

Research Article

A Quality Control Method Based on an Improved Random Forest Algorithm for Surface Air Temperature Observations

Xiaoling Ye,^{1,2} Xing Yang,¹ Xiong Xiong,¹ Yunpei Shen,¹ Man Hao,¹ and Rong Gu¹

¹*School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China*

²*Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China*

Correspondence should be addressed to Xing Yang; harryyangx@gmail.com

Received 14 March 2017; Accepted 6 June 2017; Published 10 July 2017

Academic Editor: Peng Yu

Copyright © 2017 Xiaoling Ye 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.

A spatial quality control method, ARF, is proposed. The ARF method incorporates the optimization ability of the artificial fish swarm algorithm and the random forest regression function to provide quality control for multiple surface air temperature stations. Surface air temperature observations were recorded at stations in mountainous and plain regions and at neighboring stations to test the performance of the method. Observations from 2005 to 2013 were used as a training set, and observations from 2014 were used as a testing set. The results indicate that the ARF method is able to identify inaccurate observations; and it has a higher rate of detection, lower rate of change for the quality control parameters, and fewer type I errors than traditional methods. Notably, the ARF method yielded low performance indexes in areas with complex terrain, where traditional methods were considerably less effective. In addition, for stations near the ocean without sufficient neighboring stations, different neighboring stations were used to test the different methods. Whereas the traditional methods were affected by station distribution, the ARF method exhibited fewer errors and higher stability. Thus, the method is able to effectively reduce the effects of geographical factors on spatial quality control.

1. Introduction

A number of large-scale climate datasets based on station observations have been developed in recent decades to study the mean state of the climate, its variability, and long-term climate trends; these datasets are also used in numerical weather prediction (NWP) [1]. Data assimilation technology has helped improve the accuracy of numerical weather prediction; and the quality control (QC) of observation data is an essential first step in data assimilation [2]. Climate datasets are particularly important for the study of temperature because changes in temperature are closely related to human activity [3]. However, the migration of observation stations and potential changes could lead to uncertainty errors in long-term daily observations [4], and such errors may reduce the quality of long-term observations and negatively influence subsequent research and the application of air temperature observations [5]. To mitigate these problems,

QC methods for surface air temperature observations are used [6, 7].

In general, QC for surface meteorological observations includes QC methods for single stations and multiple stations. These methods primarily include a plausible value check, a time consistency check, an internal consistency check, and a spatial consistency check [7–9], where the spatial consistency check is used primarily for multiple stations. For a single station, the QC method is mainly based on a meteorological element under the time series of the target stations as exemplified by Meek and Hatfield [10]. They proposed three screening rules for hourly and daily observations of QCs based on the number of meteorological elements at a single station, including high/low range limits, rate-of-change limits, and continuous no-observed-change over time limits. In addition, Ye and Xiong proposed a Gene Expression Programming (GEP) algorithm for the hourly surface air temperature of single station

observations [11, 12]. It represents a nondimensionalization treatment of relative humidity for the QC of air temperature, in which air temperature is based on the coupling relationship between air temperature and relative humidity. However, the method fails for a single station if there are a large number of missing observations. A multistation QC method can predict the value of a target station using observations from neighboring stations in a certain range to evaluate the reliability of the data [13]. Wade and Barnes proposed a QC method that uses the proportion of the inverse distances between multiple neighboring stations and the target station to calculate its weight [14, 15]. This method is called the inverse distance weighting (IDW) method, and it has a relatively stringent requirement for terrain. Subsequently, Hubbard and You proposed a spatial regression test (SRT) method that uses standard error calculation weights for multiple neighboring stations and a target station [16, 17]. The SRT presents a less stringent limitation regarding complex terrain than the IDW. However, both of these methods are affected by complex terrain and temperature fluctuations as well as by the logical relationship between temporal consistency and spatial consistency. Therefore, Wang and Liu proposed a comprehensive, consistent QC method that simultaneously applies a temporal consistency check and spatial consistency check [18]. The method effectively reduces the first type of error (type I error) and maintains a logical relationship between time and internal and spatial consistency. Xu et al. proposed a probabilistic spatiotemporal approach based on the SRT (SRT-PS), which can evaluate the uncertainty of temperature observations to a certain extent and eliminate the effect of temperature fluctuations [5].

The accumulation of a large amount of historical data reduces the reliance on traditional methods and allows for the application of more accurate and efficient methods. Therefore, the random forest method [19], which is a machine-learning algorithm that provides superior classification and regression capacities, is used in this paper, and parameter optimization is based on the artificial fish swarm algorithm (AFSA) [20]. The specific objectives are to (1) use evaluating indexes to assess the QC effects of the ARF method and traditional methods, (2) compare the effects of two error types with changes in the QC parameter between the ARF method and traditional methods, and (3) compare the QC performance of the ARF method with that of traditional methods for different time periods at different target stations.

2. Data

In this paper, 4 daily temperature observations recorded at six target stations, Chengdu (CD), Guangzhou (GZ), Lanzhou (LZ), Miyun (MY), Nanjing (NJ), and Taiyuan (TY), between 2005 and 2014 were selected for the spatial QC check. The 4 daily temperature observations corresponded to 02:00, 08:00, 14:00, and 20:00 Beijing time (CST). Target stations are selected by the number of neighboring stations, such as the IDW and SRT methods; the stations will not be selected as target stations if the number of neighboring stations is less than 5. In addition, a station which has typical characteristics

(terrain, climate, or something else) and special geographical location (capital, economic center, etc.) is considered as a target station usually. China's surface meteorological observation stations are divided into three levels, and the most important stations are national meteorological stations; therefore, the national meteorological stations are the primary choice of target stations and the 6 target stations in this manuscript are all national meteorological stations. The ARF method was applied to the 4 daily temperature observations for the six target stations; the 2005–2013 data were used as a training set, and the 2014 data were used as a test set. The data were obtained from the Chinese National Meteorological Center, and all data were tested with basic QC methods that are considered accurate.

The distribution of the six target stations and their neighboring stations is shown in Figure 1; the target stations are indicated by five-pointed stars, and the boundaries of the different provinces are indicated by the irregular lines. As shown in Figure 2, there are large differences in altitude between each of the LZ and TY target stations and their neighboring stations. In addition, the elevation data was only used to show the elevation difference between the target stations and their neighboring stations, and no elevation data were used in the experiment. Generally, there is a greater density of neighboring stations within 100 km of the target stations than within 200 km, and altitude varies least between the GZ and NJ target stations and their neighboring stations.

To test the feasibility of the method proposed in this paper, random errors were inserted into the observations of the target station by randomly adding artificial error. These errors were then compared with the predictions from the ARF method and the method of inserting artificial error proposed by Hubbard and You [21]. Approximately 3% of the observations were selected for the insertion of random errors, and the formula is as follows:

$$K_{\lambda} = s_{\lambda} p_{\lambda}, \quad (1)$$

where K is the value of the insertion error, s is the standard deviation of the 4 daily temperature observations from the target station, λ is the position for the error insertion, and p is a random number with uniform distribution $[-3.5, 3.5]$. In Hubbard research, there is a uniform distribution with a mean of zero and a range of ± 3.5 ; this number is then multiplied by the standard deviation of the variable in question to obtain the error magnitude and the selection of 3.5 is arbitrary but does serve to produce a large range of errors [16, 22].

3. Methods

3.1. Traditional Methods

3.1.1. Inverse Distance Weighting. The IDW method is a type of spatial QC method based on the distance between neighboring stations and the target station, and the key to IDW is the use of the ratio of the distance D_i (from the i th neighboring station to the target station) to the

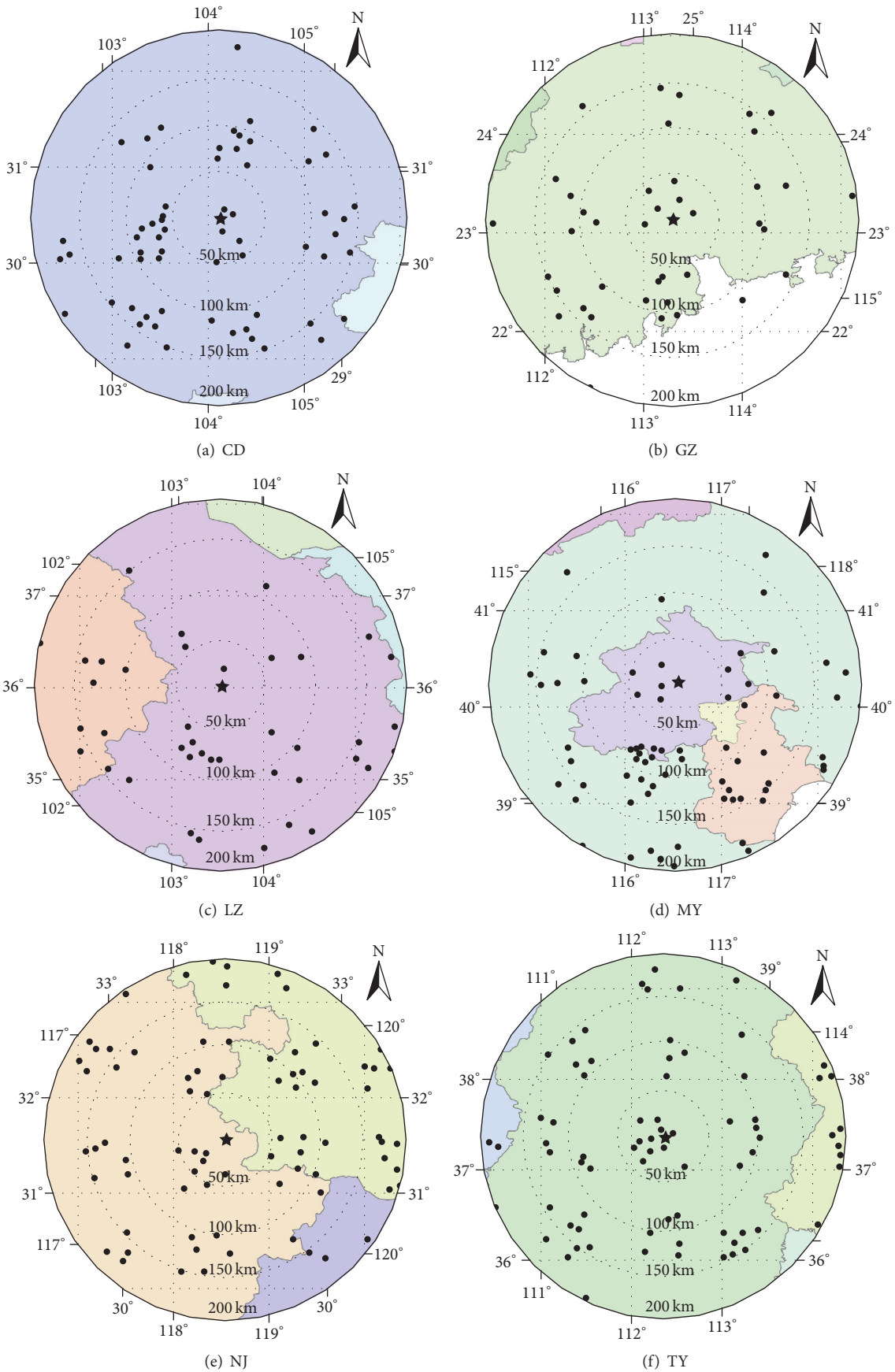


FIGURE 1: The distribution of the six target stations and their neighboring stations (within 200 km).

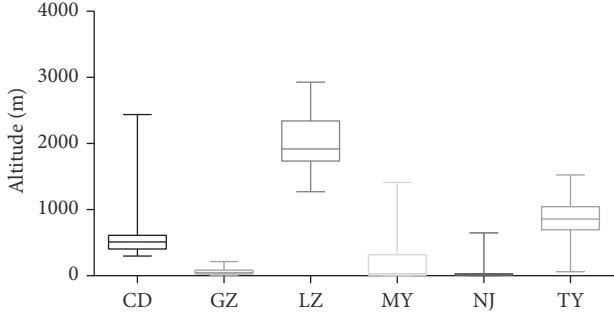


FIGURE 2: Box-and-whisker plots of altitudes for the six target stations and their neighboring stations.

whole distance D (from all neighboring stations to the target station) as the weight. The formula is as follows:

$$\hat{y}(t) = \frac{\sum_{i=1}^n D_i^{-1} y_i(t)}{D^{-1}}, \quad (2)$$

where $\hat{y}(t)$ is the predicted value for the t moment obtained by the IDW method and $y_i(t)$ is the observation value for the t moment of the i th neighboring station and the five neighboring stations that are closest to the selected target stations.

3.1.2. Spatial Regression Test. The SRT is a QC method that assigns a weight according to the standard error between the station of interest and each of the neighboring stations [21]. For each neighboring station, the linear regression is based on the following estimate:

$$x_i = a_i + b_i y_i, \quad (3)$$

where x_i is the estimate of the target station, y_i represents the data of the i th neighboring station ($i = 1, 2, \dots, n$), a_i and b_i are coefficients, and the weighted estimate x' is obtained using the standard error of estimate s :

$$x' = \sqrt{\frac{\sum_{i=1}^n x_i^2 s_i^{-2}}{\sum_{i=1}^n s_i^2}}. \quad (4)$$

The weighted standard error of estimate s' is calculated as follows:

$$s'^{-2} = N^{-1} \sum_{i=1}^n s_i^{-2}. \quad (5)$$

According to the above formulas, 10 stations with the smallest standard error relative to the target station were selected as neighboring stations.

3.2. The ARF method

3.2.1. Random Forest Algorithm. The random forest (RF) algorithm belongs to the category of ensemble learning. Schapire developed the probably approximately correct (PAC) learning model, which evaluates strong and weak

learning concepts [23]. The RF algorithm is the combination of multiple weak-classifier decision trees into a strong classifier, which is much easier than searching for a strong classifier directly and explains why the RF algorithm is widely used.

3.2.2. Artificial Fish Swarm Algorithm. The AFSA is a type of optimization algorithm inspired by the phenomenon of predation on biological clusters. The fitness function represents "food," the current status of virtual artificial fish is $M = (m_1, m_2, \dots, m_n)$, the position status of the viewpoint at a certain time is $M_v = (m_1^v, m_2^v, \dots, m_n^v)$, the Rand function generates a random number between 0 and 1, the Step function is the step length, the Visual function is the scope of vision, and the optimization process is derived from the following formulas:

$$m_i^v = m_i + \text{Visual} \cdot \text{Rand}(), \quad i = n \cdot \text{Rand}(),$$

$$M_{\text{next}} = \frac{M_v - M}{\|M_v - M\|} \times \text{Step} \times \text{Rand}(). \quad (6)$$

In this paper, the out-of-bag (OOB) error is the fitness function of the model; the OOB error is explored under different mtry values using the R platform, where mtry values indicate the number of variables randomly sampled as candidates at each split, and the ability of the permutation importance to detect influential predictor variables in sets of correlated covariates highly depends on mtry [24]. And the influence of mtry values on the model construction time is also considered. Subsequently, the most suitable mtry value is selected, and the results are rapidly derived with minimal error.

3.2.3. QC Method Using the AFSA and RF Algorithms. The QC model was constructed using the improved RF algorithm (ARF), and neighboring stations were selected within 200 km of the target station. The design of the ARF model is shown in Figure 3. Random forest is an ensemble classifier based on bagging integration, and each subset is obtained by bootstrap random sampling of the original dataset. The data that are not collected are called OOB data. The fitness function of a given mtry value and OOB error is constructed, the AFSA is used to find the optimal mtry value, and the QC model is constructed with the optimal mtry value. The QC model is used to predict the target stations of the temperature observations, and the predicted values and observed values are tested via the threshold test. The predicted values of the model were determined as follows (7):

$$|y_{\text{obs}} - y_{\text{est}}| \leq f \cdot \delta, \quad (7)$$

where y_{obs} is the observational value of the target station, which is inserted into the random error; y_{est} is obtained by the improved RF regression model to obtain the estimate of the target station; f is the QC parameter; and δ is the standard error between the observed value and the predicted value of the target station. If formula (7) holds true, then the value is correct; if it does not, then the data are recorded as suspicious, thereby accomplishing data QC.

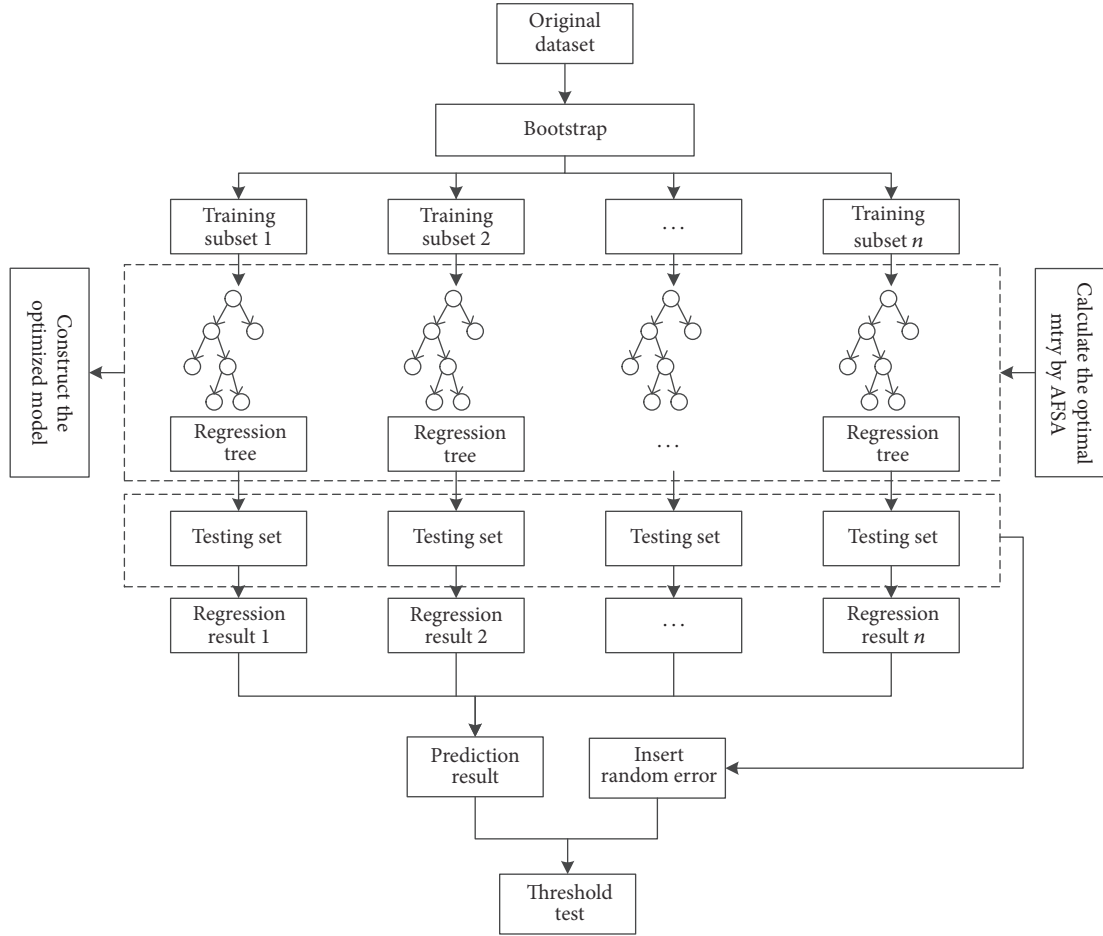


FIGURE 3: Flowchart of the QC modeling procedure ARF.

3.3. *Evaluation of Model Performance.* In this paper, the root mean square error (RMSE), the mean absolute error (MAE), and the Nash-Sutcliffe model efficiency coefficient (NSC) [25] were used to evaluate the QC model. Willmott recommends describing the average difference by the RMSE or MAE because the RMSE and MAE are among the “best” overall measures of model performance [26]. These indexes take the following form:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_{\text{obs}} - y_{\text{est}})^2}{n}},$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_{\text{est}} - y_{\text{obs}}|, \quad (8)$$

$$\text{NSC} = 1 - \frac{\sum_{i=1}^n (y_{\text{obs}} - y_{\text{est}})^2}{\sum_{i=1}^n (y_{\text{obs}} - \bar{y})^2}.$$

4. Results and Discussion

4.1. *Spatial Correlation Analysis.* This paper selected six target stations in different regions. As a regional variable, surface temperature has obvious correlation in the spatial

domain [27, 28]. Tobler considers this correlation to be a function of distance: the closer the regionalized variable, the stronger the correlation (Tobler’s First Law of Geography) [29]. The spatial correlations of these regions, as calculated by Moran’s I [30], are shown in Figure 4. The selection range of the regions is a circle with a 200 km radius and the target station at the center. The closer Moran’s I value is to 1, the higher the spatial correlation of the stations is. All regions have high Moran’s I values, which indicates high spatial correlation, except for the LZ region. The lower value for this region is due to the variation in its altitude, as illustrated in Figure 2.

4.2. *Spatial Sensitivity Analysis.* To analyze the effects of different methods on the density of stations, the neighboring stations within a radius of 20 to 200 km were selected. The results obtained by the IDW, SRT, and ARF methods are shown in Figure 5. The figure shows that the performance indexes for the ARF method are superior to those of the IDW and SRT methods. The IDW, SRT, and ARF methods are affected by the radius; at a low radius, the ARF method yields more accurate performance indexes. At a radius greater than 140 km, the performance indexes of the three methods shift slightly compared to the values identified at closer distances,

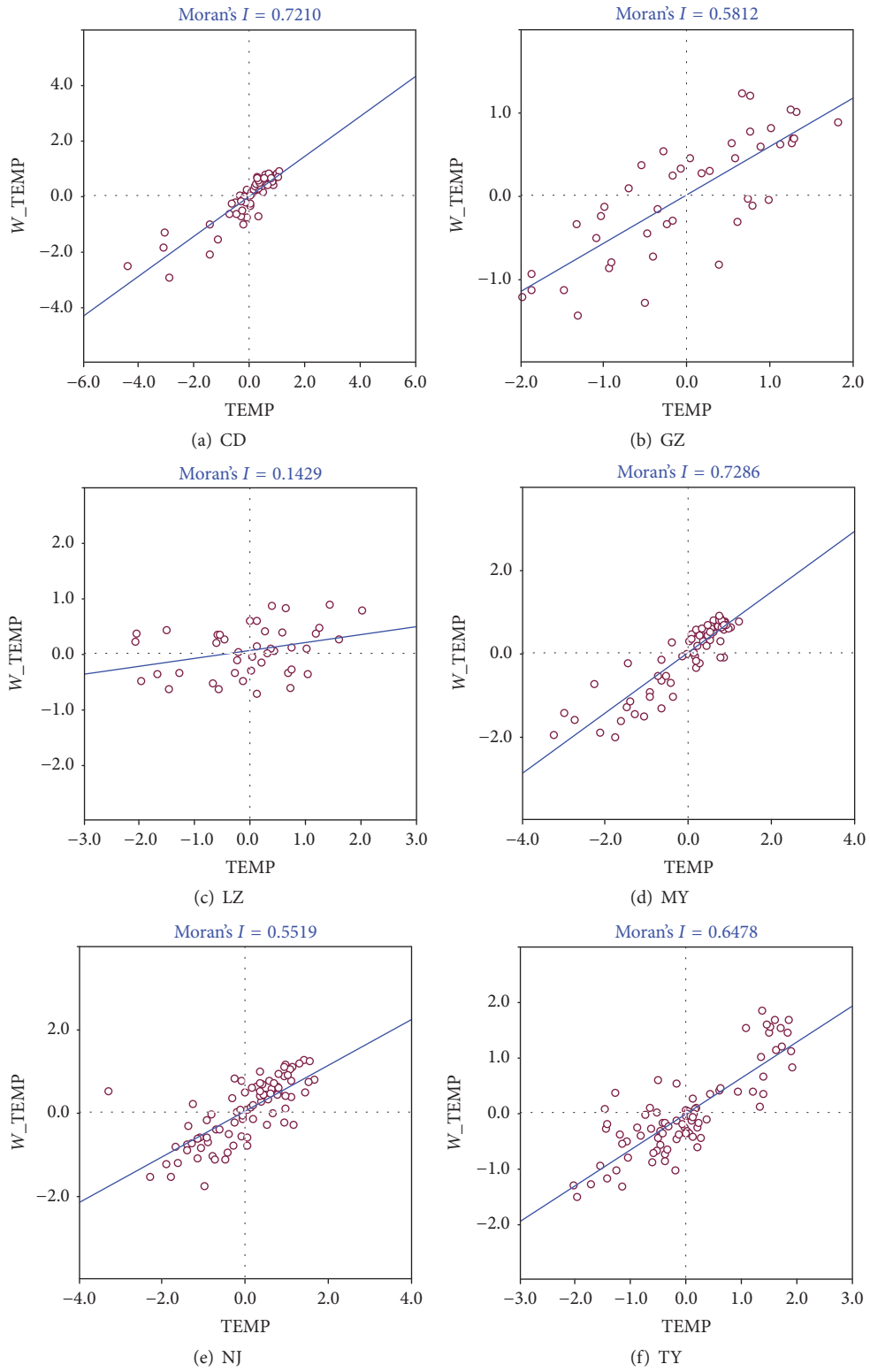


FIGURE 4: Moran's I values of six regions at some point.

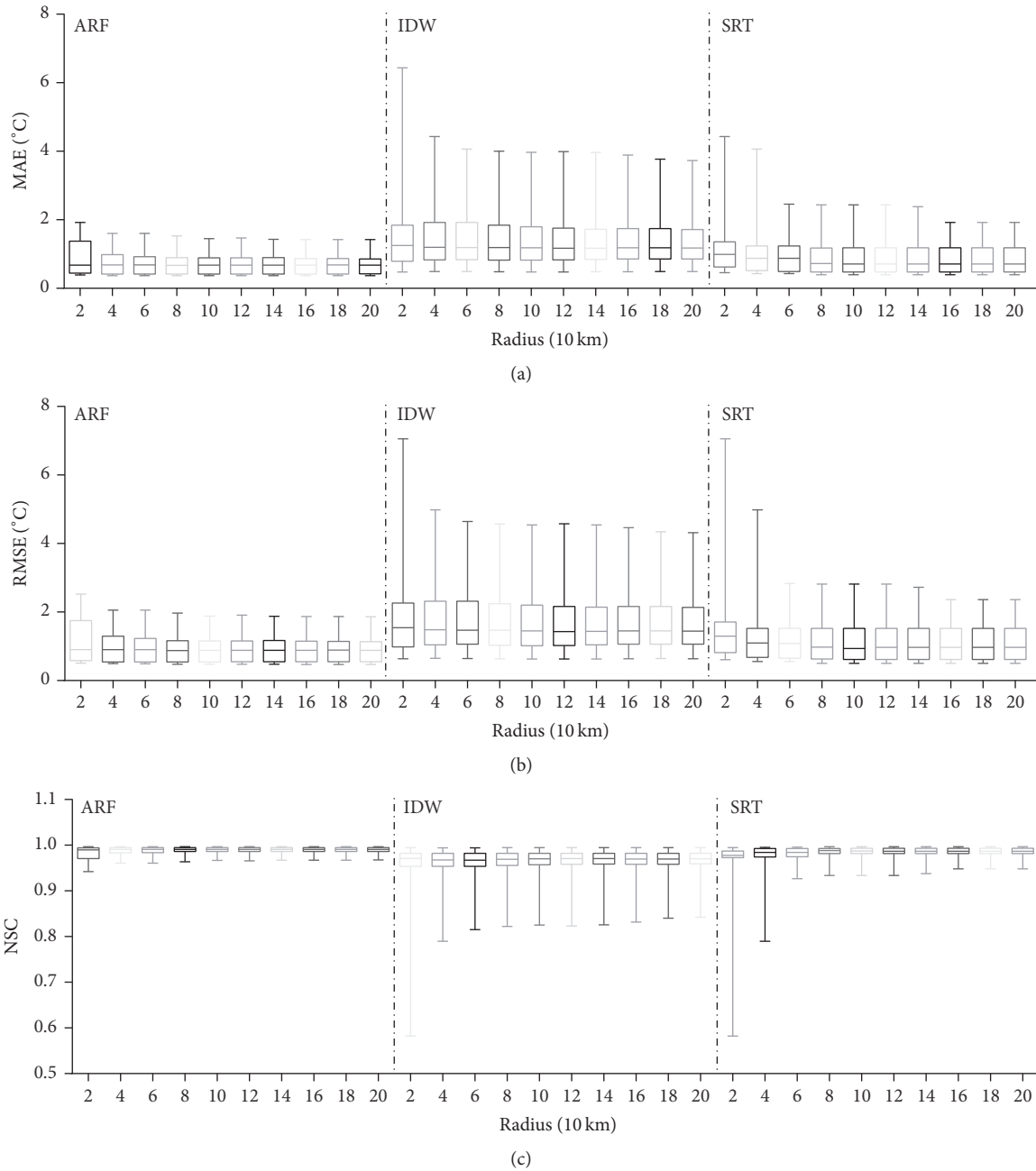


FIGURE 5: Performance indexes of different radii for six target stations at different times for the IDW, SRT, and ARF methods: (a) MAE, (b) RMSE, and (c) NSC.

which may be due to the density of stations. Small changes in the number of neighboring stations result in small changes in the performance indexes of the three methods when the radius is greater than 140 km. Therefore, it was necessary to study the performance of the IDW, SRT, and ARF methods at different target stations at the same time and with the same number of neighboring stations.

To evaluate the influence of the number of neighboring stations on the IDW, SRT, and ARF methods, the performances of the three methods at the same number of neighboring stations within 200 km of each target station

were compared. More than 10 stations were selected as neighboring stations for the IDW, SRT, and ARF methods in this paper. For the IDW method, neighboring stations were sorted by the shortest distance; and for the SRT method, stations were sorted by the smallest RMSE. The performance indexes of the IDW, SRT, and ARF methods are shown in Figure 6. As shown in Figure 6, the performance indexes of the ARF method are superior to those of the IDW and SRT methods. With an increasing number of neighboring stations, the performance of the ARF improves, whereas the performances of the IDW and SRT methods remain poor.

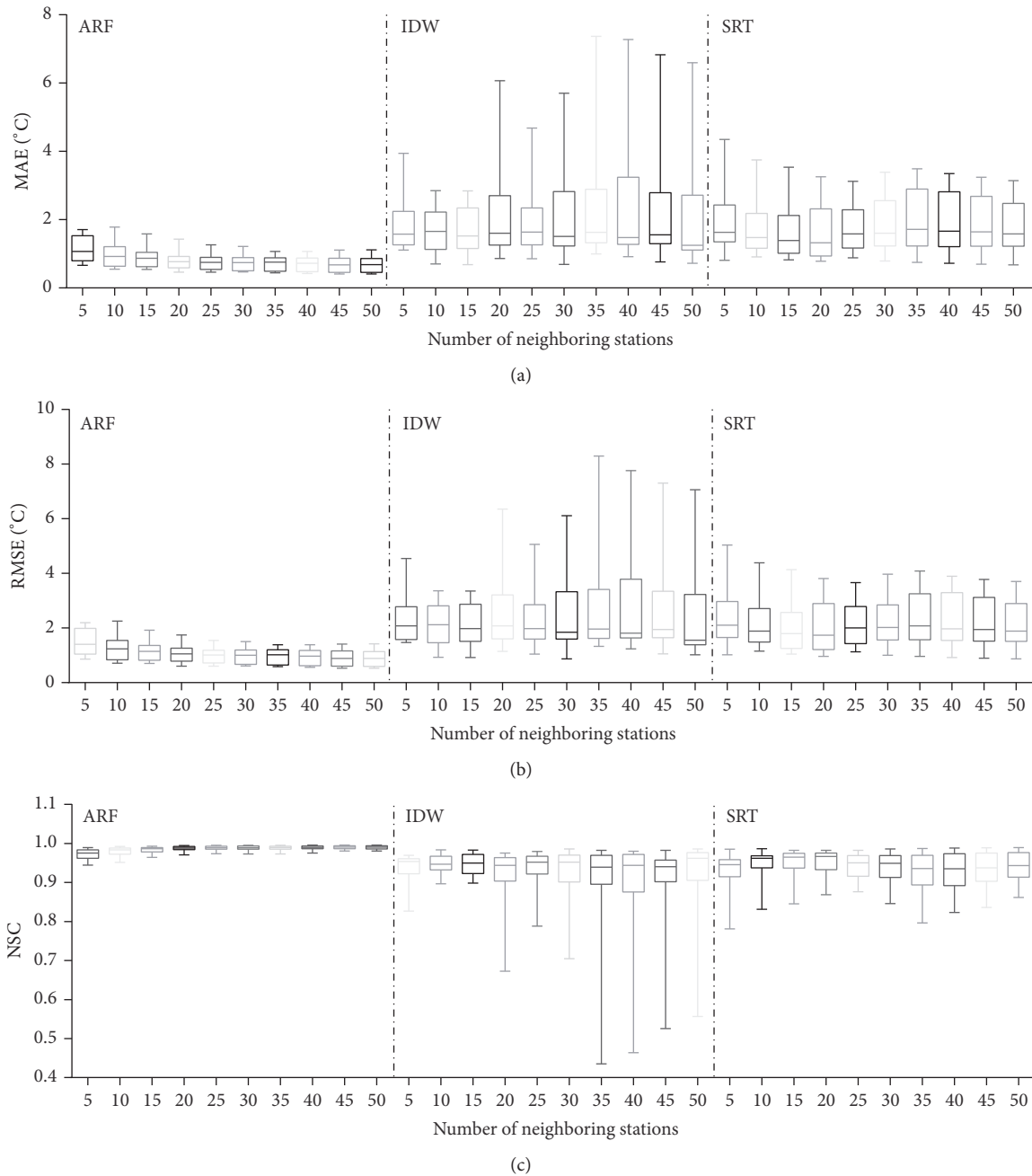


FIGURE 6: Performance indexes for different numbers of neighboring stations at different times for the IDW, SRT, and ARF methods: (a) MAE, (b) RMSE, and (c) NSC.

The performances of the IDW and SRT methods are best when 10–15 neighboring stations are used. These results are consistent with the conclusions of the two methods and demonstrate that the performance indexes used in this paper can effectively evaluate the model.

4.3. Evaluation of Model Performance

4.3.1. Evaluation of the Index Analysis. This experiment was primarily performed to compare the effect of spatial QC among the ARF method and the traditional methods (IDW

and SRT) for different types of target stations. Figure 7 shows a comparison of the performance indexes (MAE and RMSE) for the ARF, IDW, and SRT methods. The MAE and RMSE values at different target stations show that the ARF method was superior to the traditional methods, especially for LZ station. This station is surrounded by mountains related to the impact of the Gannan Plateau uplift. Thus, the climate variability at LZ station and its neighboring stations is complex, and performing spatial QC for these stations is difficult.

Compared with the IDW and SRT methods, the ARF method provided an advantage when analyzing complex

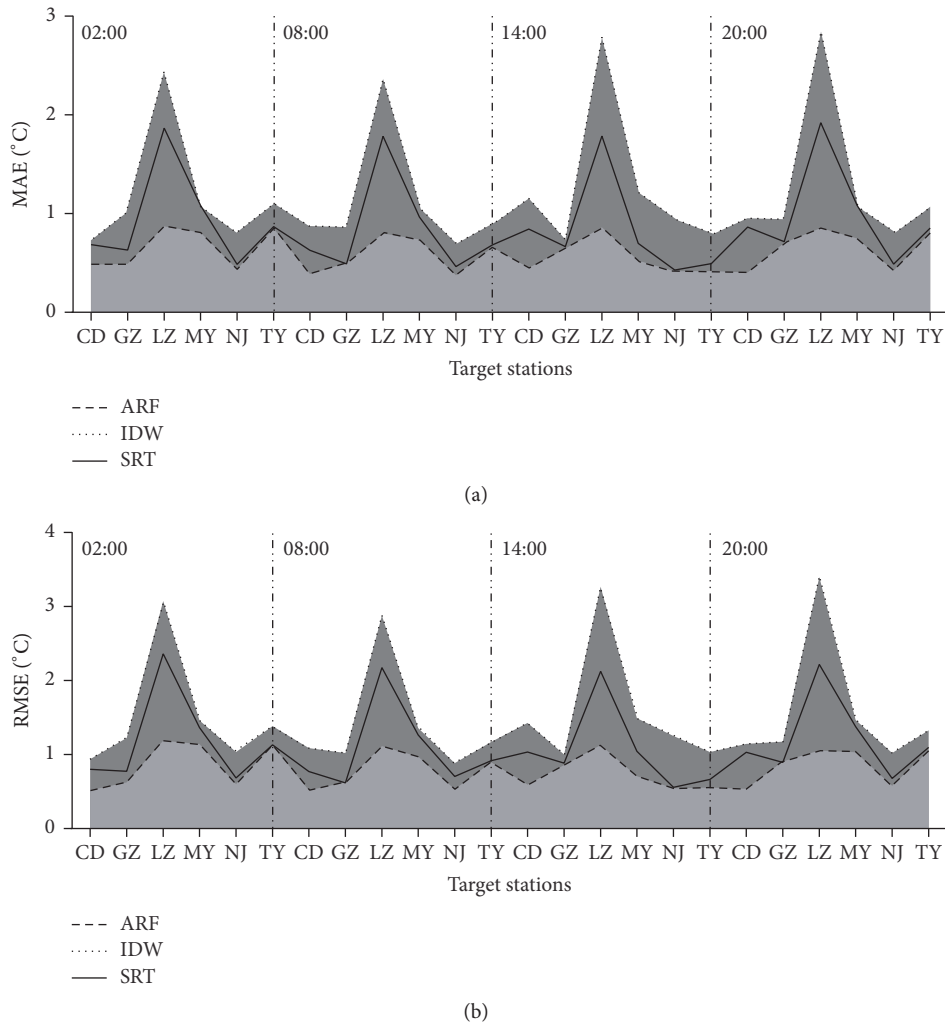


FIGURE 7: Comparisons of performance of the ARF, IDW, and SRT methods for different target stations: (a) the contrast condition of MAE at 02:00, 08:00, 14:00, and 20:00; (b) the contrast condition of RMSE at 02:00, 08:00, 14:00, and 20:00.

terrain and areas with variable climate (Figure 7). For coastal areas with relatively flat terrain, such as GZ and NJ, the deviation caused by regional differences was reduced, and the MAE and RMSE of the three methods were not very different. Overall, the MAE and RMSE of the ARF method were smaller and more stable than those of the IDW and SRT methods, and the ARF method’s predictive effect was superior. Figure 8 illustrates how the AFSA algorithm can be used to optimize the results of the RF algorithm. Although the AFSA optimization increased the time required for the RF method to generate a prediction, the error rate of the RF method decreased. Overall, the ARF method’s performance indexes were smaller and more stable.

4.3.2. *Statistical Analysis.* Statistical tests of a proposed hypothesis are generally subject to two types of errors: “type I” errors, which are produced when the correct data are treated as the wrong data, and “type II” errors, which are produced when the wrong data are treated as the correct data [31]. In the QC of surface meteorological observations,

if type I errors are large and a considerable amount of correct data is treated as the wrong data, then the detection rate may be increased, but the integrity of the original data is compromised. However, if type II errors are large and a considerable amount of wrong data that should have been detected is “passed,” then the QC meaning is obscured. A method of controlling type I errors and reducing type II errors is adopted to achieve better QC results. In this paper, the intersections between type I errors and type II errors are treated as the value of the QC parameter f .

For example, Figure 9 shows the results of type I errors and type II errors for the CD station at 02:00, 08:00, 14:00, and 20:00; and ((a)–(c), (d)–(f), (g)–(i), and (j)–(l)) in the figures present the conditions of the ARF, IDW, and SRT methods for the two types of errors at 02:00, 08:00, 14:00, and 20:00. The graphs of the two types of errors for the other target stations are omitted. The QC parameters were obtained from the intersection of the two types of errors in Figure 9, and the results are shown in Table 1.

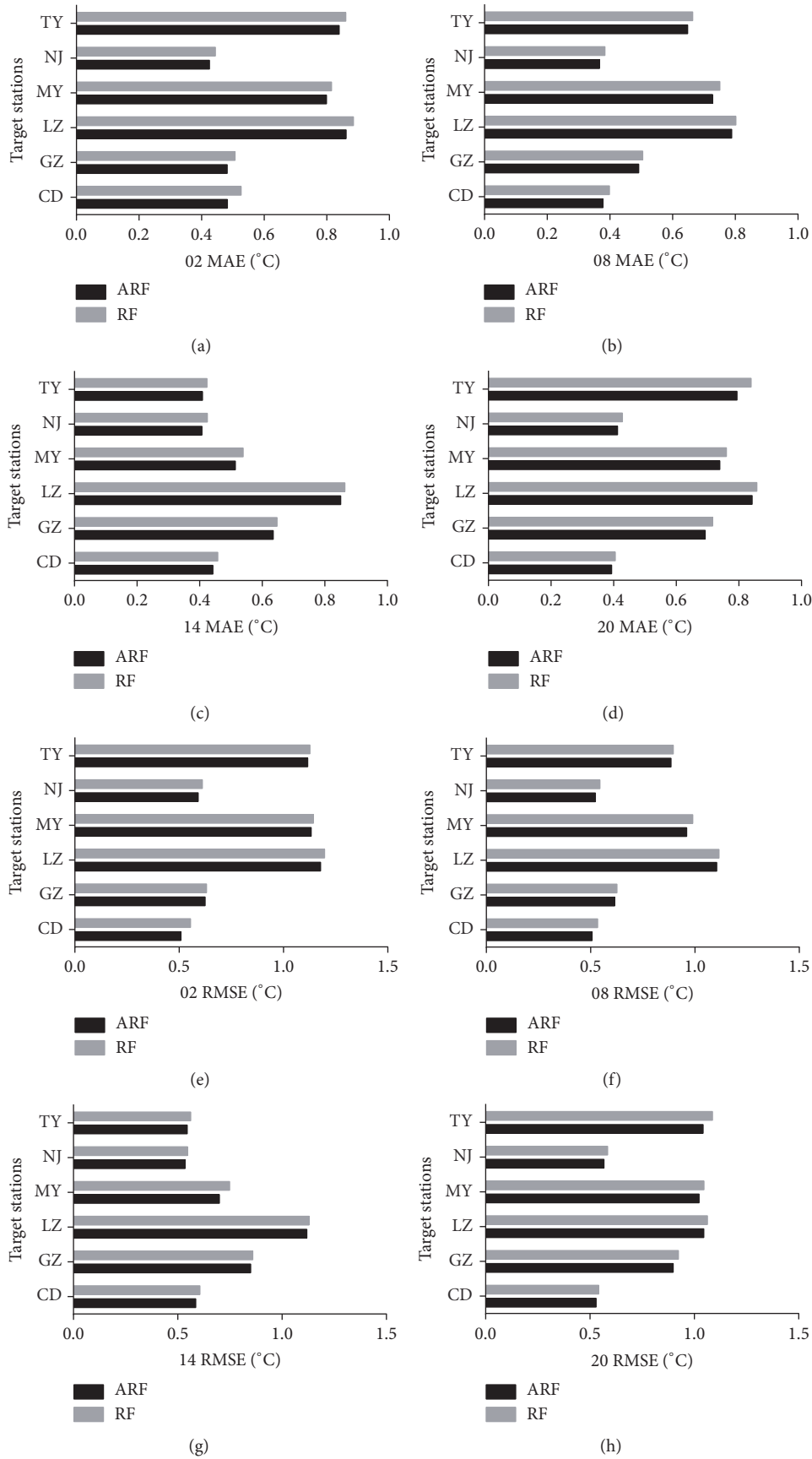


FIGURE 8: Comparisons of performance of the ARF and RF methods for different target stations: (a)–(d) the contrast condition of MAE at 02:00, 08:00, 14:00, and 20:00; (e)–(h) the contrast condition of RMSE at 02:00, 08:00, 14:00, and 20:00.

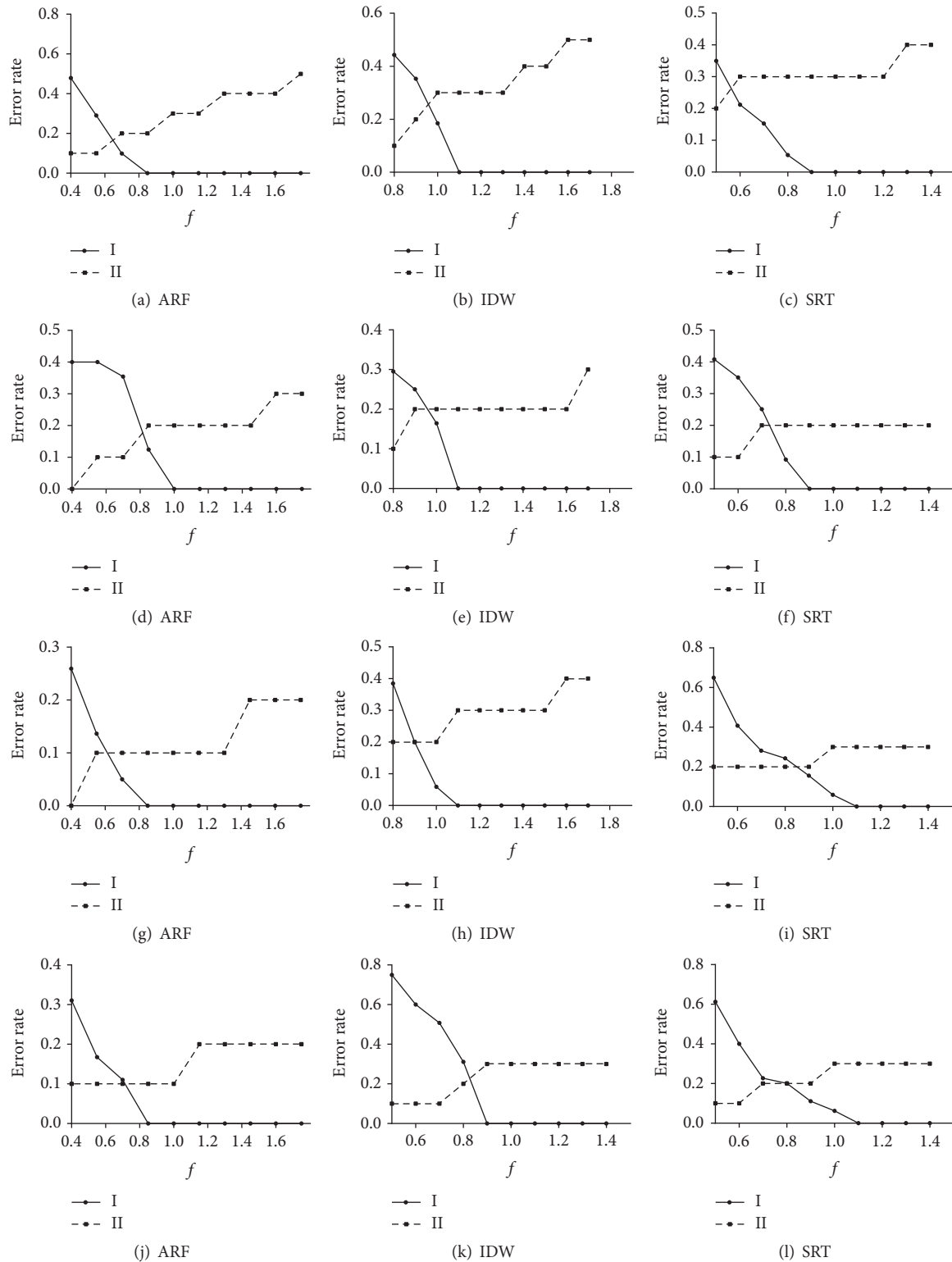


FIGURE 9: Comparisons of two types of errors for CD at 02:00 (a)–(c), 08:00 (d)–(f), 14:00 (g)–(i), and 20:00 (j)–(l).

According to formula (7), to evaluate the random insertion of an error value in the data, the detection rate is considered as equal to the magnitude of the number of detected errors divided by the total amount of error data.

The effects of different QC methods can be assessed through the comparison of the detection rate under different QC methods. As shown in Table 2, because fewer artificial errors were inserted, the detection rate gap is not obvious; however,

TABLE 1: QC parameters for CD, GZ, LZ, MY, NJ, and TY stations at 02:00, 08:00, 14:00, and 20:00.

	02:00			08:00			14:00			20:00		
	ARF	IDW	SRT	ARF	IDW	SRT	ARF	IDW	SRT	ARF	IDW	SRT
CD	0.64	0.97	0.56	0.80	0.96	0.72	0.60	0.90	0.85	0.70	0.80	0.80
GZ	0.53	0.76	0.60	0.71	0.75	0.57	0.80	0.69	0.78	0.76	0.75	0.80
LZ	0.70	1.42	1.27	0.77	1.11	0.95	0.93	1.31	1.00	0.57	1.12	0.76
MY	0.95	0.88	0.80	0.56	0.65	0.60	0.51	0.80	0.70	0.80	0.92	0.75
NJ	0.60	0.80	0.63	0.50	0.58	0.63	0.43	0.90	0.50	0.55	0.63	0.61
TY	0.80	0.90	0.80	0.62	0.80	0.70	0.47	0.66	0.65	0.63	0.72	0.62

TABLE 2: The detection rate under optimal QC parameters for CD, GZ, LZ, MY, NJ, and TY at 02:00, 08:00, 14:00, and 20:00.

	02:00			08:00			14:00			20:00		
	ARF	IDW	SRT	ARF	IDW	SRT	ARF	IDW	SRT	ARF	IDW	SRT
CD	83%	72%	74%	82%	80%	80%	90%	80%	80%	90%	78%	80%
GZ	90%	80%	90%	90%	80%	90%	90%	80%	90%	90%	80%	90%
LZ	90%	71%	74%	86%	80%	80%	90%	75%	76%	90%	80%	80%
MY	90%	80%	90%	90%	85%	90%	92%	90%	90%	83%	71%	85%
NJ	90%	80%	86%	90%	82%	85%	90%	80%	90%	90%	86%	88%
TY	90%	90%	90%	90%	90%	90%	90%	87%	80%	82%	80%	80%

the ARF method had a generally higher detection rate than did the IDW and SRT methods.

In addition, daily temperature observations from 634 stations in 10 different regions (Chengdu (CD), Guangzhou (GZ), Lanzhou (LZ), Miyun (MY), Nanjing (NJ), Taiyuan (TY), Beihai (BH), Haikou (HK), Urumqi (Um), and Changchun (CC)) between 2005 and 2014 were selected to evaluate the stability of the ARF method at different stations. As shown in Figure 10, although the performance indexes for the ARF method are greater than those for the IDW and SRT methods for some stations in BH and CC, overall, the ARF method is superior to the IDW and SRT methods, especially in LZ, MY, and UM, and the performance indexes for the ARF method are more stable than those for the IDW and SRT methods.

4.4. Edge Stations Analysis. Ideal neighboring stations, such as the 6 target stations chosen in this paper, were not available for all stations. For a station that has only a few stations in neighboring or edge areas, whether the traditional QC methods can achieve the original performance is unclear; therefore, the QC performance for the ARF method must be tested within this context. The distribution of stations in GZ and MY is shown in Figure 11. The circles represent the stations with numerous neighboring stations nearby, and the pentagons represent the edge stations that are near the sea and lack neighboring stations. Because of the special geographical location of the edge region, the IDW method for calculating the weights by distance is not applicable; therefore, only the results of the ARF and SRT methods are compared.

The performance indexes for the ARF and SRT methods for edge stations are shown in Figure 12, and the names of the stations are replaced by numbers 1 to 7 in GZ and 1–11 in MY. The letters (a) and (b) denote the performance indexes

for the MAE and RMSE and detection rate for the edge stations. The results indicate that the ARF method includes obviously smaller errors than the SRT method. In Figure 7, the ARF method exhibits nearly the same MAE and RMSE as the SRT method for GZ at different times; however, for the edge stations, the ARF method is clearly superior. Therefore, compared with the SRT method, the ARF method has less impact on the geographical edge of stations and is more conducive to eliminating the adverse effects of terrain.

5. Conclusions

The ARF method for the spatial QC of surface temperature observations was introduced in this paper. The results show that the ARF method is superior to traditional methods, especially at target stations with complex terrain and climate. Edge stations without sufficient neighboring stations were selected as target stations, and the ARF method performed better than the traditional methods for these stations. Moreover, the results show that the ARF method adapted better to different terrain and station distributions; thus, the superiority of the ARF QC method has been fully demonstrated.

The spatial correlation analysis shows that the ARF, IDW, and SRT methods are affected by spatial correlation: where spatial correlation is high, the prediction effect of the three methods is very good, whereas in regions with low spatial correlation, such as LZ, the prediction effect of the three methods is not good. However, the impact of spatial correlation is much smaller on the ARF method than on the IDW and SRT methods. With increases in the number of neighboring stations or in the radius of the surrounding area of neighboring stations, the results of the ARF method and the gain effect are consistently superior to the IDW and SRT methods in the spatial sensitivity analysis. The results

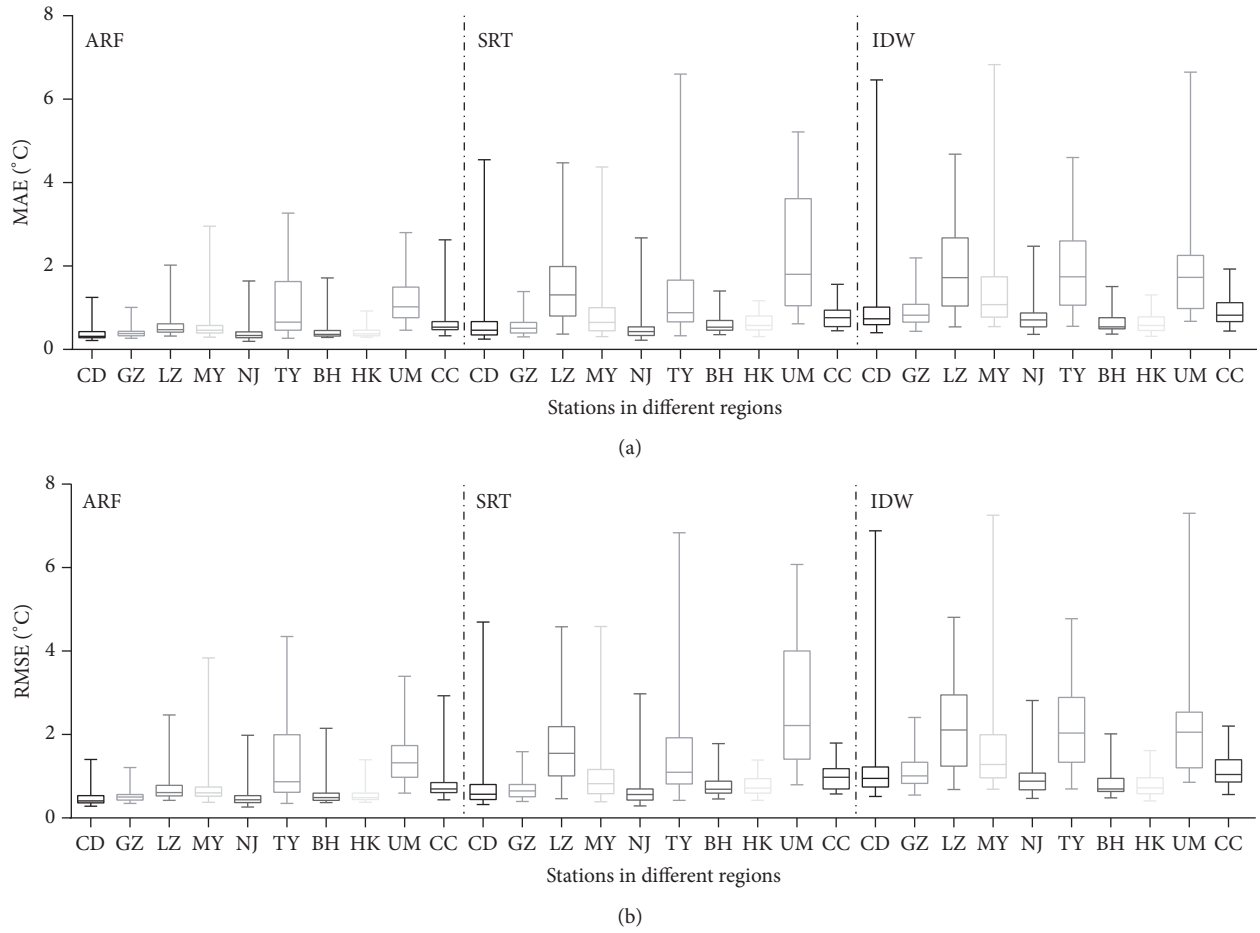


FIGURE 10: Performance comparisons for the ARF, IDW, and SRT methods for stations in different regions: (a) the contrast condition of MAE; (b) the contrast condition of RMSE.

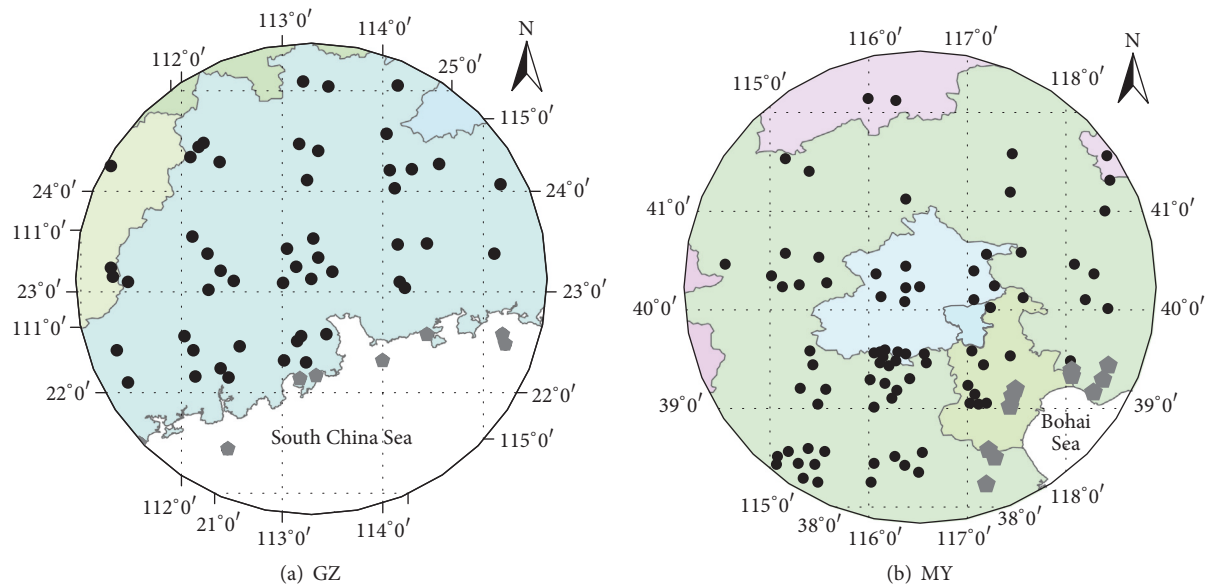


FIGURE 11: The distribution of stations in (a) Guangzhou and (b) Miyun provinces.

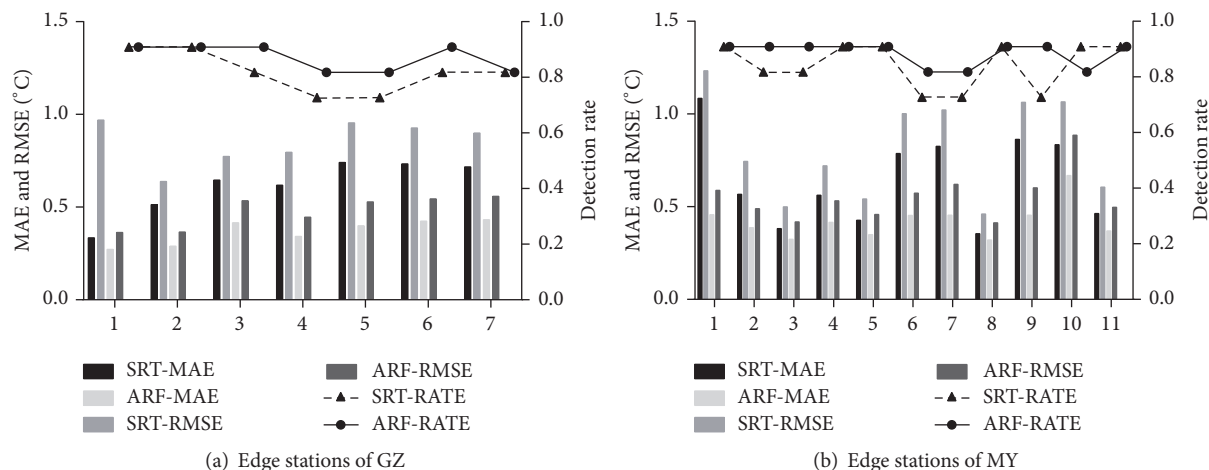


FIGURE 12: Comparisons of performance between the ARF and SRT methods for different edge stations: (a) Guangzhou and (b) Miyun.

also show that the ARF method has high dependence on the data. Although the QC effect of the ARF method is superior to the QC effects of the traditional methods evaluated in this paper, the data in this paper retained the basic integrity and credibility of long-term temperature observations. If the data had been obtained from automatic weather stations, which contain missing observations over a long time, the ARF method's performance would have suffered. In addition, the performance indexes of the ARF method have only a minor advantage over those of the IDW and SRT methods in regions with high spatial correlations and a large number of neighboring stations such as TY. Meanwhile, the ARF method takes longer than the RF and traditional methods.

To address defects in the ARF method in future research, particularly what relates to the lack of QC observations for target stations, multiple source observations for target stations that lack neighboring stations should be inserted during testing using a spatial interpolation method to simulate neighboring stations. In addition, the physical links between different meteorological elements and the relationship between surface temperature and altitude and distance of different stations should be considered for the QC of the surface temperature element at the target stations using several meteorological elements from neighboring stations.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

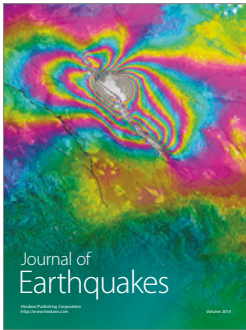
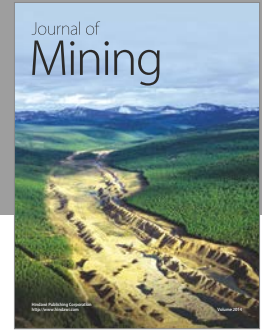
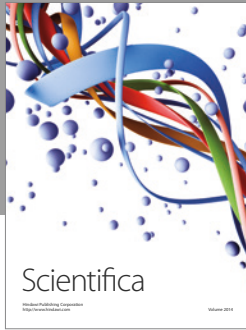
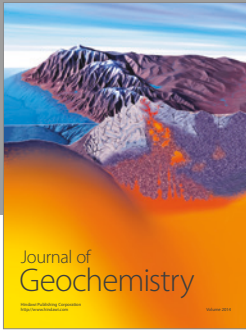
Acknowledgments

This project was supported by National Natural Science Foundation of China under Grant no. 41675156, Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grants nos. SJZZ16_0155 and SJZZ16_0156, National Public Welfare Industry (Meteorological) Scientific Research (no. GYHY20110604), and Six Talent Peaks Project in Jiangsu Province (WLW-021).

References

- [1] U. Schneider, A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, "GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle," *Theoretical and Applied Climatology*, vol. 115, no. 1-2, pp. 15–40, 2014.
- [2] M. Schindelegger and R. D. Ray, "Surface pressure tide climatologies deduced from a quality-controlled network of barometric observations," *Monthly Weather Review*, vol. 142, no. 12, pp. 4872–4889, 2014.
- [3] A. R. Cheng, T. H. Lee, H. I. Ku, and Y. W. Chen, "Quality control program for real-time hourly temperature observation in Taiwan," *Journal of Atmospheric and Oceanic Technology*, vol. 33, no. 5, pp. 953–976, 2016.
- [4] Z. Yan and P. D. Jones, "Detecting inhomogeneity in daily climate series using wavelet analysis," *Advances in Atmospheric Sciences*, vol. 25, no. 2, pp. 157–163, 2008.
- [5] C.-D. Xu, J.-F. Wang, M.-G. Hu, and Q.-X. Li, "Estimation of uncertainty in temperature observations made at meteorological stations using a probabilistic spatiotemporal approach," *Journal of Applied Meteorology and Climatology*, vol. 53, no. 6, pp. 1538–1546, 2014.
- [6] I. Durre, M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, "Comprehensive automated quality assurance of daily surface observations," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 8, pp. 1615–1633, 2010.
- [7] J. H. Lawrimore, M. J. Menne, B. E. Gleason et al., "An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3," *Journal of Geophysical Research Atmospheres*, vol. 116, no. 19, article D19, 2011.
- [8] S. Feng, Q. Hu, and W. H. Qian, "Quality control of daily meteorological data in China, 1951–2000: a new dataset," *International Journal of Climatology*, vol. 24, no. 7, pp. 853–870, 2004.
- [9] G. Sciuto, B. Bonaccorso, A. Cancelliere, and G. Rossi, "Quality control of daily rainfall data with neural networks," *Journal of Hydrology*, vol. 364, no. 1-2, pp. 13–22, 2009.
- [10] D. W. Meek and J. L. Hatfield, "Data quality checking for single station meteorological databases," *Agricultural and Forest Meteorology*, vol. 69, no. 1-2, pp. 85–109, 1994.

- [11] Y. Xiaoling, Z. Jianhua, and X. Xiong, "A GEP-based method for quality control of surface temperature observations," *Journal of Tropical Meteorology*, vol. 30, no. 6, pp. 1196–1200, 2014.
- [12] X. Xiong, X. Ye, and Y. Zhang, "A quality control method for surface hourly temperature observations via gene-expression programming," *International Journal of Climatology*, 2017.
- [13] S. Karatas and L. Yalcin, "Data Quality management," in *Proceedings of the TECO 2005*, https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Posters/P3%202833%29_Turkey_5_Karatas.pdf.
- [14] C. G. Wade, "A quality control program for surface mesometeorological data," *Journal of Atmospheric and Oceanic Technology*, vol. 4, no. 3, pp. 435–453, 1987.
- [15] S. L. Barnes, "A technique for maximizing details in numerical weather map analysis," *Journal of Applied Meteorology*, vol. 3, no. 4, pp. 396–409, 1964.
- [16] K. G. Hubbard, N. B. Guttman, J. You, and Z. Chen, "An improved QC process for temperature in the daily cooperative weather observations," *Journal of Atmospheric and Oceanic Technology*, vol. 24, no. 2, pp. 206–213, 2007.
- [17] J. You and K. G. Hubbard, "Quality control of weather data during extreme events," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 2, pp. 184–197, 2006.
- [18] H. Wang and Y. Liu, "Comprehensive consistency method of data quality controlling with its application to daily temperature," *Journal of Applied Meteorological Science*, vol. 23, no. 1, pp. 69–76, 2012.
- [19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [20] M. Neshat, G. Sepidnam, M. Sargolzaei, and A. N. Toosi, "Artificial fish swarm algorithm: a survey of the state-of-the-art, hybridization, combinatorial and indicative applications," *Artificial Intelligence Review*, vol. 42, no. 4, pp. 965–997, 2014.
- [21] K. G. Hubbard and J. You, "Sensitivity analysis of quality assurance using the spatial regression approach—a case study of the maximum/minimum air temperature," *Journal of Atmospheric and Oceanic Technology*, vol. 22, no. 10, pp. 1520–1530, 2005.
- [22] K. G. Hubbard, S. Goddard, W. D. Sorensen, N. Wells, and T. T. Osugi, "Performance of quality assurance procedures for an applied climate information system," *Journal of Atmospheric and Oceanic Technology*, vol. 22, no. 1, pp. 105–112, 2005.
- [23] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.
- [24] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC Bioinformatics*, vol. 9, no. 1, article 307, 2008.
- [25] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I—a discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970.
- [26] C. J. Willmott, "Some comments on the evaluation of model performance," *Bulletin of the American Meteorological Society*, vol. 63, no. 11, pp. 1309–1313, 1982.
- [27] I. Chollett, F. E. Müller-Karger, S. F. Heron, W. Skirving, and P. J. Mumby, "Seasonal and spatial heterogeneity of recent sea surface temperature trends in the Caribbean Sea and southeast Gulf of Mexico," *Marine Pollution Bulletin*, vol. 64, no. 5, pp. 956–965, 2012.
- [28] R. Hill, U. Schreiber, R. Gademann, A. W. D. Larkum, M. Kühl, and P. J. Ralph, "Spatial heterogeneity of photosynthesis and the effect of temperature-induced bleaching conditions in three species of corals," *Marine Biology*, vol. 144, no. 4, pp. 633–640, 2004.
- [29] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic Geography*, vol. 46, supplement 1, pp. 234–240, 1970.
- [30] M. Tiefelsdorf and B. Boots, "The exact distribution of Moran's I," *Environment and Planning A*, vol. 27, no. 6, pp. 985–999, 1995.
- [31] T. V. Perneger, "What's wrong with Bonferroni adjustments," *British Medical Journal*, vol. 316, no. 7139, article 1236, 1998.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

