

Research Article

Using Improved XGBoost Algorithm to Obtain Modified Atmospheric Refractive Index

Yanbo Mai ^{1,2}, Zheng Sheng ², Hanqing Shi ² and Qixiang Liao ²

¹College of Electronic Science and Technology, National University of Defense Technology, Changsha 410074, China

²College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410074, China

Correspondence should be addressed to Zheng Sheng; 19994035@sina.com

Received 19 June 2021; Revised 4 September 2021; Accepted 7 September 2021; Published 23 September 2021

Academic Editor: Stefania Bonafoni

Copyright © 2021 Yanbo Mai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Atmospheric refraction is a special meteorological phenomenon mainly caused by gas molecules and aerosol particles in the atmosphere, which can change the propagation direction of electromagnetic waves in the atmospheric environment. Atmospheric refractive index, an index to measure atmospheric refraction, is an important parameter for electromagnetic wave. Given that it is difficult to obtain the atmospheric refractive index of 100 meters (m)–3000 m over the ocean, this paper proposes an improved extreme gradient boosting (XGBoost) algorithm based on comprehensive learning particle swarm optimization (CLPSO) operator to obtain them. Finally, the mean absolute percentage error (MAPE) and root mean-squared error (RMSE) are used as evaluation criteria to compare the prediction results of improved XGBoost algorithm with backpropagation (BP) neural network and traditional XGBoost algorithm. The results show that the MAPE and RMSE of the improved XGBoost algorithm are 39% less than those of BP neural network and 32% less than those of the traditional XGBoost. Besides, the improved XGBoost algorithm has the strongest learning and generalization capability to calculate missing values of atmospheric refractive index among the three algorithms. The results of this paper provide a new method to obtain atmospheric refractive index, which will be of great reference significance to further study the atmospheric refraction.

1. Introduction

Atmospheric refraction, which is mainly caused by the gas molecules and aerosol particles, is a special phenomenon of atmospheric environment; it can change the propagation route in the atmospheric environment and then cause the abnormal propagation of electromagnetic wave. Especially under the condition of trapping refraction, the electromagnetic wave will be caught in a certain thickness of atmosphere and propagate back and forth through the upper and lower gas layers; this propagation phenomenon is called atmospheric ducts' propagation, and the atmosphere which causes this abnormal phenomenon is called atmospheric duct. Because atmospheric duct can significantly change the route of electromagnetic wave, it has a great impact on the radar system, navigation system, and communication system [1]. If we make full use of the atmospheric duct's effect, we can realize the beyond-the-horizon transmission and

target detection. Atmospheric refractive index is an important index to evaluate the degree of atmospheric refraction, and the change of atmospheric refractive index is one of the important bases to judge the appearance of atmospheric ducts. The reason for the formation of atmospheric duct is that the atmospheric temperature and humidity decrease sharply with the height, which makes the atmospheric refractive index decrease gradually and then leads to the formation of the atmospheric ducts [2].

So far, many scholars have studied the atmospheric refractive index [3–13]. The traditional way to obtain the refractive index of the atmosphere is to measure profiles of temperature and humidity by radiosonde and then estimate the atmospheric refractive index [14, 15]; later, Mathai and Harrison [16] used the measured volume scattering coefficient and particle size distribution to estimate the refractive index of atmospheric aerosol. Besides, radar phase information can also be used to retrieve atmospheric refractive

index [17, 18], and some estimation methods of the atmospheric structure constant also provide reference for obtaining atmospheric refractive index [19–21]. In recent years, some scholars use new methods to measure the refractive index of the atmosphere. For instance, Dinar et al. [22] used a cavity ring-down aerosol spectrometer to obtain the complex refractive index of the atmosphere; Xie et al. [23] used one-dimensional variational assimilation algorithm to obtain atmospheric refractive index from ground-based Global Positioning System (GPS) phase delay. Besides, with the development of computer, machine learning is applied more and more in the field of meteorology [24–33]. Machine learning is a process of using computers to summarize the existing data, get its general rules, and establish the corresponding mapping relationship [34]. It is very suitable for data analysis, especially in the case of large amount of data [35]. Extreme gradient boosting (XGBoost) is an integrated model based on a decision tree [36, 37]. It is a typical representative of machine learning, which shows a strong generalization force in big data prediction in recent years. Because the meteorological data are recorded by sea surface buoys generally below 100 m, and occultation data and meteorological satellite remote sensing data are not sufficiently accurate above the ocean at a height below 3000 m, the meteorological data between 100 m and 3000 m are often missing; hence, the atmospheric refractive index at the certain height cannot be calculated, and then the determination of atmospheric ducts is difficult and inefficient; so calculating the atmospheric refractive index 100 m–3000 m over the ocean by using machine learning to determine the occurrence of atmospheric ducts is a focus of research in meteorological data processing.

In this paper, we propose an improved XGBoost based on the CLPSO operator, which has a fast convergence speed and strong ability to jump out of the local optimum solution for parameter optimization combination to obtain the atmospheric refractive index. To test its effectiveness, it is applied to filling missing values in the modified atmospheric refractive index over the ocean. The modified atmospheric refractive indices in the lower (0–100 m) and upper (3000 m–10000 m) layers are used as input, and those in the middle layer (100 m–3000 m) are used as output. The improved XGBoost algorithm is then used for learning and training to fill in missing values of modified atmospheric refractive index in the middle layer. Finally, to verify its feasibility for calculating the modified atmospheric refractive index, the mean absolute percentage error (MAPE) and root mean-squared error (RMSE) are used as evaluation criteria to compare the prediction results of the proposed method with backpropagation (BP) neural network and traditional XGBoost algorithm; the results show that the improved XGBoost algorithm significantly improves accuracy and reduces operation time and has stronger learning and generalization capability to calculate missing values of modified atmospheric refractive index than other algorithms. The results of this paper provide a new method to obtain atmospheric refractive index, which will be of great reference significance to avoid the loss of electromagnetic wave propagation in the future.

2. Materials and Methods

The test data used in this paper are high-resolution sounding balloon data [38] collected from 1998 to 2008 at an observation station in Hawaii, USA, at 155.1°W longitude and 19.7°N latitude; the sounding balloon is released at UTC = 12 every day, and the vertical meteorological data are recorded once a day. The balloon recorded such meteorological parameters as temperature, pressure, and humidity at different height with a vertical resolution of 20 m. The average thickness of the atmospheric ducts is between tens and hundreds of meters [1]. Therefore, this sounding data can be used to study the atmospheric refractive index and atmospheric duct over the ocean.

Atmospheric duct is an abnormal phenomenon in the tropospheric atmospheric environment. It is mainly due to the uneven distribution of atmospheric temperature and humidity in the vertical direction, which leads to the significant difference of atmospheric refractive index in the vertical direction and then leads to the occurrence of atmospheric ducts [1]. In order to treat the earth's surface as a plane and judge the change of atmospheric refractive index conveniently and then evaluate the gradient of atmospheric refractive index and its influence on electromagnetic wave propagation more easily, modified atmospheric refractive index M is defined. The calculation formula of modified atmospheric refractive index M is as follows [1]:

$$M = N + 0.157Z = 77.6 \times \frac{P}{T} + 3.73 \times 10^5 \times \frac{e}{T^2} + 0.157 \times Z. \quad (1)$$

In equation (1), N is the atmospheric refractive index, P is the air pressure and its unit is *pha*, T is the temperature and its unit is *K*, e is the water vapor pressure and its unit is *pha*, and Z is the altitude and its unit is *m*. According to the variation of M with height, atmospheric ducts are mainly divided into four categories [1]: surface ducts, surface-based duct, elevated ducts, and evaporation ducts, as shown in Figure 1. Surface ducts mainly appear on the land, and their height is usually less than 300 meters. Elevated ducts mainly appear at high altitude over land and ocean, usually between 300 meters and 3000 meters. Evaporation ducts mainly appear over the ocean, usually below 40 meters in height.

3. BP Neural Network

3.1. Introduction to BP Neural Network. Artificial neural networks, which are widely used in data prediction, use mathematical methods to simulate the structure, function, and mode of processing of biological neural systems to build a complete information processing system. In this paper, the BP neural network model [39, 40], which is relatively mature and widely used, is used to fill the modified atmospheric refractive index of the middle layer (100 m–3000 m) by inputting the modified atmospheric refractive index at the lower (0–100 m) and higher (3000 m–10000 m) layers.

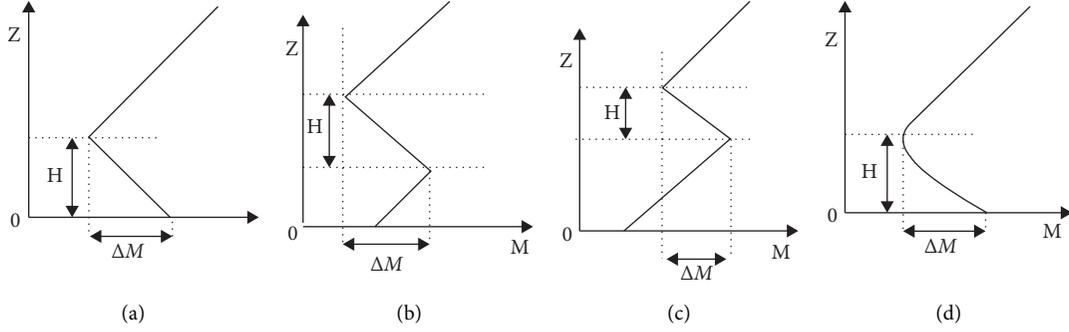


FIGURE 1: Categories of atmospheric ducts: (a) surface duct, (b) surface-based duct, (c) elevated duct, and (d) evaporation duct.

3.2. Model Building

3.2.1. Sample Selection. Because the number of samples is very large, the dimensionality of the input vector will become too large, and the generalization ability will be poor if the traditional sample selection method is used. So we need to reconstruct the model's input vector. In this paper, the input vector is constructed by calculating the correlation coefficients of different time samples, and sequences with large correlation coefficients are included in the input samples. The correlation coefficient is calculated as follows: if S and U are used to represent two sequences, their correlation coefficients of S and U are as follows [34]:

$$r_{SU} = \frac{\sum_{t=1}^l (S_t - \bar{S})(U_t - \bar{U})}{\sqrt{\sum_{t=1}^l (S_t - \bar{S})^2} \sqrt{\sum_{t=1}^l (U_t - \bar{U})^2}} \quad (2)$$

where l is sample size and \bar{S} and \bar{U} are sample means. The larger the correlation coefficient r_{SU} is, the greater the correlation is. In this way, the sequence with a large correlation coefficient can be included in the input sample to reduce its dimensionality and make the best use of historical information.

3.2.2. Data Normalization. To prevent output saturation caused by the excessive absolute value of the net input, normalization processing is needed to cause the data vector to fall within the $[0, 1]$ or $[-1, 1]$ range before the training samples are input to the input layer of the neural network, and the antinormalization processing is carried out when the neural network outputs the data. Scale normalization can effectively distance the input and output vector data to prevent the data from being too dense. The transformation formula is as follows:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

In equation (3), x_i is the input or output data, x_{\min} is the minimum value in the data, and x_{\max} is the maximum value.

After comprehensive consideration, we formulate a four-layer BP neural network including an input layer, two hidden layers, and an output layer. The input layer is set to one neuron, the first hidden layer to 30 neurons, the second

hidden layer to 20 neurons, and the output layer to one neuron. The network structure is shown in Figure 2. Besides, the BP neural network used the standard BP algorithm, and the excitation function of nodes of the hidden layer of the BP neural network is the rectified linear units (ReLU), a piecewise linear function, which can change all negative values to zero and maintain the positive values.

4. Improved XGBoost Algorithm Using CLPSO Operator

4.1. Introduction to XGBoost Algorithm. XGBoost [31, 32] is an integrated model based on a decision tree. Its basic principles are as follows [31].

For any given training set of n samples,

$$D = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), \quad (4)$$

$(|D| = n, x_i \in R^m, y_i \in R).$

The model for the integration of the tree uses functions constructed by K superimposed tree models to predict the output. The formula is as follows:

$$\hat{y} = \sum_{k=1}^K f_k(x_i), \quad f_k \in F, \quad (5)$$

where F represents the instance space of all regression trees.

To obtain the optimal model \hat{y} through learning and consider the influence of overfitting on predictive accuracy, the following objective functions are constructed and minimized:

$$\begin{aligned} \xi &= \sum_i^n (l(y_i, \hat{y}_i)) + \sum_{k=1}^K \Omega(f_k), \Omega(f) \\ &= \alpha * N + \frac{\beta * \|w\|}{2}, \end{aligned} \quad (6)$$

where l represents a differentiable convex function used to measure the error between the predicted value \hat{y}_i and the actual value, N represents the number of nodes in the regression tree, $w = (w_1, w_2, \dots, w_N)$ represents the weight of each node in N leaf nodes, and $\Omega(f)$ is the penalty function that punishes the complexity of the model, such as the number of leaf nodes of the generated regression tree and the weight of each leaf node. The model represented by equation

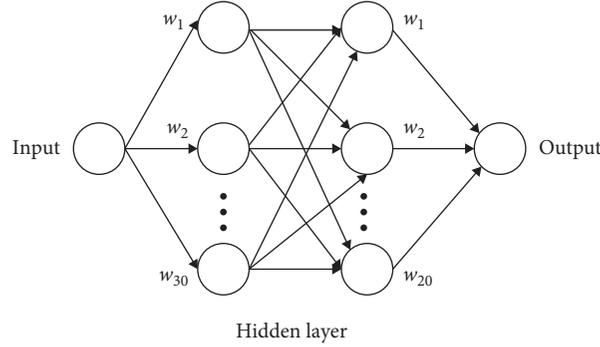


FIGURE 2: The structure of the double hidden layer neural network. °

(6) is an integrated tree model that cannot be solved in Euclidean space by traditional optimization method, so we use iterative approximation to solve it and use $y_i^{(t)}$ to represent the prediction function formed using t iterations. In the $t + 1$ -th iteration, we optimize the following equation:

$$\xi^{(t+1)} = \sum_i^n (l(y_i, y_i^{(t)}) + f_{t+1}(x_i)) + \Omega(f_{t+1}). \quad (7)$$

By optimizing the above equation, the predictions of the model are gradually improved through iteration. To obtain the optimal solution to the above equation, Taylor expansion is carried out at point $((y_1^{(t)}, y_2^{(t)}, \dots, y_n^{(t)}))$ of equation (7) to form the following expression:

$$\begin{aligned} \xi^{(t+1)} = & \sum_i^n \left(l(y_i, y_i^{(t)}) + \partial_{\hat{y}} l(y_i, \hat{y}^{(t)}) \right) * f_{t+1}(x_i) \\ & + \frac{1}{2} \partial_{\hat{y}}^2 l(y_i, \hat{y}^{(t)}) * f_{t+1}^2(x_i) + \Omega(f_{t+1}). \end{aligned} \quad (8)$$

By calculating $\partial_{\hat{y}} l(y_i, \hat{y}^{(t)})$ and $\partial_{\hat{y}}^2 l(y_i, \hat{y}^{(t)})$ and substituting them into equation (7), we can conclude, for any decision tree function $f_{t+1}(x)$, the following:

$$\xi^{(t+1)} = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} \partial_{\hat{y}} l(y_i, \hat{y}^{(t)}) \right)^2}{\sum_{i \in I_j} \partial_{\hat{y}}^2 l(y_i, \hat{y}^{(t)})} + \alpha * T, \quad (9)$$

where $I_j = \{i | q(x_i) = j\}$ represent all sample sets belonging to node j in the training set and q represents the structure of the decision tree corresponding to the decision tree function $f_{t+1}(x)$. Because different decision tree structures have different values of ξ^{t+1} , it can be used to calculate the new decision tree in the $t + 1$ -th iteration.

To better solve the problem of overfitting, XGBoost adds a regular term to the cost function that contains the number of leaf nodes of the tree and the square sum of the modules of the weight of each leaf node W . In addition, XGBoost uses column subsampling technology to reduce computation and avoid overfitting. After each iteration, XGBoost multiplies the weight of the leaf node by a coefficient to weaken the influence of each tree so that there is more learning space in later stages. The XGBoost tool supports parallelism, and one of the most time-consuming steps in learning the decision

tree is to sort the values of the features. However, before training, XGBoost sorts the data in advance and saves it as a block structure that is repeatedly used in later iterations to significantly reduce the amount of calculation. This block structure also makes parallelism possible; when splitting nodes, the gain in each feature needs to be calculated and can be obtained by parallel calculation. Finally, the feature with the largest gain is selected for splitting.

4.2. Model Building

4.2.1. Selecting Superparameters. The selection of the superparameters of XGBoost is directly related to the effect of the results of the algorithm. XGBoost needs to adjust nine superparameters and traditional method of adjusting parameters is to use the grid search method. Grid search algorithm is a method that divides the superparameters to be searched into grids in a certain space and searches for the optimal superparameters by traversing all the points in the grid. This method can find the global optimal solution when the optimization interval is large enough and the step distance is small enough. However, because the classification accuracy of most superparameter groups in the grid is very low, and the classification accuracy of superparameter groups only in a relatively small interval is very high, all parameter groups in the traversal grid are easy to fall into local optimum, and it is quite a waste of time. To solve this problem that the traditional XGBoost method is easy to fall into the local optimum and improve the calculation efficiency, we used the CLPSO operator to optimize the selection of superparameters.

4.2.2. Introduction to CLPSO Operator. The CLPSO [41, 42] can be described as a D -dimensional minimum optimization problem:

$$\min F(x), \quad x = x_1, x_2, \dots, x_D, x \in \{x_{\min}, x_{\max}\}. \quad (10)$$

CLPSO stands for ‘‘comprehensive learning particle swarm optimization,’’ which is an improved version of the traditional particle swarm optimization (PSO) algorithm. It enhances learning between particles and accelerates the convergence of the population. It solves the defect in the

original PSO algorithm whereby it can easily fall into the local optimum. Formulae to update the speed and position of the traditional PSO algorithm are as follows:

$$V_i^d = V_i^d + c_1 * \text{rand}(0, 1) * (pbest_i^d - X_i^d) + c_2 * \text{rand}(0, 1) * (gbest^d - X_i^d), \quad (11)$$

$$X_i^d = X_i^d + V_i^d, \quad (12)$$

where $pbest_i = \{pbest_i^1, pbest_i^2, \dots, pbest_i^D\}$ represents the historical optimal value of the i -th particle and $gbest = (gbest^1, gbest^2, \dots, gbest^D)$ is the global optimal value of all particles. Equations (11) and (12) show that each particle swarm learns from its historical and global optimal values in each learning process to continuously improve its fitness.

The CLPSO algorithm changes velocity equation (11) of the PSO so that the learning object can learn from different particles in different dimensions and at different times. The updated formula is as follows:

$$V_i^d = V_i^d + c * \text{rand}(0, 1) * (pbest_{f_i(d)}^d - X_i^d), \quad (13)$$

where $f_i = \{f_i(1), f_i(2), \dots, f_i(D)\}$ and $f_i(d)$ indicate that particle i needs to learn from the historical optimal value of particle number $f_i(d)$ in dimension D .

4.2.3. Optimizing XGBoost Algorithm by CLPSO Operator. The process of optimizing CLPSO consists of three steps [42]:

Step 1: set parameters of the CLPSO algorithm and initialize the particle swarm

Step 2: after initialization, update the speed and position of each particle, take the current positional information as the XGBoost superparameters, run the experiment, take the experimental result as the fitness value of the particle, and update its historical and global optimal values according to its fitness value

Step 3: repeat step 2 N times

The flowchart of this process is shown in Figure 3. There are nine superparameters of XGBoost during the experiment, which are the number of iterations ($n_estimators$), the learning rate ($learning_rate$), the maximum depth (max_depth), the sampling ratio ($subsample$ and $colsample_bytree$), the sum of sample weights of minimum leaf nodes (min_child_weight), the decline value of loss function ($gamma$), L_1 regularization (reg_alpha), and L_2 regularization (reg_lambda). Finally, we use the CLPSO algorithm and conduct three cross validations to find the optimal superparameters. The final results are shown in Table 1.

5. Results and Discussion

After analyzing the modeling, to measure the effect of filling in missing values in the data, 160 profiles were randomly selected for testing, and the rest of the profiles were used for

training. The RMSE and MAPE were used as evaluation indices. They are calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_{obs,i} - X_{model,i})^2}{N}}, \quad (14)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|X_{obs,i} - X_{model,i}|}{X_{obs,i}} \times 100\%.$$

Table 2 shows the overall error of 160 profiles randomly selected and the total convergence time for training set, and Figures 4 and 5 show the average error of the 160 profiles at each height.

Table 2 shows that the MAPE and RMSE of the improved XGBoost algorithm are about 39% less than those of the BP neural network and about 32% less than those of traditional XGBoost. The operating time of the improved XGBoost algorithm is about 99% shorter than that of BP neural network and 18% shorter than that of traditional XGBoost. The proposed algorithm not only improves accuracy but also reduces the time needed for calculation to a significant extent to improve the efficiency of calculation. From Figures 4 and 5, it is clear that most MAPE of BP neural network exceeds 1.5% and most RMSE exceeds 8 M. The MAPE of traditional XGBoost exceeds 1% at most heights and some even exceed 1.5% at both ends' height; the RMSE of traditional XGBoost is about 8 M at most heights and some even exceed 10 M at both ends' height. However, the MAPE of the improved XGBoost algorithm is less than 1% for results above 1000 m and less than 1.5% for those below 1000 m; all RMSE of the improved XGBoost algorithm are less than 8 M. The MAPE and RMSE of the proposed method are smaller than those of the BP neural network and traditional XGBoost at the same height, which further proves its advantages in filling missing values in the data. Figure 6 shows the fitting diagram of a randomly selected profile from the 160 test profiles.

From Figure 6, we can conclude that although the three prediction algorithms have the same trend as the true value, the fitting effect of the improved XGBoost algorithm at the fluctuation of the curve is in good agreement with the true value. To further measure the effect of a single profile, Figures 7 and 8 show the MAPE and RMSE at different heights of the selected profile.

Figures 7 and 8 show that the MAPE and RMSE from 0 m to 1000 m, 1500 m–2000 m and 2500 m–3000 m, when using the improved XGBoost algorithm, are much less than those of the BP neural network and traditional XGBoost. We can see that the 1500 m–2000 m is precisely the area where the profile of the modified atmospheric index fluctuated from Figure 6. According to the definition of large atmospheric ducts, this is the area where they occur.

Through the above analysis, we can conclude that the improved XGBoost algorithm is considerably superior to the BP neural network and traditional XGBoost in terms of filling in missing values of the modified atmospheric refractive index, including the profiles of the index fluctuation. According to the definition of atmospheric

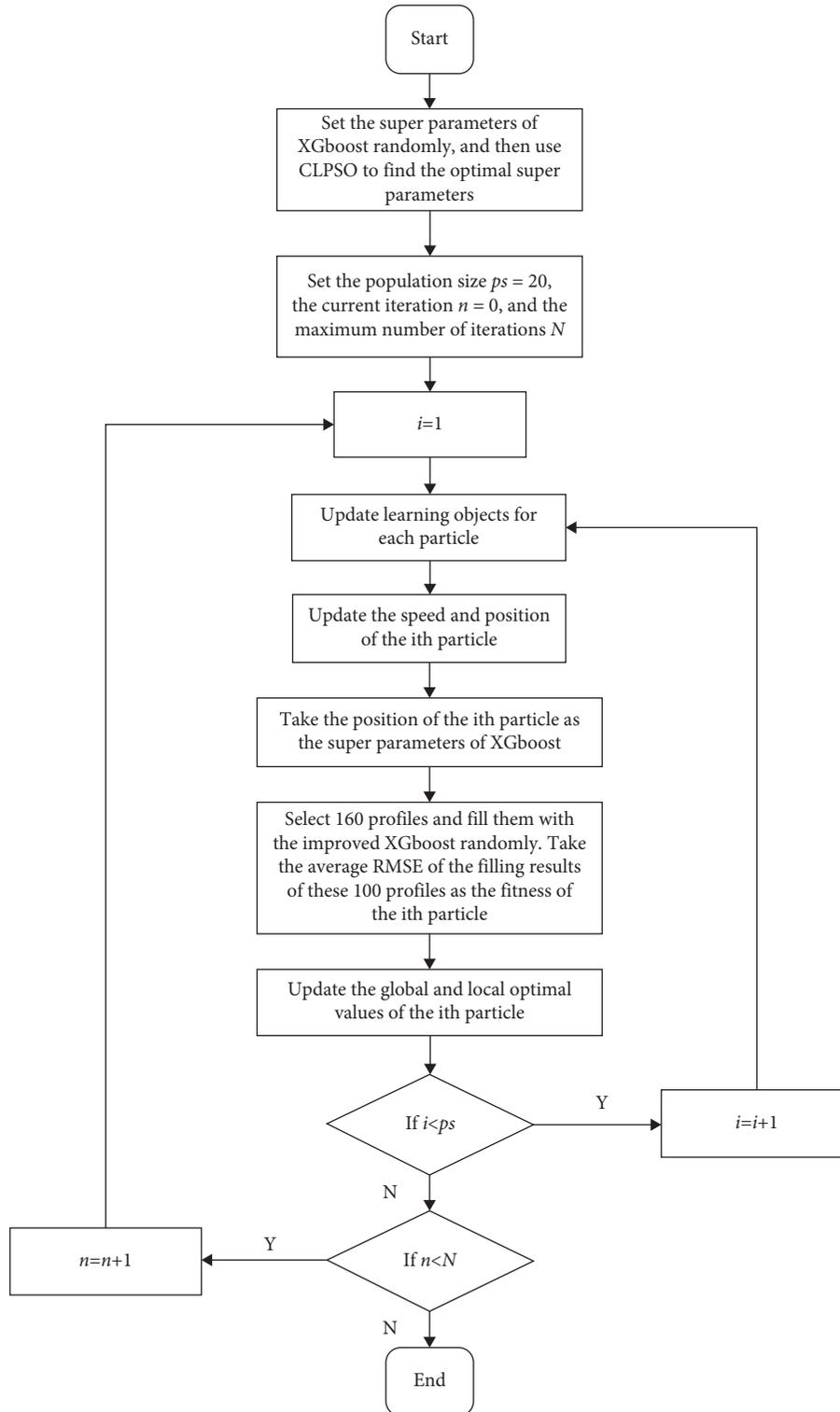


FIGURE 3: Flowchart of CLPSO optimization.

TABLE 1: The final superparameters of XGBoost algorithm.

Parameter	Value
<i>learning_rate</i>	0.112
<i>max_depth</i>	11
<i>subsample</i>	0.72
<i>min_child_weight</i>	0.55
<i>gamma</i>	0.217
<i>reg_alpha</i>	0.145
<i>reg_lambda</i>	0.5
<i>colsample_bytree</i>	0.83

TABLE 2: The total error and time of the selected test set.

	MAPE (%)	RMSE (M)	T (s)
BP neural network	1.6255	10.4606	28577.901
Traditional XGBoost	1.4288	9.6324	207.1768
Improved XGBoost	0.9757	6.4484	170.5095

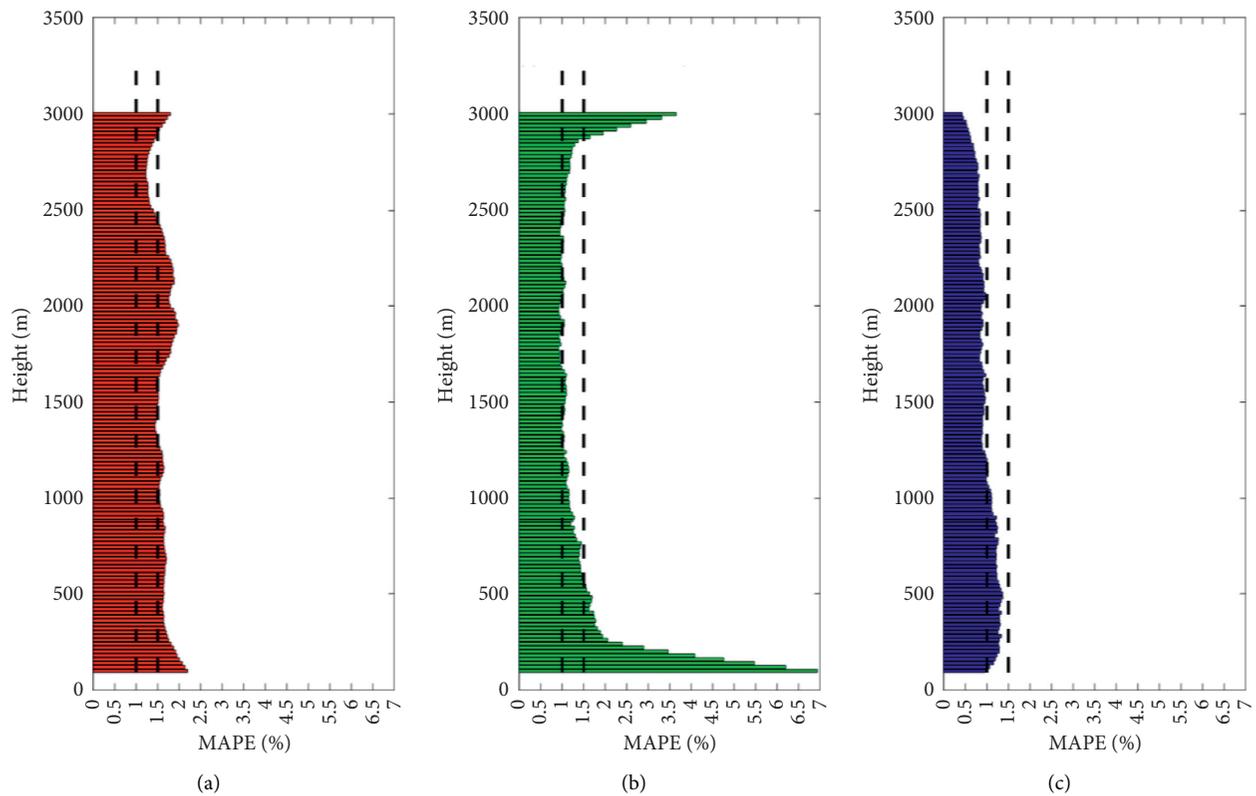


FIGURE 4: The average MAPE of the three algorithms at each height.

ducts, the region of fluctuation is where atmospheric ducts occur, so the correct judgment of modified atmospheric refractive index provides a favorable support for the

subsequent use of the modified atmospheric refractive index judgment and the analysis of the occurrence of atmospheric ducts.

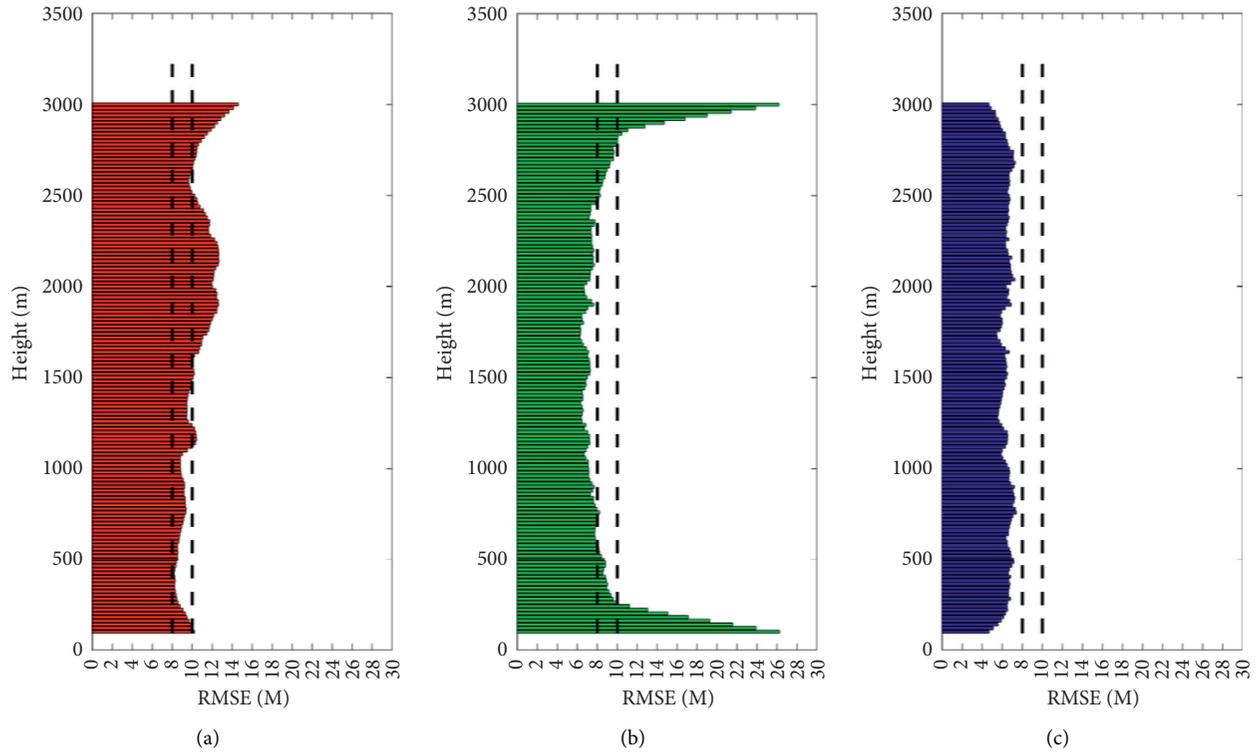


FIGURE 5: The average RMSE of the three algorithms at each height.

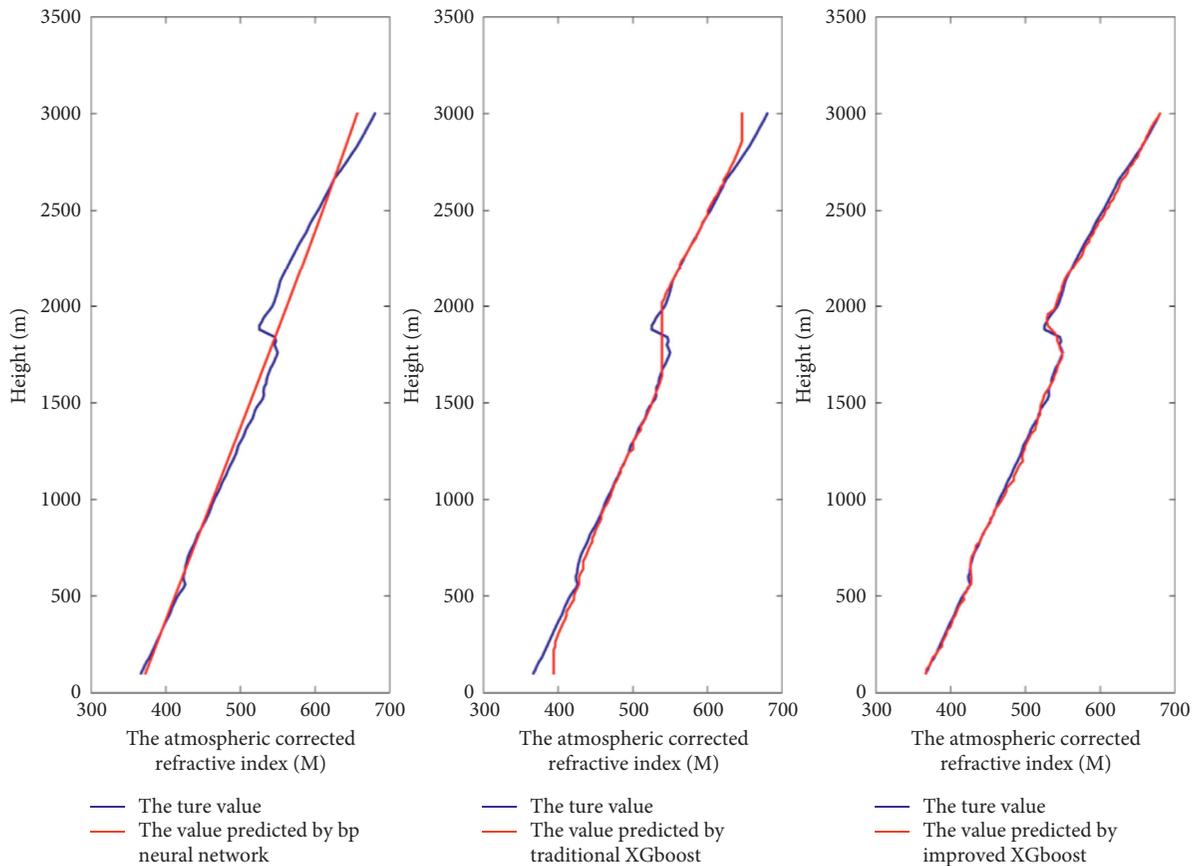


FIGURE 6: The results for a randomly selected vertical profile.

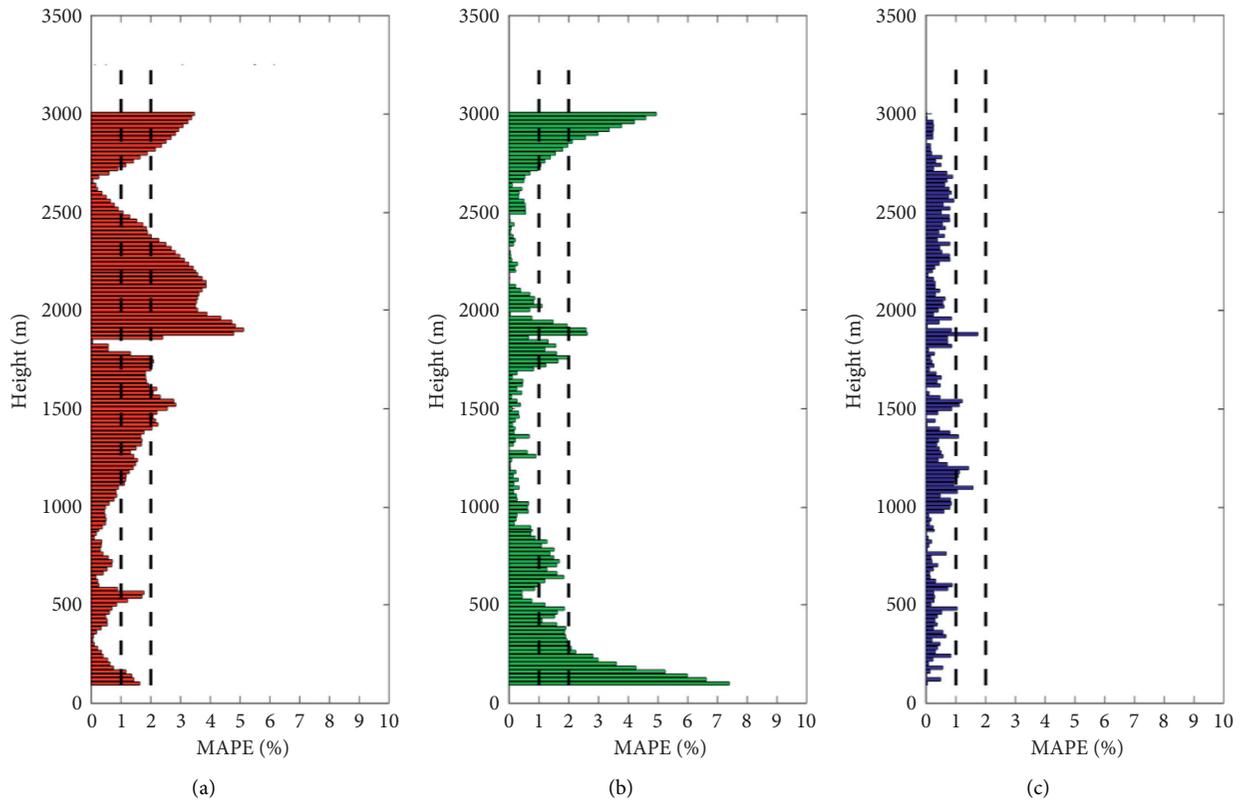


FIGURE 7: The MAPE of the selected profile.

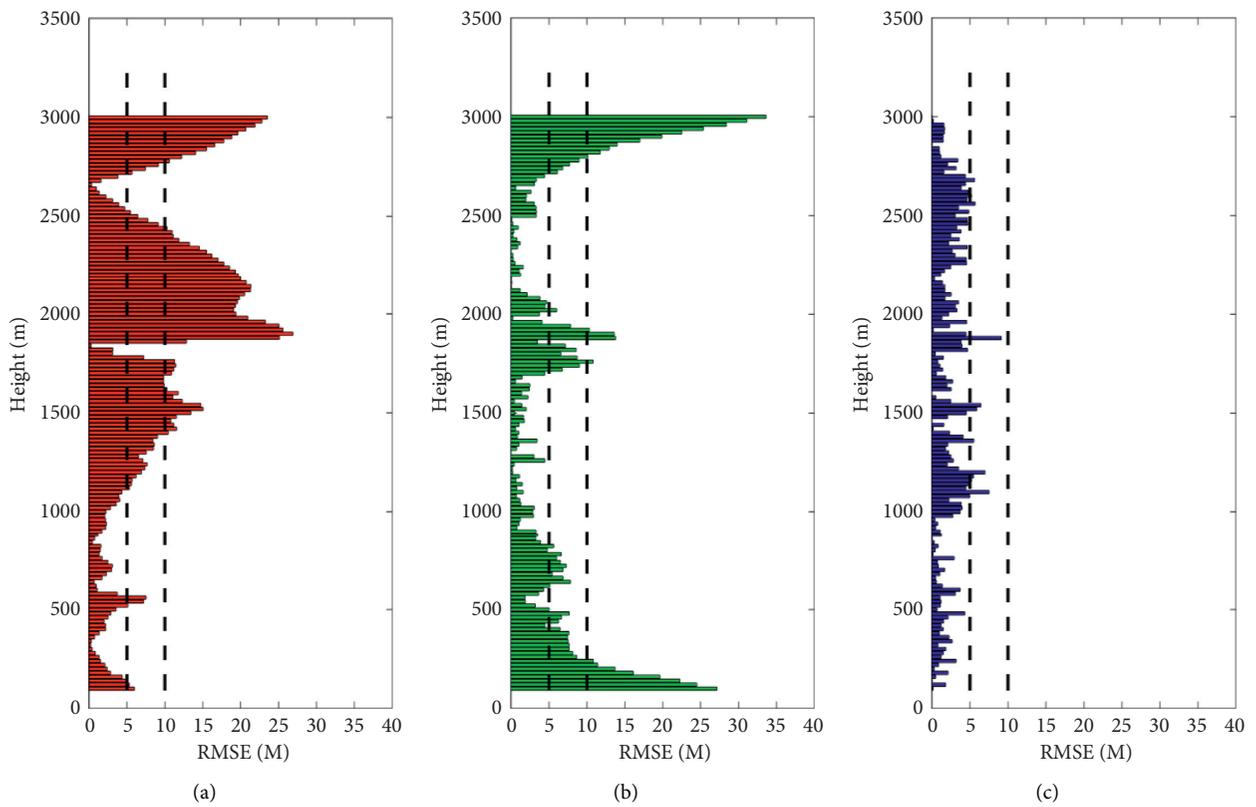


FIGURE 8: The RMSE of the selected profile.

6. Conclusions

In order to solve the problem that obtaining the atmospheric refractive index's efficiency is low in atmospheric environment, this paper proposes an improved XGBoost based on the CLPSO operator, which has high convergence speed and strong ability to jump out of the local optimum solution for parameter optimization combination; the CLPSO can not only enhance learning between particles but also accelerate the convergence of the population. To test its effectiveness, it is applied to filling missing values in the modified atmospheric refractive index over the ocean. High-resolution sounding balloon data were chosen as experimental data, and the modified atmospheric refractive index was calculated at different altitudes. The modified atmospheric refractive indices in the lower (0–100 m) and upper (3000 m–10000 m) layers were used as input, and that in the middle layer (100 m–3000 m) was used as the output. The improved XGBoost algorithm was used for learning and training to fill in the missing values of the modified refractive index of the atmosphere in the middle layer (100 m–3000 m).

The results of the proposed method were compared with those of the BP neural network and traditional XGBoost through experiment. The MAPE and RMSE of the improved XGBoost algorithm are 39% less than those of BP neural network and 32% less than those of the traditional XGBoost, and its operating time is 99% shorter than that of the BP neural network and 18% than that of the traditional XGBoost. Compared with the BP neural network and traditional XGBoost, the improved XGBoost algorithm records significant improvements in terms of both accuracy and operating time and has a significant advantage in terms of providing an accurate modified atmospheric refractive index, especially when the profile fluctuates. According to the definition of atmospheric ducts, the region of fluctuation is precisely where atmospheric ducts occur, so the correct judgment of modified atmospheric refractive index provides a favorable support for the subsequent judgment of the atmospheric ducts. The results of this paper show a new method to obtain the atmospheric refractive index over the ocean, which will be of great significance to the subsequent judgment of the atmospheric ducts over the ocean. In addition, we can also use the results to realize the beyond-the-horizon transmission and target detection of electromagnetic wave over the ocean.

However, the proposed algorithm fills in only values of the modified atmospheric refractive index from the perspective of machine learning and does not consider such physical processes as weather. In future work, we intend to consider such processes to further optimize the proposed algorithm to fill in missing values of the modified atmospheric refractive index better.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Contract no. 41875045 and Hunan Provincial Innovation Foundation for Postgraduate under Grant no. CX20200093.

References

- [1] S. Kang, Y. Zhang, and H. Wang, "The characteristics and detection of atmospheric duct," in *Atmospheric Ducts in the Troposphere*, Z. Wang, Ed., pp. 44–68, Science, Beijing, China, 3rd edition, 2014.
- [2] Y. Mai, Z. Sheng, H. Shi, Q. Liao, and W. Zhang, "Spatio-temporal distribution of atmospheric ducts in Alaska and its relationship with the arctic vortex," *International Journal of Antennas and Propagation*, vol. 2020, Article ID 9673289, 13 pages, 2020.
- [3] K. L. Ho, N. D. Mavroukoulakis, and R. S. Cole, "Determination of the atmospheric refractive index structure parameter from refractivity measurements and amplitude scintillation measurements at 36 GHz," *Journal of Atmospheric and Terrestrial Physics*, vol. 40, no. 6, pp. 745–747, 1978.
- [4] I. Sokolik, A. Andronova, and T. C. Johnson, "Complex refractive index of atmospheric dust aerosols," *Atmospheric Environment. Part A. General Topics*, vol. 27, no. 16, pp. 2495–2502, 1993.
- [5] Y. Li, Z. Gao, D. H. Lenschow, and F. Chen, "An improved approach for parameterizing surface-layer turbulent transfer coefficients in numerical models," *Boundary-Layer Meteorology*, vol. 137, no. 1, pp. 153–165, 2010.
- [6] C. Qing, X. Wu, H. Huang et al., "Estimating the surface layer refractive index structure constant over snow and sea ice using Monin-Obukhov similarity theory with a mesoscale atmospheric model," *Optics Express*, vol. 24, no. 18, pp. 20424–20436, 2016.
- [7] A. Y. Shikhovtsev, P. G. Kovadlo, L. A. Bol'basova, and V. P. Lukin, "Features of the formation of wavefront slopes on the telescope aperture at different vertical profiles of optical atmospheric turbulence," *Atmospheric and Oceanic Optics*, vol. 33, no. 2, pp. 141–145, 2020.
- [8] P. G. Kovadlo, V. P. Lukin, and A. Y. Shikhovtsev, "Development of the model of turbulent atmosphere at the large solar vacuum telescope site as applied to image adaptation," *Atmospheric and Oceanic Optics*, vol. 32, no. 2, pp. 202–206, 2019.
- [9] S. Gang, W. Ning-Quan, W. X. Li-Ming, and W. Yi, "Profile and character of atmospheric structure constant of refractive index Cn2," *Atmospheric and Oceanic Science Letters*, vol. 5, no. 3, pp. 270–272, 2012.
- [10] S. Vratolis, P. Fetfatzis, A. Argyrouli et al., "A new method to retrieve the real part of the equivalent refractive index of atmospheric aerosols," *Journal of Aerosol Science*, vol. 117, pp. 54–62, 2018.
- [11] F. Xie, J. Li, J. Zhang et al., "Variations in North Pacific sea surface temperature caused by Arctic stratospheric ozone anomalies," *Environmental Research Letters*, vol. 12, no. 11, Article ID 114023, 2017.

- [12] S. L. Odintsov, V. A. Gladkikh, A. P. Kamardin, V. P. Mamyshev, and I. V. Nevzorova, "Estimates of the refractive index and regular refraction of optical waves in the atmospheric boundary layer: Part 1, refractive index," *Atmospheric and Oceanic Optics*, vol. 31, no. 5, pp. 437–444, 2018.
- [13] J. Bian, D. Li, Z. Bai, Q. Li, D. Lyu, and X. Zhou, "Transport of Asian surface pollutants to the global stratosphere from the Tibetan Plateau region during the Asian summer monsoon," *National Science Review*, vol. 7, no. 3, pp. 516–533, 2020.
- [14] A. P. Deam, "Radiosonde for atmospheric refractive index measurements," *Review of Scientific Instruments*, vol. 33, no. 4, pp. 438–441, 1962.
- [15] B. M. Herman, R. S. Browning, and J. J. De Luisi, "Determination of the effective imaginary term of the complex refractive index of atmospheric dust by remote sensing: the diffuse-direct radiation method," *Journal of the Atmospheric Sciences*, vol. 32, no. 5, pp. 918–925, 1975.
- [16] C. V. Mathai and A. W. Harrison, "Estimation of atmospheric aerosol refractive index," *Atmospheric Environment (1967)*, vol. 14, no. 10, pp. 1131–1135, 1980.
- [17] S. Park and F. Fabry, "Estimation of near-ground propagation conditions using radar ground echo coverage," *Journal of Atmospheric and Oceanic Technology*, vol. 28, no. 2, pp. 165–180, 2011.
- [18] R. N. Lopez and V. S. del Rio, "High temporal resolution refractivity retrieval from radar phase measurements," *Remote Sensing*, vol. 10, no. 6, 2018.
- [19] G. B. Scharmer and T. I. M. van Werkhoven, "S-DIMM plus height characterization of day-time seeing using solar granulation," *Astronomy & Astrophysics*, vol. 513, 2010.
- [20] Z. Y. Wang, L. Q. Zhang, L. Kong et al., "A modified S-DIMM plus: applying additional height grids for characterizing daytime seeing profiles," *Monthly Notices of the Royal Astronomical Society*, vol. 478, no. 2, pp. 1459–1467, 2018.
- [21] A. Y. Shikhovtsev, A. V. Kiselev, P. G. Kovadlo, D. Y. Kolobov, V. P. Lukin, and V. E. Tomin, "Method for estimating the altitudes of atmospheric layers with strong turbulence," *Atmospheric and Oceanic Optics*, vol. 33, no. 3, pp. 295–301, 2020.
- [22] E. Dinar, A. A. Riziq, C. Spindler, C. Erlick, G. Kiss, and Y. Rudich, "The complex refractive index of atmospheric and model humic-like substances (HULIS) retrieved by a cavity ring down aerosol spectrometer (CRD-AS)," *Faraday Discussions*, vol. 137, pp. 279–295, 2007.
- [23] Y. Xie, X. Wu, and X. Hu, "Using one-dimensional variational assimilation algorithm to obtain atmospheric refractive index from ground-based GPS phase delay," *Geo-spatial Information Science*, vol. 43, no. 7, pp. 1042–1047, 2018.
- [24] P. Guo, Y.-H. Kuo, S. V. Sokolovskiy, and D. H. Lenschow, "Estimating atmospheric boundary layer depth using COSMIC radio occultation data," *Journal of the Atmospheric Sciences*, vol. 68, no. 8, pp. 1703–1713, 2011.
- [25] M. L. Shen, N. Keenlyside, F. Selden, W. Wiegner, and G. S. Duane, "Dynamically combining climate models to "supermodel" the tropical Pacific," *Geophysical Research Letters*, vol. 43, no. 1, pp. 359–366, 2016.
- [26] S. Yang, S. Liu, X. Li, Y. Zhong, X. He, and C. Wu, "The short-term forecasting of evaporation duct height (EDH) based on ARIMA model," *Multimedia Tools and Applications*, vol. 76, no. 23, pp. 24903–24916, 2017.
- [27] J. Rhee and J. Im, "Meteorological drought forecasting for ungauged areas based on machine learning: using long-range climate forecast and remote sensing data," *Agricultural and Forest Meteorology*, vol. 237–238, pp. 105–122, 2017.
- [28] J. Yao, S. M. Raffuse, M. Brauer et al., "Predicting the minimum height of forest fire smoke within the atmosphere using machine learning and data from the CALIPSO satellite," *Remote Sensing of Environment*, vol. 206, pp. 98–106, 2018.
- [29] I. Ebtehaj and H. Bonakdari, "A support vector regression-firefly algorithm-based model for limiting velocity prediction in sewer pipes," *Water Science and Technology*, vol. 73, no. 9, pp. 2244–2250, 2016.
- [30] A. Gholami, H. Bonakdari, A. A. Akhtari, and I. Ebtehaj, "A combination of computational fluid dynamics, artificial neural network and support vectors machines model to predict flow variables in curved channel," *Scientia Iranica*, vol. 26, no. 2, pp. 726–741, 2019.
- [31] Y.-G. Ham, J.-H. Kim, and J.-J. Luo, "Deep learning for multi-year ENSO forecasts," *Nature*, vol. 573, no. 7775, pp. 568–572, 2019.
- [32] X. Zhao, Z. Sheng, J. W. Li, H. Yu, and K. J. Wei, "Determination of the "wave turbopause" using a numerical differentiation method," *Journal of Geophysical Research: Atmospheres*, vol. 124, 2019.
- [33] S. Chang, Z. Sheng, H. Du, W. Ge, and W. Zhang, "A channel selection method for hyperspectral atmospheric infrared sounders based on layering," *Atmospheric Measurement Techniques*, vol. 13, no. 2, pp. 629–644, 2020.
- [34] P. Márquez-Neila, C. Fisher, R. Sznitman, and K. Heng, "Supervised machine learning for analysing spectra of exoplanetary atmospheres," *Nature Astronomy*, vol. 2, no. 9, pp. 719–724, 2018.
- [35] M. Hughes, J. K. Kodros, J. R. Pierce, M. West, and N. Riemer, "Machine learning to predict the global distribution of aerosol mixing state metrics," *Atmosphere*, vol. 9, no. 1, 2018.
- [36] C. Zopluoglu, "Detecting examinees with item preknowledge in large-scale testing using extreme gradient boosting (XGBoost)," *Educational and Psychological Measurement*, vol. 79, no. 5, pp. 931–961, 2019.
- [37] X. Zhang and A. I. Luo, "XGBOOST based stellar spectral classification and quantized feature," *Spectroscopy and Spectral Analysis*, vol. 39, no. 10, pp. 3292–3296, 2019.
- [38] Y. He, Z. Sheng, and M. He, "Spectral analysis of gravity waves from near space high-resolution balloon data in northwest China," *Atmosphere*, vol. 11, no. 2, p. 133, 2020.
- [39] P. Zhang, J. Liu, C. Chen, and Y. Li, "The algorithm study for using the back propagation neural network in CT image segmentation," in *International Conference on Innovative Optical Health Science*, X. Li and Q. Luo, Eds., vol. 245, 2017.
- [40] P. Singh, V. K. Singh, A. Lala, and A. K. Bhoi, "Design and analysis of microstrip antenna using multilayer feed-forward back-propagation neural network (MLPFBP-ANN)," in *Advances in Communication, Devices and Networking*, R. Bera, S. K. Sarkar, and S. Chakraborty, Eds., vol. 462pp. 393–398, 2018.
- [41] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 281–295, 2006.
- [42] Y. Cao, H. Zhang, W. Li, M. Zhou, Y. Zhang, and W. A. Chaovalitwongse, "Comprehensive learning particle swarm optimization algorithm with local search for multimodal functions," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 4, pp. 718–731, 2019.