux changes the temperature of a larger volume of soil and ST is reduced further.

Soil thermal conductivity and soil moisture content (related to soil texture) have a strong influence on soil temperature gradients. Many hydrological, biogeochemical, biological processes and atmospheric water cycle are influenced by temperature and soil moisture. By definition, soil moisture refers to the total water volume (water vapor included) in unsaturated soil. Many microbes depend on water for their activity and survival. The biogeochemical environment for microorganisms is determined by soil moisture dynamics, as it affects dissolved nutrient availability, including organic carbon, ammonium, and nitrate. Soil moisture is highly important for microbial activity and diversity regulation. Despite the importance of soil moisture, it has not been considered as a parameter for weather prediction due to the complexity of its routine measurement over large areas. Accurate prediction of surface soil temperature can be achieved using analytical models as long as all the necessary input parameters are available, although this process is not efficient always. Based on the developed hybrid intelligence model, ANFIS-mSG, can be an effective and simple technique of soil temperature measurement since it does not require several input parameters.

V. CONCLUSION

This study aimed to predict ST using a univariate modeling scheme by incorporating only the air temperature information. It was evaluated the ability of the proposed hybrid intelligence model (i.e., ANFIS-mSG) to predict ST at different

IEEE *Access*

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

meteorological stations (i.e., Baker, Beach, Cando, Crary and Fingal) of North Dakota (ND), USA. The proposed hybrid intelligence model was validated against several benchmark models (i.e., ANFIS, ANFIS-DA, ANFIS-GA, ANFIS-GO, ANFIS-GWO, ANFIS-PSO and ANFIS-SSA). The training and testing modeling phases were conducted based on historical information of maximum, mean, and minimum air temperature using daily records for nine years (1 of January 2010 - 31 of December 2018). Multiple evaluation techniques including statistical and graphical were used to present the predictive models. The modeling results acknowledged the feasibility of hybridizing ANFIS model with optimization algorithms (mSG) for building a robust and efficient predictive model for soil temperature. The convergence time of the proposed ANFIS-mSG was relatively acceptable; however, the standalone ANFIS reported minimal time but with a high level of prediction error. At Carary station, there was a slight difference in the predictive performance among the developed hybrid models. The RMSE metrics was enhanced by 73%, 74.4%, 71.2%, 76.7% and 80.7% using the developed ANFIS-mSG over the standalone ANFIS model during the testing phase at Baker, Beach, Cando, Crary, and Fingal meteorological stations, respectively. In conclusion, the proposed hybrid artificial intelligence (i.e., ANFIS-mSG) model was found to be an efficient predictive model for soil temperature based on univariate air temperature scenario.

CONFLICT OF INTEREST
The authors declare no conflict of interest to any party.

REFERENCES
[1] J. R. Teasdale and C. L. Mohler, "Light transmittance, soil temperature, and soil moisture under residue of hairy vetch and rye," *Agronomy J.*, vol. 85, no. 3, pp. 673–680, May 1993.

[2] M. Béhaegel, P. Sailhac, and G. Marquis, "On the use of surface and ground temperature data to recover soil water content information," *J. Appl. Geophys.*, vol. 62, no. 3, pp. 234–243, Jul. 2007.

[3] V. M. Sayão, J. A. M. Demattê, L. G. Bedin, M. R. Nanni, and R. Rizzo, "Satellite land surface temperature and reflectance related with soil attributes," *Geoderma*, vol. 325, pp. 125–140, Sep. 2018.

[4] G. Mihalakakou, "On estimating soil surface temperature profiles," *Energy Buildings*, vol. 34, no. 3, pp. 251–259, Mar. 2002.

[5] O. Kisi, H. Sanikhani, and M. Cobaner, "Soil temperature modeling at different depths using neuro-fuzzy, neural network, and genetic programming techniques," *Theor. Appl. Climatol.*, vol. 129, pp. 833–848, May 2016.

[6] H. Sanikhani, R. C. Deo, Z. M. Yaseen, O. Eray, and O. Kisi, "Non-tuned data intelligent model for soil temperature estimation: A new approach," *Geoderma*, vol. 330, pp. 52–64, Nov. 2018.

[7] M. Bilgili, "Prediction of soil temperature using regression and artificial neural network models," *Meteorol. Atmos. Phys.*, vol. 110, nos. 1–2, pp. 59–70, Dec. 2010.

[8] T. Yilmaz, A. Özbek, A. Yilmaz, and O. Büyükalaca, "Influence of upper layer properties on the ground temperature distribution," *Isi Bilimi ve Teknigi Dergisi/J. Thermal Sci. Technol.*, vol. 29, no. 2, pp. 43–51, 2009.

[9] K. Koçak, L. Şaylan, and J. Eitzinger, "Nonlinear prediction of near-surface temperature via univariate and multivariate time series embedding," *Ecol. Model.*, vol. 173, no. 1, pp. 1–7, Mar. 2004.

[10] Y. Feng, N. Cui, W. Hao, L. Gao, and D. Gong, "Estimation of soil temperature from meteorological data using different machine learning models," *Geoderma*, vol. 338, pp. 67–77, Mar. 2019.

[11] C. Chatfield, "Time-series forecasting," *Significance*, vol. 2, no. 3, pp. 131–133, 2005.

[12] B. Aljoumani, J. A. Sànchez-Espigares, N. Cañameras, R. Josa, and J. Monserrat, "Time series outlier and intervention analysis: Irrigation management influences on soil water content in silty loam soil," *Agricult. Water Manage.*, vol. 111, pp. 105–114, Aug. 2012.

[13] P. Zou, J. Yang, J. Fu, G. Liu, and D. Li, "Artificial neural network and time series models for predicting soil salt and water content," *Agricult. Water Manage.*, vol. 97, no. 12, pp. 2009–2019, Nov. 2010.

[14] M. Kumar, A. Kumar, N. C. Mahanti, C. Mallik, and R. K. Shukla, "Surface flux modelling using ARIMA technique in humid subtropical monsoon area," *J. Atmos. Solar-Terr. Phys.*, vol. 71, no. 12, pp. 1293–1298, Aug. 2009.

[15] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *J. Hydrol.*, vol. 569, pp. 387–408, Feb. 2019.

[16] H. Tabari, A.-A. Sabziparvar, and M. Ahmadi, "Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region," *Meteorol. Atmos. Phys.*, vol. 110, nos. 3–4, pp. 135–142, Jan. 2011.

[17] S. Mehdizadeh, J. Behmanesh, and K. Khalili, "Evaluating the performance of artificial intelligence methods for estimation of monthly mean soil temperature without using meteorological data," *Environ. Earth Sci.*, vol. 76, no. 8, p. 325, Apr. 2017.

[18] S. M. R. Kazemi, B. Minaei Bidgoli, S. Shamshirband, S. M. Karimi, M. A. Ghorbani, K.-W. Chau, and R. Kazem Pour, "Novel genetic-based negative correlation learning for estimating soil temperature," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 506–516, Jan. 2018.

[19] M. Zeynoddin, H. Bonakdari, I. Ebtehaj, F. Esmaeilbeiki, B. Gharabaghi, and D. Zare Haghi, "A reliable linear stochastic daily soil temperature forecast model," *Soil Tillage Res.*, vol. 189, pp. 73–87, Jun. 2019.

[20] H. Zare Abyaneh, M. Bayat Varkeshi, G. Golmohammadi, and K. Mohammadi, "Soil temperature estimation using an artificial neural network and co-active neuro-fuzzy inference system in two different climates," *Arabian J. Geosci.*, vol. 9, no. 5, p. 377, May 2016.

[21] O. Kisi and Z. M. Yaseen, "The potential of hybrid evolutionary fuzzy intelligence model for suspended sediment concentration prediction," *Catena*, vol. 174, pp. 11–23, Mar. 2019.

[22] Z. M. Yaseen, A. M. Al-Juboori, U. Beyaztas, N. Al-Ansari, K.-W. Chau, C. Qi, M. Ali, S. Q. Salih, and S. Shahid, "Prediction of evaporation in arid and semi-arid regions: A comparative study using different machine learning models," *Eng. Appl. Comput. Fluid Mech.*, vol. 14, no. 1, pp. 70–89, Jan. 2020.

[23] H. Tabari, P. Hosseinzadeh Talaee, and P. Willems, "Short-term forecasting of soil temperature using artificial neural network," *Meteorol. Appl.*, vol. 22, no. 3, pp. 576–585, Jul. 2015.

[24] W. Wu, X.-P. Tang, N.-J. Guo, C. Yang, H.-B. Liu, and Y.-F. Shang, "Spatiotemporal modeling of monthly soil temperature using artificial neural networks," *Theor. Appl. Climatol.*, vol. 113, nos. 3–4, pp. 481–494, Aug. 2013.

[25] M. Ozgoren, M. Bilgili, and B. Sahin, "Estimation of global solar radiation using ANN over turkey," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5043–5051, Apr. 2012.

[26] H. Bonakdari, H. Moeeni, I. Ebtehaj, M. Zeynoddin, A. Mahoammadian, and B. Gharabaghi, "New insights into soil temperature time series modeling: Linear or nonlinear?" *Theor. Appl. Climatol.*, vol. 135, nos. 3–4, pp. 1157–1177, Feb. 2019.

[27] L. Xing, L. Li, J. Gong, C. Ren, J. Liu, and H. Chen, "Daily soil temperatures predictions for various climates in united states using data-driven model," *Energy*, vol. 160, pp. 430–440, Oct. 2018.

[28] S. Samadianfard, E. Asadi, S. Jarhan, H. Kazemi, S. Kheshtgar, O. Kisi, S. Sajjadi, and A. A. Manaf, "Wavelet neural networks and gene expression programming models to predict short-term soil temperature at different depths," *Soil Tillage Res.*, vol. 175, pp. 37–50, Jan. 2018.

[29] J. Behmanesh and S. Mehdizadeh, "Estimation of soil temperature using gene expression programming and artificial neural networks in a semiarid region," *Environ. Earth Sci.*, vol. 76, no. 2, p. 76, Jan. 2017.

[30] R. Moazenzadeh and B. Mohammadi, "Assessment of bio-inspired Metaheuristic optimisation algorithms for estimating soil temperature," *Geoderma*, vol. 353, pp. 152–171, Nov. 2019.

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

IEEE*Access*

[31] M. A. Ghorbani, R. C. Deo, V. Karimi, Z. M. Yaseen, and O. Terzi, "Implementation of a hybrid MLP-FFA model for water level prediction of lake egirdir, turkey," *Stochastic Environ. Res. Risk Assessment*, vol. 32, no. 6, pp. 1683–1697, Jun. 2018.

[32] S. Naganna, P. Deka, M. Ghorbani, S. Biazar, N. Al-Ansari, and Z. Yaseen, "Dew point temperature estimation: Application of artificial intelligence model integrated with nature-inspired optimization algorithms," *Water*, vol. 11, no. 4, p. 742, 2019.

[33] I. Ebtehaj and H. Bonakdari, "Design of a fuzzy differential evolution algorithm to predict non-deposition sediment transport," *Appl. Water Sci.*, vol. 7, no. 8, pp. 4287–4299, Dec. 2017.

[34] H. Azimi, H. Bonakdari, I. Ebtehaj, B. Gharabaghi, and F. Khoshbin, "Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length," *Acta Mech.*, vol. 229, no. 3, pp. 1197–1214, Mar. 2018.

[35] H. Azimi, H. Bonakdari, I. Ebtehaj, S. H. Ashraf Talesh, D. G. Michelson, and A. Jamali, "Evolutionary Pareto optimization of an ANFIS network for modeling scour at pile groups in clear water condition," *Fuzzy Sets Syst.*, vol. 319, pp. 50–69, Jul. 2017.

[36] H. Azimi, S. Shabanlou, I. Ebtehaj, H. Bonakdari, and S. Kardar, "Combination of computational fluid dynamics, adaptive neuro-fuzzy inference system, and genetic algorithm for predicting discharge coefficient of rectangular side orifices," *J. Irrigation Drainage Eng.*, vol. 143, no. 7, Jul. 2017, Art. no. 04017015.

[37] I. Ebtehaj, H. Bonakdari, and B. Gharabaghi, "Development of more accurate discharge coefficient prediction equations for rectangular side weirs using adaptive neuro-fuzzy inference system and generalized group method of data handling," *Measurement*, vol. 116, pp. 473–482, Feb. 2018.

[38] H. Azimi, H. Bonakdari, I. Ebtehaj, and D. G. Michelson, "A combined adaptive neuro-fuzzy inference system–firefly algorithm model for predicting the roller length of a hydraulic jump on a rough channel bed," *Neural Comput. Appl.*, vol. 29, pp. 249–258, Aug. 2016.

[39] I. Ebtehaj and H. Bonakdari, "Assessment of evolutionary algorithms in predicting non-deposition sediment transport," *Urban Water J.*, vol. 13, no. 5, pp. 499–510, Jul. 2016.

[40] K. Mohammadi, S. Shamshirband, A. Kamsin, P. C. Lai, and Z. Mansor, "Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure," *Renew. Sustain. Energy Rev.*, vol. 63, pp. 423–434, Sep. 2016.

[41] O. Castillo and P. Melin, "Optimization of type-2 fuzzy systems based on bio-inspired methods: A concise review," *Inf. Sci.*, vol. 205, pp. 1–19, Nov. 2012.

[42] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems," *Adv. Eng. Softw.*, vol. 114, pp. 163–191, Dec. 2017.

[43] H. Faris, M. M. Mafarja, A. A. Heidari, I. Aljarah, A. M. Al-Zoubi, S. Mirjalili, and H. Fujita, "An efficient binary salp swarm algorithm with crossover scheme for feature selection problems," *Knowl.-Based Syst.*, vol. 154, pp. 43–67, Aug. 2018.

[44] G. I. Sayed, G. Khoriba, and M. H. Haggag, "A novel chaotic salp swarm algorithm for global optimization and feature selection," *Int. J. Speech Technol.*, vol. 48, no. 10, pp. 3462–3481, Oct. 2018.

[45] R. Abbassi, A. Abbassi, A. A. Heidari, and S. Mirjalili, "An efficient salp swarm-inspired algorithm for parameters identification of photovoltaic cell models," *Energy Convers. Manage.*, vol. 179, pp. 362–372, Jan. 2019.

[46] M. Mafarja, I. Aljarah, A. A. Heidari, A. I. Hammouri, H. Faris, A. M. Al-Zoubi, and S. Mirjalili, "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems," *Knowl.-Based Syst.*, vol. 145, pp. 25–45, Apr. 2018.

[47] S. Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, and I. Aljarah, "Grasshopper optimization algorithm for multi-objective optimization problems," *Int. J. Speech Technol.*, vol. 48, no. 4, pp. 805–820, Apr. 2018.

[48] M. C. Peel, B. L. Finlayson, and T. A. Mcmahon, "Updated world map of the Köppen-geiger climate classification," *Hydrol. Earth Syst. Sci. Discuss.*, vol. 4, no. 2, pp. 439–473, 2007.

[49] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, Jun. 1993.

[50] Z. Yaseen, I. Ebtehaj, S. Kim, H. Sanikhani, H. Asadi, M. Ghareb, H. Bonakdari, W. Wan Mohtar, N. Al-Ansari, and S. Shahid, "Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis," *Water*, vol. 11, no. 3, p. 502, 2019.

[51] N. Walia, H. Singh, and A. Sharma, "ANFIS: Adaptive neuro-fuzzy inference System- a survey," *Int. J. Comput. Appl.*, vol. 123, no. 13, pp. 32–38, 2015.

[52] J. C. Sutherland, "Linear dichroism of DNA: Characterization of the orientation distribution function caused by hydrodynamic shear," *Anal. Biochem.*, vol. 523, pp. 24–31, Apr. 2017.

[53] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: Theory and application," *Adv. Eng. Softw.*, vol. 105, pp. 30–47, Mar. 2017.

[54] S. M. Rogers, "Mechanosensory-induced behavioural gregarization in the desert locust schistocerca gregaria," *J. Exp. Biol.*, vol. 206, no. 22, pp. 3991–4002, Nov. 2003.

[55] M.-Y. Chen, "A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering," *Inf. Sci.*, vol. 220, pp. 180–195, Jan. 2013.

[56] A. Sarkheyli, A. M. Zain, and S. Sharif, "A multi-performance prediction model based on ANFIS and new modified-GA for machining processes," *J. Intell. Manuf.*, vol. 26, no. 4, pp. 703–716, Aug. 2015.

[57] H. Tao, L. Diop, A. Bodian, K. Djaman, P. M. Ndiaye, and Z. M. Yaseen, "Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso," *Agricult. Water Manage.*, vol. 208, pp. 140–151, Sep. 2018.

[58] W. Chen, M. Panahi, and H. R. Pourghasemi, "Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling," *Catena*, vol. 157, pp. 310–324, Oct. 2017.

LIU PENGHUI received the bachelor's degree in computer science from Yan'an University, in 1992, and the M.Sc. degree from Rio Grande University, USA, in 2008. He is currently the Dean of the School of Computer Science, Baoji University of Arts and Sciences, the Dean of the Baoji Institute of Digital Economy, the Director of the Baoji Artificial Intelligence Industry Development Promotion Center, and also the Director of the National Association of Artificial Intelligence and Big Data Innovation. His research areas cover the Internet of Things, rehabilitation robots, 3D pattern recognition, and integrated micro intelligent monitoring systems.

AHMED A. EWEES received the Ph.D. degree from Damietta University, in 2012. He co-advises master's and Ph.D. students, and leading and supervising various graduation projects. He currently works as an Associate Professor of computer science with Damietta University, Egypt. He has published many scientific research articles in international journals and conferences, and also has written several book chapters. His current research interests include machine learning, artificial intelligence, image processing, metaheuristic optimization techniques, text mining, and natural language processing.

BESTE HAMIYE BEYAZTAS received the Ph.D. degree in statistics from Dokuz Eylul University. She is currently an Assistant Professor with the Department of Statistics, İstanbul Medeniyet University, Turkey. Her research interests are time series analysis, resampling methods, forecasting, and panel data analysis.

IEEE Access

L. Penghui *et al.*: Metaheuristic Optimization Algorithms Hybridized With Artificial Intelligence Model for ST Prediction

CHONGCHONG QI received the B.S. degree from the China University of Mining and Technology and the Ph.D. degree from The University of Western Australia. He is currently a Professor with Central South University, Changsha, China. He has authored more than 40 articles. His research interests include the application of artificial intelligence in civil and mining engineering. His awards and honors include the Australian Research Award (Western Australia Chinese Scientists Association) and Second Prize for Undergraduate Thesis in Jiangsu Province (Jiangsu Province Education Department, China).

SINAN Q. SALIH received the B.Sc. degree in information systems from the University of Anbar, Anbar, Iraq, in 2010, the M.Sc. degree in computer sciences from Universiti Tenaga National (UNITEN), Malaysia, in 2012, and the Ph.D. degree in soft modeling and intelligent systems from Universiti Malaysia Pahang (UMP). His current research interests include optimization algorithms, nature-inspired metaheuristics, machine learning, and feature selection problem for real world problems.

NADHIR AL-ANSARI received the B.Sc. and M.Sc. degrees from the University of Baghdad, in 1968 and 1972, respectively, and the Ph.D. degree in water resources engineering from Dundee University, in 1976. He worked at Baghdad University, from 1976 to 1995, and also with Al-Bayt University, Jordan, from 1995 to 2007. He served in several academic administrative posts including the Dean of the Head of Department. He is currently a Professor with the Department of Civil, Environmental, and Natural Resources Engineering, Luleå University of Technology, Sweden. His publications include more than 424 articles in international/national journals, chapters in books and 13 books. He has one patent on physical methods for the separation of iron oxides. He supervised more than 66 postgraduate students at Iraq, Jordan, U.K., and Australia universities. He executed more than 60 major research projects in Iraq, Jordan, and U.K. His research interests are mainly in geology, water resources, and environment. He was a member of several scientific societies, e.g., the International Association of Hydrological Sciences, the Chartered Institution of Water and Environment Management, the Network of Iraqi Scientists Abroad, and the Founder and the President of the Iraqi Scientific Society for Water Resources. He was a member of the Editorial Board of ten international journals. He awarded several scientific and educational awards, among them is the British Council on its 70th Anniversary awarded him top 5 scientists in Cultural Relations.

SURAJ KUMAR BHAGAT received the master's degree (Hons.) in environmental engineering from the Rajasthan Technical University, India, in 2014. He is currently pursuing the Ph.D. degree in civil engineering from Ton Duc Thang University, Ho Chi Minh, Vietnam. He also received a research fellowship at IIT Delhi, India, in 2013. In 2014, he joined the Department of Civil Engineering, Institute of Technology, Ambo University, Ethiopia, as a Lecturer, where he also led a funded project which received the Energy Globe Award, issued by the Austrian Energy Pioneer Wolfgang Neumann. He is currently an Operations Director (Volunteer) of the Agua International Water Relief (NGO), USA.
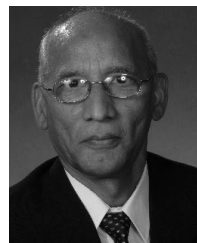
ZAHER MUNDHER YASEEN received the master's and Ph.D. degrees in hydrology, water resources engineering, hydrological processes modeling, environmental engineering, and climate from the National University of Malaysia (UKM), Malaysia. He is currently a Senior Lecturer and a Senior Researcher in civil engineering with Ton Duc Thang University. In addition, he has an excellent expertise in machine learning and advanced data analytics. He has published more than 100 articles in international journals with a Google Scholar H-index of 24, and a total of 1724 citations.

VIJAY P. SINGH received the B.S., M.S., Ph.D., and D.Sc. degrees in engineering. He is currently a Distinguished Professor, a Regents Professor, and the Caroline and William N. Lehrer Distinguished Chair in water engineering with Texas A&M University. He is also a Registered Professional Engineer, a Registered Professional Hydrologist, and an Honorary Diplomate of ASCE-AAWRE. He has published extensively in the area of hydrology and water resources, including 30 textbooks; *Handbook of Applied Hydrology*; *Encyclopedia of Snow, Ice and Glaciers*; 71 edited books; 114 book chapters; 1270 refereed journal articles; 330 conference proceedings articles; 13 special edited journal issues; 50 book reviews; and 72 technical publications and reports. For his seminal contributions, he has received more than 92 national and international awards, and three honorary doctorates. He is a member of 11 international science/engineering academies. He has served/serves as an Editor-in-Chief of three journals and two book series and serves on editorial boards of more than 25 journals and three book series. He has served as the President for the American Institute of Hydrology (AIH), the Chair of the Watershed Council, American Society of Civil Engineers, and the Vice President of the Indian Association of Hydrologists and of Association of Global Groundwater Scientists. He is currently the President of the American Academy of Water Resources Engineers.

●●●