

Research Article

Detecting Anomalies in Meteorological Data Using Support Vector Regression

Min-Ki Lee ¹, Seung-Hyun Moon, ¹ Yourim Yoon ², Yong-Hyuk Kim ³,
and Byung-Ro Moon ¹

¹School of Computer Science and Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

²Department of Computer Engineering, College of Information Technology, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do 13120, Republic of Korea

³School of Software, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 01897, Republic of Korea

Correspondence should be addressed to Byung-Ro Moon; moon@snu.ac.kr

Received 20 February 2018; Accepted 10 May 2018; Published 26 June 2018

Academic Editor: Alastair Williams

Copyright © 2018 Min-Ki Lee 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.

Significant errors exist in automated meteorological data, and identifying them is very important. In this paper, we present a novel method for determining abnormal values in meteorological observations based on support vector regression (SVR). SVR is used to predict the observation value from a spatial perspective. The difference between the estimated value and the actual observed value determines if the observed value is abnormal or not. In addition, SVR input variables are deliberately selected to improve SVR performance and shorten computing time. In the selection process, a multiobjective genetic algorithm is used to optimize the two objective functions. In experiments using real-world data sets collected from accredited agencies, the proposed estimation method using SVR reduced the RMSE by an average of 45.44% whilst maintaining competitive computing times compared to baseline estimators.

1. Introduction

Meteorological observations play an important role in weather forecasting, disaster warning, and policy formulation in agriculture and various industries [1–4]. In addition, meteorological observations are used for efficient operation of alternative energy sources such as solar power, hydro-power, and wind power [5–7]. In recent years, as climate change due to global warming has accelerated, the extent of damage due to abnormal weather phenomena is increasing and becoming more difficult to predict. Consequently, there is a greater need for accurate and quantitative climate data based on meteorological observations. The collection of meteorological information, which was previously done manually, has been automated in line with computational advances. An automatic weather station (AWS) is an automated system that allows a computer to observe and collect numerical values of multiple meteorological elements. The development of AWS

has enabled (i) real-time information retrieval, (ii) reduced maintenance costs, (iii) increased accuracy of observations, (iv) a larger amount of data, and (v) easier weather observations in poorly accessible regions.

However, meteorological data gathered by AWS often includes errors, and unusual metrics can be observed for a variety of reasons. Causes of unusual values include sensor malfunction, hardware error, power supply error, ambient environment change, and, in some rare circumstances, abnormal weather phenomena. Therefore, a quality control procedure is required for the collected data. Abnormal data identified during the quality control process are examined thoroughly by an expert and may become the subject of further research. If an abnormality is detected due to an error in the measurement process, it is necessary to replace the observed value with a corrected value [8–11]. Quality control of meteorological observations can also be regarded as anomaly detection [12] because anomalous values are of

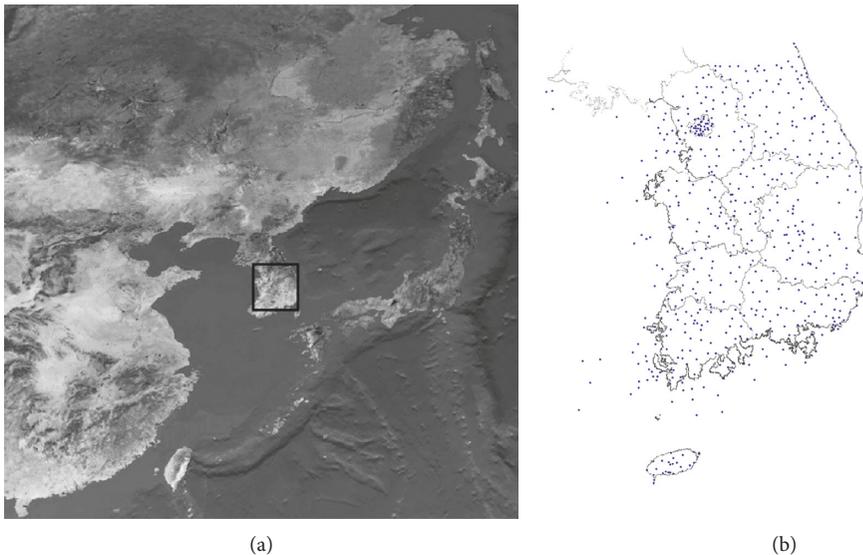


FIGURE 1: Locations of the 572 automatic weather stations (AWSs) in South Korea [18].

substantial interest to researchers. As the installation of AWS is expanding, and the amount and types of collected data are increasing, a fast and reliable quality control algorithm must be developed.

Quality control is achieved by several methods, ranging from simple discrimination using criteria related to physical limits to relatively complex discrimination related to spatiotemporal relationships with other observations [13]. Daly et al. [14] performed quality control of meteorological data metrics using climate statistics and spatial interpolation, and Sciuto et al. [15] proposed a spatial quality control procedure for daily rainfall data using a neural network. We propose a spatial quality control method, which uses values obtained from observational stations surrounding the target observational station to determine spatial compatibility and estimate the value of the observation station. It is possible to determine if an observed value is abnormal or not, based on differences with the estimated value. The developed spatial quality control method uses support vector regression (SVR) and a genetic algorithm. It can be applied to a wide range of meteorological elements to reflect the geographic and climatic characteristics of observation stations by studying past data through SVR. During preprocessing of the SVR, input variables, that is, the surrounding observation stations are selected according to two objective functions: similarity and spatial dispersion. Multiobjective optimization is required to simultaneously optimize the objective function that could be incompatible with these two functions. This is effectively performed by the genetic algorithm, which improves SVR performance and reduces execution time in this study. To verify the performance of the proposed method, we applied it to observational data measured by the Korea Meteorological Administration (KMA) for one year in 2014. Experiments on real-world data sets show that the performance of the proposed method is superior to previous methods such as the Cressman method [16] and the Barnes method [17], which have previously been used for spatial quality control.

TABLE 1: Meteorological elements in automatic weather station (AWS) data.

Meteorological element	Unit
Wind direction	°
Wind speed	m/s
Temperature	°C
Humidity	%
Atmospheric pressure	hPa
Hourly precipitation	
Rainfall occurrence	0 or 1

The remainder of this paper is organized as follows. Section 2 describes the problem that we attempt to solve in this study. Spatial quality control is defined, and we describe real-world data sets used to test the performance of the proposed methods. Section 3 introduces the methods previously used in spatial quality control, Section 4 describes an estimation model using SVR, Section 5 describes the algorithm for selecting SVR input variables, and Section 6 presents the experimental results using the real-world dataset. Finally, Section 7 discusses our conclusions.

2. Problem Description

2.1. Data Sets. This study covers data from 572 AWSs operated by KMA in South Korea. Figure 1 shows the locations of the target AWSs.

Target data include meteorological information measured every 1 minute from January 1, 2014 at 00:00 to December 31, 2014, at 23:59. In one year, 525,600 pieces of observational data are collected for each meteorological element at each station. We selected seven major meteorological elements for analysis: 10-minute average wind direction, 10-minute average wind speed, 1-minute average temperature, 1-minute average humidity, 1-minute average pressure, 1-hour cumulative amount of precipitation, and precipitation. Table 1 shows the types and units of meteorological elements used in this study.

The wind direction is expressed in degrees; however, this leads to a large error in the algorithm internal operation. For example, 1° and 359° differ only by 2° , but arithmetically, they differ by 358° . In addition, the average of the two wind directions should be regarded as 0° , but arithmetically it is 180° . Therefore, to avoid these problems, wind direction was converted into a two-dimensional unit vector, as expressed in [8].

$$\mathbf{v} = \left(\cos\left(\theta \cdot \frac{\pi}{180}\right), \sin\left(\theta \cdot \frac{\pi}{180}\right) \right), \quad (1)$$

where \mathbf{v} is the transformed two-dimensional vector and θ is the original wind direction in degrees. In the quality control process, the components of the two vectors are processed separately. When a quantitative comparison of wind direction is required, the wind direction represented by the vector must be converted back to degrees:

$$\theta = \text{atan2}(v, u) \cdot \frac{180}{\pi}, \quad (2)$$

where u and v are \mathbf{v} 's first and second components, respectively, and atan2 is defined as follows:

$$\text{atan2}(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right), & x > 0 \\ \arctan\left(\frac{y}{x}\right) + \pi, & x < 0, y \geq 0 \\ \arctan\left(\frac{y}{x}\right) - \pi, & x < 0, y < 0 \\ +\frac{\pi}{2}, & x = 0, y > 0 \\ -\frac{\pi}{2}, & x = 0, y < 0 \\ \text{undefined}, & x = 0, y = 0. \end{cases} \quad (3)$$

2.2. Quality Control. During quality control, when the observed value is determined as not normal, it is then classified as “suspect” or “error” according to its level of abnormality. “Suspect” means that the value is likely to be abnormal, and “error” indicates that it is definitely anomalous.

2.2.1. Basic Quality Control. Basic quality control is a relatively simple quality control procedure performed in real-time. Abnormalities are determined using only the observed value itself or observations of a short time before and after. The data used in this study were first filtered through the following four basic quality control procedures. Each test was performed sequentially. If any test failed, the data were classified as an error, and subsequent tests were not performed. Each test and the numerical criteria are the same as those used by KMA.

- (i) *Physical limit test:* If the observed value is higher or lower than the physically possible upper or lower

TABLE 2: Limits for physical limit test.

Meteorological element	Lower limit	Upper limit
Wind direction	0°	360°
Wind speed	0 m/s	75 m/s
Temperature	-80°C	60°C
Humidity	1	100
Atmospheric pressure	500 hPa	1080 hPa
Precipitation	0	400
Rainfall occurrence	0	1

TABLE 3: Maximum amount of change for step test.

Meteorological element	Maximum amount of change
Wind speed	10 m/s
Temperature	1°C
Humidity	10
Atmospheric pressure	2 hPa

TABLE 4: Minimum amount of change for persistence test.

Meteorological element	Minimum amount of change
Wind speed	0.5 m/s
Temperature	0.1°C
Humidity	1.0
Atmospheric pressure	0.1 hPa

limit, respectively, it is classed as an error. The physical limit test is performed on all meteorological elements. Table 2 shows the physical limits of each meteorological element, which are based on World Meteorological Organization (WMO) standards [19].

- (ii) *Step test:* The step test is performed for wind speed, temperature, humidity, and atmospheric pressure. If the difference between the current observation value and the value one minute prior is more than a certain value, it is classed as an error. Table 3 shows the maximum variation of each meteorological element.
- (iii) *Persistence test:* The persistence test is performed for wind speed, temperature, humidity, and atmospheric pressure. A value is classified as an error when the accumulated change in the observed value within 60 minutes is smaller than a certain value. Table 4 shows the minimum variation within 60 minutes for each meteorological element.
- (iv) *Internal consistency test:* The internal consistency test is performed for pairs of wind direction and wind speed data and pairs of precipitation and rainfall occurrence data. If any one of the factors in each pair is determined to be an error in another test, the other factor is also perceived as an error. Also, if the rainfall occurrence value is 0 but the precipitation value is not 0, both values are classed as suspects.

Table 5 shows the percentages of normal, error, and suspect values, respectively, after performing each test on the KMA dataset. If the observed meteorological element is not available due to an absence of observational equipment, or if

TABLE 5: Results of basic quality control.

Meteorological element	Normal (%)	Limit Error (%)	Step Error (%)	Persistence Error (%)	Consistency		Uninspected (%)
					Suspect (%)	Error (%)	
Wind direction	81.97	$1.68e^{-3}$	N/A	N/A	0.00	2.80	15.23
Wind speed	81.90	$2.85e^{-4}$	$1.36e^{-3}$	$8.47e^{-1}$	0.00	2.80	14.45
Temperature	96.41	$3.22e^{-3}$	$7.73e^{-3}$	$1.02e^{-1}$	N/A	N/A	3.48
Humidity	54.47	$2.23e^{-2}$	$3.38e^{-3}$	5.43	N/A	N/A	40.07
Atmospheric pressure	38.30	$1.83e^{-1}$	$5.72e^{-5}$	$3.69e^{-2}$	N/A	N/A	61.48
Hourly precipitation	93.08	0.00	N/A	N/A	1.94	0.00	4.98
Rainfall occurrence	93.08	$2.34e^{-4}$	N/A	N/A	1.94	0.00	4.98

the observed value is missing, it is classified as uninspected. All subsequent experiments were performed only on data determined as normal after basic quality control.

2.2.2. Spatial Quality Control. Spatial quality control determines whether the observation data of the target station are abnormal based on the values of other observation stations around the target station. It is sometimes referred to as a spatial consistency test [13]. Because this test is based on a large amount of data, it involves more time and resources than basic quality control. Therefore, spatial quality control is often performed in quasi-real-time. Typical spatial quality control process is as follows:

- (i) Estimate the value of the target station using the values of surrounding observation stations.
- (ii) If the difference between the observed and the predicted value of target station is greater than the critical value, the observation is considered as abnormal.

The meteorological elements of the KMA dataset, excluding rainfall occurrence, consist of continuous values; therefore, the predicted value can be estimated naturally via the interpolation or regression model. In the case of rainfall occurrence, it has a value of 0 or 1, so the value should be taken as 0 if the estimated value is less than 0.5, and 1 if the estimated value is 0.5 or more. The acceptable range for the difference between the observed value and the predicted value is generally determined using the standard deviation of the surrounding stations, which we set to the observation stations within 30 km of the target station. If the standard deviation is 0, because the observation values of all neighboring stations are the same, it is difficult to determine the acceptable range; therefore, the test is not performed. In the KMA dataset, this was often the case for elements such as precipitation and rainfall occurrence, which are always zero during periods of nonrain. Moreover, if there are less than three stations within 30 km, spatial quality control does not proceed because reliable standard deviations cannot be calculated. Also, observations that are missing or identified as abnormal during basic quality control are not considered for spatial quality control.

If the tolerance of the difference between the observed and predicted value is the same, the accuracy of the predicted value estimation will determine the reliability of spatial quality control. In this study, we aim for more accurate spatial prediction and thus improved spatial quality control performance. Traditional spatial prediction methods include

spatial interpolation methods such as the Cressman method [16] and the Barnes method [17]. However, these methods do not reflect the geographical features of each region because they depend only on relative position to estimate the predicted value. Here, we propose a method to improve the accuracy of estimates by overcoming the shortcomings of existing methods by using supervised learning techniques.

3. Previous Methods

This section describes the spatial interpolation methods used in this study: the Cressman method and the Barnes method. The two methods have been slightly modified by the KMA to detect meteorological anomalies in South Korea. Actual observations are compared with estimates generated by the spatial interpolation methods. If there is a significant difference between observed and predicted values, the observation is classed as “suspect” or “error” according to the degree of difference.

3.1. Cressman Method. The Cressman method performs spatial interpolation on a two-dimensional distribution of meteorological elements. Meteorological elements at each station are irregularly distributed in two dimensions and converted into estimated values of the grid points at regular intervals. In this study, the grid interval is 0.2° for both longitude and latitude. The estimated values of the grid points are called the background field and are calculated with respect to the effective radius r . The effective radius is the control parameter describing the maximum station distance when estimating each grid point. Let z_i be the observed value of the station i , and d_{ei} denote the distance between the grid point e and the station i . Then, $Z_r(e)$, the estimated value of the grid point e , is the weighted average of the observations within the effective radius r (Figure 2):

$$Z_r(e) = \frac{\sum w_r(i) \cdot z_i}{\sum w_r(i)}, \quad (4)$$

where $w_r(i)$, the weight of the station i , depends only on the distance:

$$w_r(i) = \begin{cases} \frac{(r^2 - d_{ei}^2)}{(r^2 + d_{ei}^2)} & \text{if } d_{ei} \leq r \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

To obtain $Z(i)$, the estimated value of a station i , the estimates of the four closest grid points from the station are

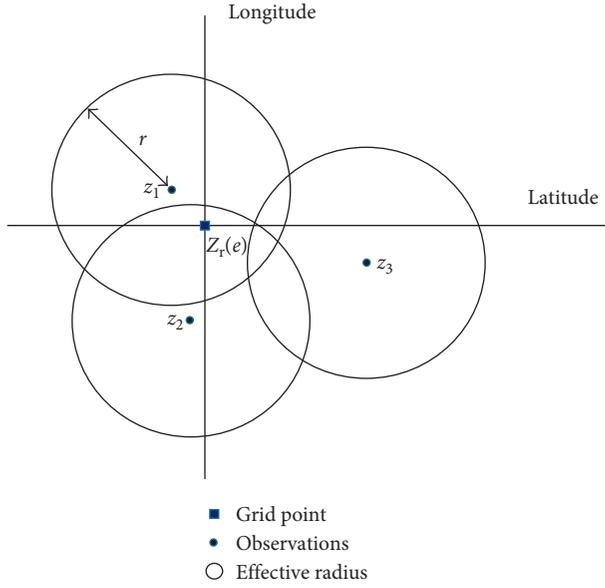


FIGURE 2: Calculation of the $Z_r(e)$, the estimated value of the grid point e in the Cressman method. Only observations of stations located within the effective radius r are used. In this example, z_1 and z_2 are used to calculate $Z_r(e)$, but z_3 is not used.

averaged. After calculating the estimates of all the stations, the background field can be recalculated using the estimates instead of the observations. The estimates of the stations can also be recalculated over the new background field. This process can be repeated as many times as desired. We set the effective radius to 50 km, 30 km, and 10 km and updated the background field and the estimates of the stations.

Let σ_i be the standard deviation of the observations at all stations located within the final effective radius of the station i . If $|z_i - Z(i)|$ is greater than $3 \cdot \sigma_i$, z_i is classified as an error. If $|z_i - Z(i)|$ is greater than $2 \cdot \sigma_i$, z_i is classified as a suspect.

3.2. Barnes Method. The Barnes method is a statistical technique that can derive accurate two-dimensional distribution from randomly distributed data in space. It is similar, in many respects, to the Cressman method, but instead uses a Gaussian function in the weight function:

$$w_r(i) = \begin{cases} \exp\left(\frac{-d_{ij}^2}{2r^2}\right) & \text{if } d_{ij} \leq r \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where d_{ij} is the distance between stations i and j . The KMA uses observations without using grid points when calculating the estimates by the Barnes method:

$$Z(i) = \frac{\sum w_r(i) \cdot z_i}{\sum w_r(i)}, \quad (7)$$

where r is set to 30 km. The process of determining whether or not observations are normal is almost identical to that of the Cressman method. Let σ_i be the standard deviation of the observations at all stations located within 30 km of the station i .

If $|z_i - Z(i)|$ is greater than $3 \cdot \sigma_i$, z_i is classified as an error. If $|z_i - Z(i)|$ is greater than $2 \cdot \sigma_i$, z_i is classified as a suspect.

4. SVR-Based Approach

In this section, we propose a method using support vector regression (SVR) to overcome the spatial prediction limitations of the Cressman and Barnes methods for a target observation station from a spatial quality control perspective. The support vector machine (SVM) is a supervised machine learning method developed by Vapnik and Lerner in 1963 [20]. In the 1990s, nonlinear classification using SVM became popular as an alternative to artificial neural networks, and SVM generally has less overfitting problems than artificial neural networks. In the SVM, learning proceeds in the direction of maximizing the margin of the support vector, which is a hyperplane that divides each class of the given data. During early research, only linear classification was possible; however, nonlinear classification became feasible by mapping data to a higher dimensional space using a kernel function. A typical nonlinear kernel function is a radial basis function (RBF). The RBF function transforms the original space into an infinite dimension Hilbert space. In this study, the RBF function was used because the RBF function is better, on average, than the linear or polynomial function. Support vector regression (SVR) was proposed, which uses SVM for regression with continuous values as the output [21]. Preliminary study on meteorological elements has shown that the estimation capability of SVR is superior to other machine learning techniques [8]. In this study, the SVMlight [22] library was used for C language implementation.

The input and output of the SVR model for spatial quality control are as follows:

- (i) *Input*: observations of stations surrounding the target station
- (ii) *Output*: observation value of the target station.

In the input, values that are missing or classified as errors during basic quality control are replaced by the temporally closest values. The wind speed converted into a 2D vector representation was learned and tested by two separate models for each dimension. Once the model is learned from the values of the target station and the surrounding observation stations in the past, the predicted value of the target station can be estimated for the input that has not been learned. Because past observation values are not labeled as normal or abnormal with respect to spatial quality control, they are learned regardless of normal and abnormal values. Therefore, this approach assumes that most observations are normal and abnormalities are few.

Once the predicted value of the target station is estimated, the process of determining whether the observed value of the target station is normal is the same as the Cressman method or Barnes method. Let z_i and $Z(i)$ be the observations value and the estimated value by SVR of station i , respectively, and let σ_i be the standard deviation of the observations from all stations within a radius of 30 km of station i . If $|z_i - Z(i)|$ is greater than $3 \cdot \sigma_i$, z_i is classified as an error. If $|z_i - Z(i)|$ is greater than $2 \cdot \sigma_i$, z_i is classified as a suspect.

The SVR model can implicitly capture the geographic characteristics of the target station while learning past data.

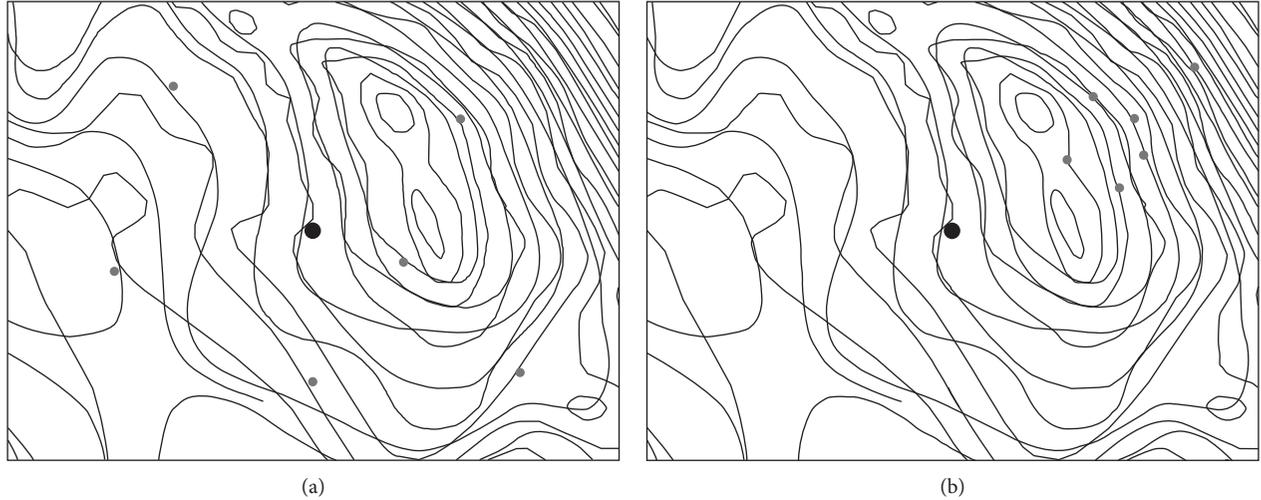


FIGURE 3: Examples of neighbor selection. (a) Neighbor stations with high spatial dispersion. (b) Neighbor stations with low spatial dispersion.

Through this process, the combination of each station and each meteorological element has its own specific model. This is an advantage of SVR over non-ML approaches. However, an approach based on machine learning also has its drawbacks; specifically, that it takes a long time to learn. A method to overcome this is introduced in Section 5.

5. Selecting Neighboring Stations

The input of the SVR model uses the observations of neighboring AWSs within a certain radius of the target AWS. However, if there are too many neighbors, the learning time of SVR becomes too long. Also, some neighboring stations act as noise instead of helping to estimate the value of the target station. Therefore, it is necessary to select the best core neighbors to estimate the value of the target station while reducing the number of neighbors used in the input.

5.1. Similarity and Spatial Dispersion. Two criteria were applied to select key neighbors. The first considered the similarity of the observations between the target station and the neighbor station. Observations at locations with similar meteorological phenomena are helpful in deriving observations at the target site. The second considered how widespread the neighboring stations were in space. If one constructs a core neighborhood at stations concentrated in a narrow area, the model cannot be flexible to various situations. For example, if there is a peculiar meteorological phenomenon within a narrow area (e.g., a local storm), the estimate will be misled. Spatial dispersion ensures statistical robustness of the model. Figure 3 shows two different choices of neighboring stations. When the amount of rainfall in target station is estimated, the amount of rainfall in neighboring stations is used. If localized heavy rain happens in an area including neighboring stations with low spatial dispersion, the estimated amount of rainfall will be inclined to be very high even though target station is out of influence of localized heavy rain. On the other hand, it reflects overall surrounding circumstances of rainfall when spatial dispersion of neighboring stations is high.”

To measure the similarity of stations according to their meteorological elements, the time series values of the elements are expressed as vectors, and the distance between them is measured in various ways. We used the L_1 distance, L_2 distance, Pearson correlation coefficient, and mutual information to measure the similarity between two vectors. After the distance of all the station pairs was calculated, the smallest value was zeroed and the largest value was normalized to one. The L_1 distance, known as the Manhattan distance or taxicab distance, between two vectors \mathbf{x} and \mathbf{y} was calculated as follows:

$$\|\mathbf{x} - \mathbf{y}\|_1 = \sum_{i=1}^n |x_i - y_i|, \quad (8)$$

where x_i is the i -th element of \mathbf{x} . We used $(1-L_1$ distance) as a similarity measure to ensure that larger measurement was given to two vectors which were more similar. The L_2 distance, known as the Euclidean distance, between two vectors \mathbf{x} and \mathbf{y} was calculated as follows:

$$\|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (9)$$

We used $(1-L_2$ distance) as a similarity measure to ensure that larger measurement was given to two vectors which were more similar. The Pearson correlation coefficient is used to measure the degree of the linear relationship between two variables. It has a value 1 when there is a perfect positive linear correlation and -1 when there is a perfect negative linear correlation. The Pearson correlation coefficient is calculated as follows, where $\bar{x} = \sum_{i=1}^n x_i/n$:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (10)$$

Mutual information measures the mutual dependence between two random variables X and Y . It quantifies the reduction in uncertainty of one of the variables due to knowing the other. Mutual information is calculated as

```

non-dominated set  $E = \emptyset$ ;
initialize population  $P$ ;
repeat
    assign a fitness value to each solution in  $P$ ;
    select  $2N$  parents from  $P$ ;
    create  $N$  offspring applying crossover on the parents;
    mutate offspring;
    repair offspring;
    local-optimize offspring;
     $P \leftarrow$  offspring;
    update  $E$ ;
    remove  $n_E$  solutions from  $P$ ;
    add  $n_E$  solutions from  $E$  to  $P$ ;
until stopping condition;
return  $E$ ;

```

FIGURE 4: The framework of our hybrid multiobjective genetic algorithm.

follows, where $p(x, y)$ is the joint probability function of x and y and $p(x)$ and $p(y)$ are the marginal probability density functions of x and y , respectively:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right). \quad (11)$$

We computed the mutual information from the observed frequency of two vectors, \mathbf{x} and \mathbf{y} , assuming that these vectors constitute an independent and identically distributed sample of (X, Y) .

As a measure of spatial dispersion, we used the average of the geographical distance from the nearest station [23]. If the set of target stations and selected neighbors is x , and $d_{x_i x_j}$ is the normalized geographic distance between the two stations x_i and x_j , then the spatial dispersion is calculated as

$$\text{dispersion}(x) = \frac{\sum_{x_i \in x} \min_{x_j \in x, x_i \neq x_j} d_{x_i x_j}}{\sum_{x_i \in x} 1}. \quad (12)$$

The larger the spatial dispersion is, the better the neighborhood selection is. The two criteria of similarity and spatial dispersion often conflict. In general, similarities in climatic characteristics are often due to geographic proximity. Therefore, the key neighborhood screening problem is a multiobjective optimization problem that simultaneously optimizes two or more objectives that are not independent of each other. In this study, we solve the multiobjective optimization problem using genetic algorithms.

5.2. Multiobjective Genetic Algorithm. The genetic algorithm (GA) is a global optimization technique developed by Holland, which mimics the natural evolution of biological selection [24]. It is used to find a solution with high (or low) fitness while repeating a genetic operation that imitates processes such as selection, crossover, and mutation, which are important elements of evolution. GA is a type of metaheuristic that does not depend substantially on the nature of the problem. It can search all ranges and is less likely to fall into a local optimum. Pure GA is disadvantageous in that it takes a long time to

converge. The hybrid genetic algorithm solves this problem by combining the local optimization algorithm with the GA. Several successful attempts have been made to solve multiobjective problems using GA [25–29]. Among them, NSGA-II by Deb [30] is the most well-known. To maximize the function f_1, f_2, \dots, f_n with n number of objects, if solution x and solution y satisfy the following condition, then it can be said that solution y dominates solution x :

$$f_i(x) \leq f_i(y) \quad \forall i \text{ and } \exists j : f_j(x) < f_j(y). \quad (13)$$

When a solution is not dominated by any other solution, the solution is called Pareto optimal. To improve an objective function in a Pareto optimal solution, one has to sacrifice another objective function. The multiobjective genetic algorithm (MOGA) does not output one solution but several Pareto optimal solutions. The final solution selection is performed by the decision-maker.

5.3. Our GA Framework. In this study, we tested the SVR with several Pareto optimal solutions for each meteorological element, and selected the best solution on average. Figure 4 shows the structure of the GA used in this study.

- (i) *Encoding*: In a genetic algorithm, one solution is expressed as a set of genes, or a chromosome. Here, one chromosome is represented by a one-dimensional binary string. Each gene corresponds to one station. If the value of the gene is “0,” the observation value of the corresponding station is not used as the input of the SVR. If it is “1,” it is selected as an input of the SVR.
- (ii) *Fitness*: This indicates the validity of the solution for a given problem. When the individual objective function is f_1, f_2, \dots, f_n , the fitness value of solution x is calculated as

$$f(x) = w_1 f_1(x) + w_2 f_2 + \dots + w_n f_n(x), \quad (14)$$

where w_1, \dots, w_n are nonnegative and $\sum w_i = 1$, each weight w_i is randomly set for every

generation, not as a fixed value. This allows the algorithm to search for various Paretooptimal solutions [31]. This method is more intuitive than the algorithm that uses Pareto ranking-based fitness evaluation [32] and easier to be combined with a local optimization algorithm [33]. In this problem, $n = 2$, and $f_1(x)$ and $f_2(x)$ correspond to similarity and spatial dispersion, respectively, as described in Section 5.

- (iii) *Population*: Population is a set of chromosomes. Chromosomes in the population interact each other to generate new solutions and cull existing solutions. In this study, the size of the population was set to 50. The initial population consisted of 50 randomly generated chromosomes.
- (iv) *Selection*: This is the operator used to select the parent chromosome for the crossover. To mimic the principle of survival of the fittest in nature, chromosomes with high fitness are selected with high probability. In this study, roulette-wheel selection, one of the most widely used selection operators, was used. The probability that the best fitness solution will be selected is four times the probability that the lowest fitness solution will be selected.
- (v) *Crossover*: A key operator of the genetic algorithm. In inheriting the features of the parents, we expect that the different advantageous traits combine to produce an offspring chromosome that is superior to the parents. In this study, we used a two-point crossover with two cut points.
- (vi) *Mutation*: This statistically modifies a portion of the offspring chromosome to increase solution diversity and prevent premature convergence. In this study, each gene was flipped with a probability of 10%.
- (vii) *Repair*: After crossover and mutation, offspring still may not meet the constraints of the problem. That is, the number of genes with a value of “1” in the chromosome may be different from the number of stations to be selected. If the number of genes with a value of “1” is insufficient, we repeat the process of changing the value of the randomly selected gene among genes with a value of “0” to “1.” On the other hand, if the number of genes with a value of “0” is insufficient, we repeat the process of changing randomly selected genes to “0” among genes with a value of “1.”
- (viii) *Local optimization*: This exchanges the values of two genes whose fitness value increases when exchanged. This process is repeated until the exchange of any two gene values can no longer increase the fitness value.
- (ix) *Replacement and elitism*: We used a generational GA to generate offspring as large as the size of the population and replace the entire population. Among the solutions found so far, nondominant solutions that are closest to the Paretooptimal are stored in an external archive. This nondominant solution archive

updates every time a new solution is created. In other words, the solution that is dominated by the new solution is removed from the existing nondominant solution archive, and the new solution is stored in the archive when it is a nondominant solution. As survival of good solutions within a population can result in a good solution for the next generation, some of the population are replaced by solutions in nondominant solution archive. In this algorithm, 20% of the entire population was randomly replaced with solutions in nondominant solution archive.

- (x) *Stopping condition*: The genetic algorithm stops when 1,000 generations have passed.

Figure 5 describes the whole spatial quality control process including neighbor selection.

6. Experimental Results

In this section, (i) detailed good parameters are selected, (ii) the performances of the estimation methods are compared, and (iii) the results of the proposed spatial quality control procedure are presented using meteorological data collected by the KMA for a year in 2014. The root mean square error (RMSE) was used as a measure to compare the accuracy of each estimation method. RMSE is a standard metrics for dealing with errors between model-estimated values and observed values in a real environment, including meteorology [34]. If θ is the observed vector and $\hat{\theta}$ is the estimated vector, then the RMSE of $\hat{\theta}$ is calculated as

$$\text{RMSE}(\hat{\theta}) = \sqrt{\mathbf{E}((\hat{\theta} - \theta)^2)}. \quad (15)$$

The lower the RMSE value, the better the model estimate. As the accuracy of estimates should be based on normal observations, only observations classed as normal by the model are used to calculate RMSE. When comparing the RMSE of two or more models, only those observations determined as normal by all models were used.

Performance evaluation of SVR estimation models was achieved through 10-fold cross-validation. All data were divided into 10 folds, of which 9 were used as the training set, and the other was used as the test set. Learning and test are performed 10 times so that each fold can be used as a test set. Due to there being 7 meteorological elements in 572 AWSs, and 10 models must be learned each time, a total of 40,040 models were created for each experiment. The entire training set consists of 473,040 data sets. Because of the large number of models and the overly long total execution time, we sampled 5,000 data sets and used them as training sets for the parameter optimization experiments. We then describe the change in performance and time caused by increasing the size of the training set once the final parameter is determined.

The experiment was performed on an Intel i7 quad-core 2.93 GHz CPU. Each experiment used only one core. Experiments with a long execution time were performed by dividing each of the seven machines by observatories, and the total execution time included the execution time of each machine.

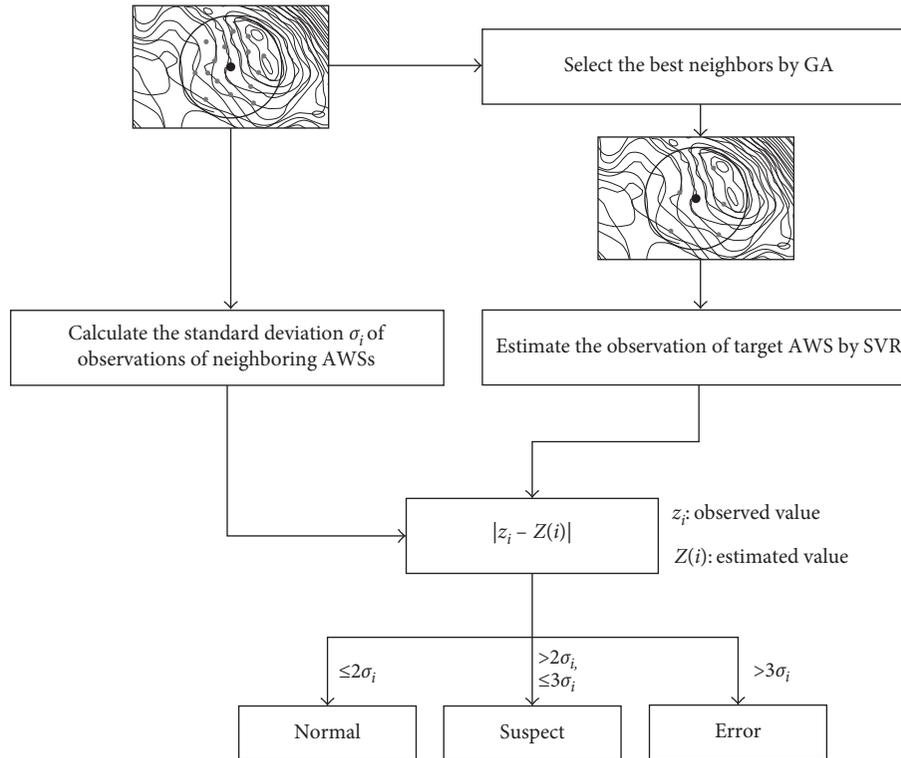


FIGURE 5: The proposed spatial quality control process.

TABLE 6: Accuracy of estimates according to wind direction representation (RMSE).

Representation	RMSE
Degree	92.17
Vector	68.28

6.1. *Representation of Wind Direction.* Section 2.1 describes the process of converting wind direction expressions from degrees to 2D vectors. Table 6 compares the accuracy of SVR estimates for each wind direction representation. The accuracy of the estimate is much higher when expressed in terms of vector expression than degrees. Thus, all subsequent experiments used a vector expression to represent wind direction.

6.2. *Similarity Measure.* Section 5 describes four measures used to calculate the similarity between two observation stations. To compare the usefulness of each measure, the accuracy of the estimates predicted by the Madsen-Allerup method [35] is examined. The Madsen-Allerup technique selects the stations similar to the target station and then uses the observed values of selected stations to obtain the estimate of the target station; therefore, the higher the quality of the similarity measure, the more accurate the estimate. Table 7 shows the estimation accuracy of the Madsen-Allerup method for each similarity measure. In all subsequent experiments, we used the highest quality similarity measure for each meteorological element. Figure 6 shows the connected station pairs with a similarity greater than 0.5.

TABLE 7: Accuracy of estimates for each similarity measure.

Meteorological element	L_1	L_2	PCC ¹	MI ²
Wind direction	104.574	105.624	102.786	101.424
Wind speed	1.228	1.224	1.306	1.317
Temperature	1.241	1.241	1.327	1.319
Humidity	8.085	8.086	8.757	8.829
Atmospheric pressure	6.497	6.497	8.134	7.256
Hourly precipitation	1.074	1.065	1.066	1.155
Rainfall occurrence	0.151	0.151	0.152	0.157

¹Pearson correlation coefficient. ²Mutual information.

6.3. *Selecting Neighboring Stations.* In Section 5, we proposed MOGA to select input variables to improve SVR performance and speed. Figure 7 shows the accuracy of estimates based on the number of neighboring stations selected by MOGA. The greater the number of parameters (over a certain amount), the worse the performance of the SVR and the longer it takes to train. The optimal number of neighboring stations with the best performance differs with the meteorological element. Table 8 shows the optimal number of neighboring stations according to each meteorological element. All subsequent experiments were fixed using the optimal number of neighbors. Table 9 compares the estimation accuracy of SVR when neighboring stations were selected randomly, with the accuracy of SVR when neighboring stations were selected using MOGA. We confirm that selection of neighbors using MOGA improves the estimation performance of SVR.

6.4. *Comparison of Estimation Models.* Table 10 shows the accuracy of estimates for each estimation model. Estimation

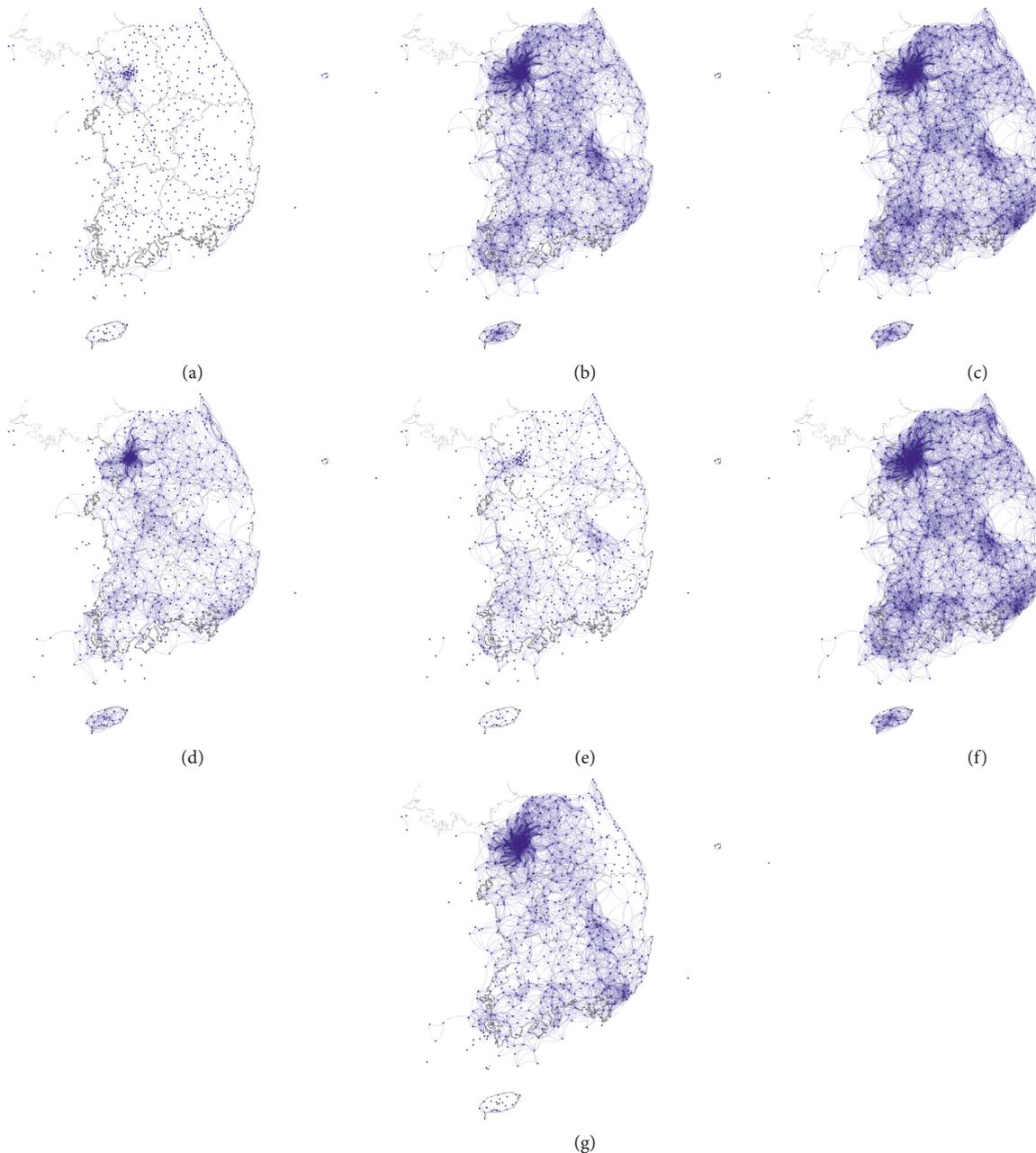


FIGURE 6: Similarity map for different meteorological elements. (a) Wind direction. (b) Wind speed. (c) Temperature. (d) Humidity. (e) Atmospheric pressure. (f) Precipitation. (g) Rainfall occurrence.

using SVR model is better than that using Cressman or Barnes algorithms. Hourly precipitation does not show much improvement compared to other meteorological elements. Because there are many more days without rain than with rain, there is rather sparse data distribution for rainy days, which results in learning difficulties.

Table 11 shows the execution time of spatial quality control according to each estimation model. The execution time might be considered of little importance as a single process of spatial quality control can be executed in a very small time. But if a quality control process should be performed in a centralized single facility, a large number of meteorological data from every observational station need to

be inspected in real time. For example, in our test data, there are 572 stations, and they collect 7 kinds of meteorological observation data. It takes about 5.77 seconds to inspect every data from every station using the Cressman method, and it should be executed in every minute. Moreover, the execution time becomes more important as the number of stations and the kind of data get bigger and the time interval for collecting data gets shortened.

Spatial quality control is fastest using the Barnes algorithm, but the accuracy of the estimation is very poor. Spatial quality control using SVR is approximately 6 times faster than that using the Cressman algorithm, but more time is required to learn the SVR model. However, as it does not

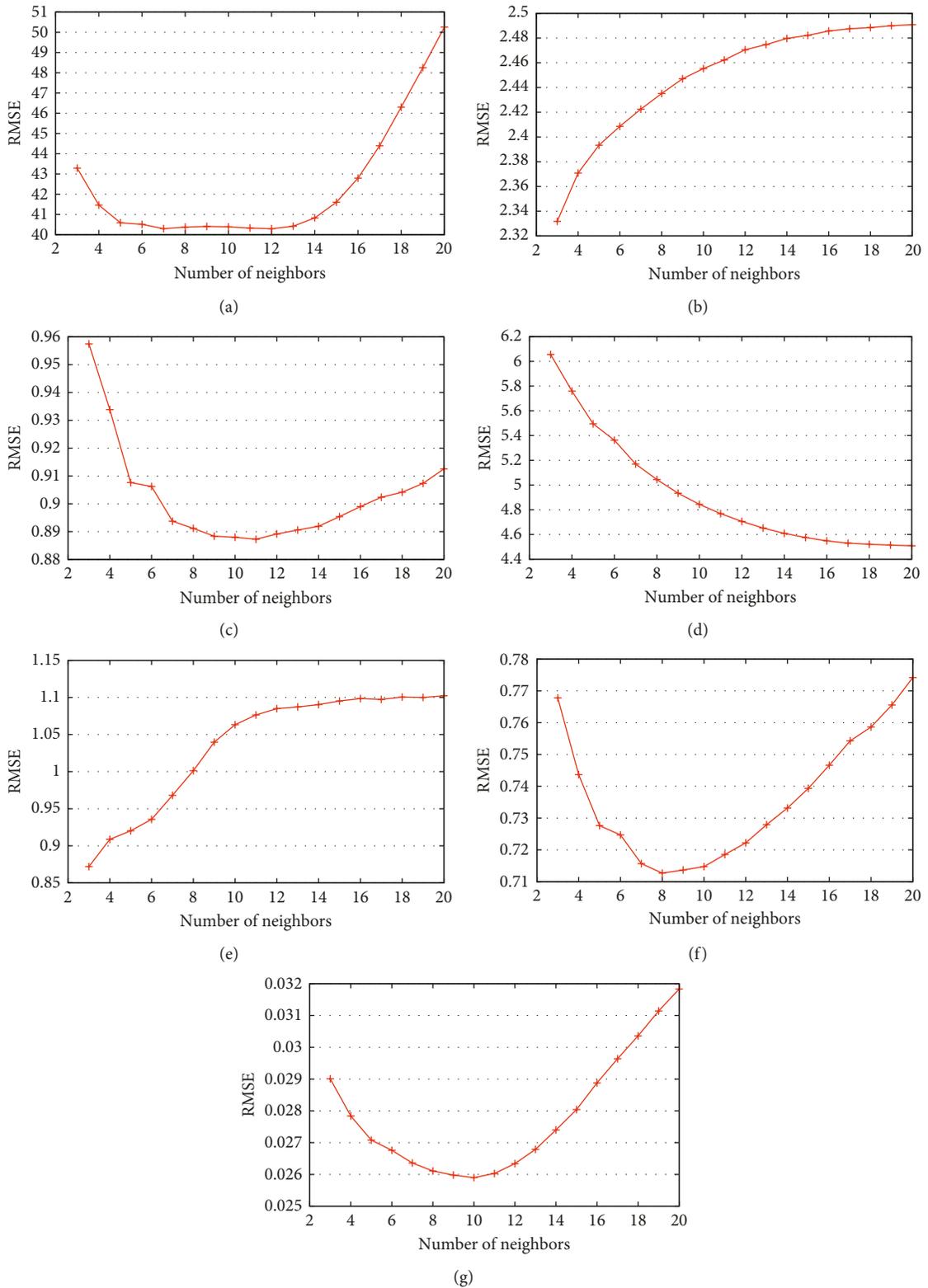


FIGURE 7: Accuracy of estimates according to the number of selected neighboring stations (RMSE). (a) Wind direction. (b) Wind speed. (c) Temperature. (d) Humidity. (e) Atmospheric pressure. (f) Precipitation. (g) Rainfall occurrence.

give weight to more recent data in the learning process, there is no need to learn the model every time the spatial quality control is performed. If the model uses sufficient previous

data, the performance of spatial quality control is not adversely affected, even if the learning cycle for model updates is only once a week or a month.

TABLE 8: Optimal number of neighboring stations per meteorological element.

Meteorological element	Optimal number of neighbors
Wind direction	7
Wind speed	3
Temperature	11
Humidity	20
Atmospheric pressure	3
Hourly precipitation	8
Rainfall occurrence	10

TABLE 9: Comparison of SVR estimation accuracy with neighboring stations selected randomly or by MOGA.

Weather element	Random	MOGA
Wind direction	50.390	48.499
Wind speed	2.523	2.513
Temperature	0.970	0.902
Humidity	5.216	5.038
Atmospheric pressure	1.066	1.063
Hourly precipitation	0.847	0.762
Rainfall occurrence	0.028	0.026

TABLE 10: Comparison of estimation accuracy based on estimation model (RMSE).

Meteorological element	Cressman	Barnes	SVR
Wind direction	53.568	75.470	48.341
Wind speed	2.347	2.315	2.179
Temperature	1.180	2.583	0.880
Humidity	6.755	12.767	4.582
Atmospheric pressure	5.663	11.601	0.847
Hourly precipitation	0.583	0.833	0.583
Rainfall occurrence	0.071	0.137	0.021

TABLE 11: Execution time for spatial quality control based on estimation model.

Estimator	Average time spent in learning one model (second)	Average time spent in determining one observation (second)
Cressman	—	$1.442e^{-3}$
Barnes	—	$8.427e^{-5}$
SVR	6.839	$2.303e^{-4}$

6.5. Size of Training Set. In general, the higher the number of training samples in the SVR, the higher the accuracy of the estimate but the longer the learning time. Table 12 shows the accuracy of estimates based on the number of training samples. Exceptionally, in the case of wind speed, the performance tends to decrease as the number of training samples increases. Figure 7 also shows that the fewer input variables of SVR, the better the performance with regards to wind speed. In the present model structure, it is difficult to learn wind speed; thus, overfitting seems to occur if the model becomes overly complicated.

Figure 8 shows the learning time according to the number of training samples, and Figure 9 shows the time taken purely for spatial quality control, excluding learning

TABLE 12: Accuracy of estimated values based on the size of the training set (RMSE).

Meteorological element	5000	10000	15000	20000	25000	30000
Wind direction	43.820	42.831	42.298	41.948	41.691	41.481
Wind speed	2.363	2.365	2.367	2.369	2.369	2.370
Temperature	0.902	0.879	0.870	0.863	0.860	0.857
Humidity	4.710	4.330	4.130	3.998	3.904	3.831
Atmospheric pressure	0.871	0.837	0.817	0.807	0.797	0.785
Hourly precipitation	0.763	0.746	0.736	0.732	0.727	0.724
Rainfall occurrence	0.026	0.025	0.024	0.024	0.024	0.024

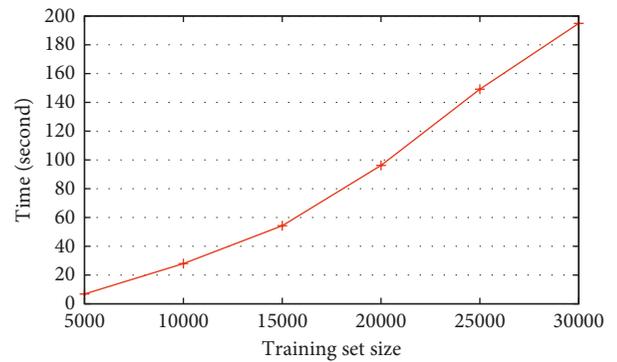


FIGURE 8: Average time spent on learning one model depending on the size of the training set.

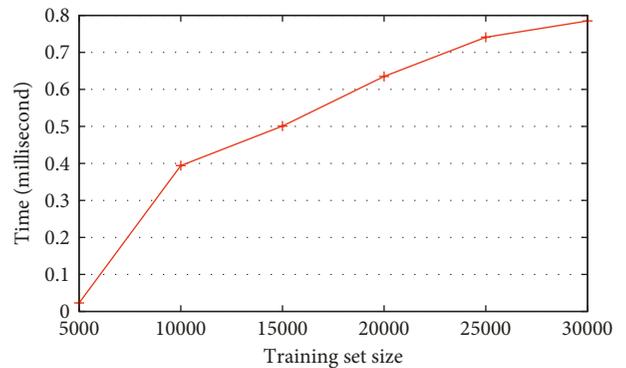


FIGURE 9: Time spent on determining one value depending on the size of the training set.

time. Theoretically, the time taken to test the SVR model is not affected by the size of the training set, but as the training set grows, the complexity of the model becomes larger (e.g., the number of support vectors increases), and the time required for the test also increases. However, as the number of samples increases, the increase in test time gradually decreases. The test time is expected not to increase after the number of samples reaches a certain point. Experiments on all observation stations using 30,000 samples took approximately 15 days on seven machines. Due to time limitations, we could not experiment with more samples, but there seems to be room for further performance improvement. In this

TABLE 13: Results of the proposed spatial quality control method.

Meteorological element	Normal (%)	Suspect (%)	Error (%)	Uninspected (%)
Wind direction	72.9	6.32	$6.31e^{-1}$	20.2
Wind speed	75.7	3.49	$6.56e^{-1}$	20.2
Temperature	93.8	$3.98e^{-1}$	$9.08e^{-2}$	5.67
Humidity	52.8	$2.08e^{-1}$	$5.33e^{-2}$	47.0
Atmospheric pressure	36.4	$4.75e^{-4}$	$1.60e^{-4}$	63.6
Hourly precipitation	87.1	$8.73e^{-1}$	2.99	9.04
Rainfall occurrence	89.2	1.38	$4.16e^{-1}$	9.04

study, all the stations were analyzed together, but the burden of the learning time would not be as great if each test were conducted separately for each observation station.

6.6. Result of Spatial Quality Control. Table 13 shows the results of applying the proposed spatial quality control procedure to actual data. As described in Section 2, the spatial quality control applies only to observations that are determined as normal during basic quality control. Therefore, values that did not pass the basic quality control are classed as uninspected during spatial quality control. The high ratio of uninspected observations of humidity and atmospheric pressure is due to the lack of measuring instruments for those elements in many observation stations.

7. Conclusion

In this study, we proposed a method to detect the abnormality of meteorological data using SVR. First, the value of the corresponding station was predicted using observations made in the surrounding area, and then any abnormality was detected by checking whether the observation differs from the predicted value outside of a predetermined range. SVR was used to create a model to predict the value of observation stations. In addition, we used MOGA to select SVR input variables to improve model performance and to reduce computation time.

Experiments on actual weather data collected by KMA show that using SVR is more accurate than the existing Cressman or Barnes methods for estimating the value of an observation station. Therefore, more accurate anomaly detection is possible through more accurate predictions. If the model can be learned in advance for a fixed cycle rather than learning the model every time, the proposed method has an acceptable execution time. A limitation of the method is that preaccumulated data are required, but it was confirmed through experiments that data collected over approximately one year provide sufficiently high performance.

In future study, the proposed method can be applied to additional meteorological elements such as sunshine duration and cloud height. Other valuable research could examine whether state-of-the-art learning techniques such as deep learning can yield more accurate predictions than SVR, which was not attempted here due to limitations of the system environment. In addition to accurate predictions, additional studies are required on the acceptable difference between the observation and the estimate which we set using the standard deviation. Furthermore, it will be interesting to

compare the anomaly detection technique with unsupervised learning technology as opposed to that based on prediction using supervised learning.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Research for the Forecasting Technology and Its Application, through the National Institute of Meteorological Sciences of Korea, in 2015 (NIMR-2012-B-1). This work was also partly supported by BK21 Plus for Pioneers in Innovative Computing (Department of Computer Science and Engineering, SNU) funded by the National Research Foundation of Korea (NRF) (21A20151113068), by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2017-0-01630) supervised by the IITP (Institute for Information & Communications Technology Promotion) and by a grant (KCG-01-2017-05) through the Disaster and Safety Management Institute funded by Korea Coast Guard of the Korean government. The Institute of Computer Technology (ICT) at the Seoul National University provides research facilities for this study.

References

- [1] J. H. Seo and Y. H. Kim, "Genetic feature selection for very short-term heavy rainfall prediction," in *Convergence and Hybrid Information Technology*, pp. 312–322, Springer, Berlin, Germany, 2012.
- [2] Y. H. Kim and Y. Yoon, "Spatiotemporal pattern networks of heavy rain among automatic weather stations and very-short-term heavy-rain prediction," *Advances in Meteorology*, vol. 2016, Article ID 4063632, 13 pages, 2016.
- [3] P. Cortez and A. D. J. R. Morais, "A data mining approach to predict forest fires using meteorological data," in *Proceedings of the 13th Portuguese Conference of Artificial Intelligence*, pp. 512–523, Guimarães, Portugal, December 2007.
- [4] A. Stoppa and U. Hess, "Design and use of weather derivatives in agricultural policies: the case of rainfall index insurance in Morocco," in *Proceedings of the International Conference on Agricultural Policy Reform and the WTO: Where are We Heading*, Capri, Italy, June 2003.
- [5] M. Kubik, P. J. Coker, J. F. Barlow, and C. Hunt, "A study into the accuracy of using meteorological wind data to estimate turbine generation output," *Renewable Energy*, vol. 51, pp. 153–158, 2013.

- [6] H. Yang, L. Lu, and J. Burnett, "Weather data and probability analysis of hybrid photovoltaic-wind power generation systems in hong kong," *Renewable Energy*, vol. 28, no. 11, pp. 1813–1824, 2003.
- [7] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," *Applied Energy*, vol. 67, no. 1, pp. 17–35, 2000.
- [8] M. K. Lee, S. H. Moon, Y. H. Kim, and B. R. Moon, "Correcting abnormalities in meteorological data by machine learning," in *Proceedings of the 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 888–893, San Diego, CA, USA, April 2014.
- [9] N. Y. Kim, Y. H. Kim, Y. Yoon, H. H. Im, R. K. Choi, and Y. H. Lee, "Correcting air-pressure data collected by mems sensors in smartphones," *Journal of Sensors*, vol. 2015, Article ID 245498, 10 pages, 2015.
- [10] Y.-H. Kim, J.-H. Ha, Y. Yoon et al., "Improved correction of atmospheric pressure data obtained by smartphones through machine learning," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 9467878, 12 pages, 2016.
- [11] J.-H. Ha, Y.-H. Kim, H.-H. Im, N.-Y. Kim, S. Sim, and Y. Yoon, "Error correction of meteorological data obtained with mini-AWSs based on machine learning," *Advances in Meteorology*, vol. 2018, Article ID 7210137, 8 pages, 2018.
- [12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.
- [13] J. Estévez, P. Gavilán, and J. V. Giráldez, "Guidelines on validation procedures for meteorological data from automatic weather stations," *Journal of Hydrology*, vol. 402, no. 1, pp. 144–154, 2011.
- [14] C. Daly, W. Gibson, M. Doggett, J. Smith, and G. Taylor, "A probabilistic-spatial approach to the quality control of climate observations," in *Proceedings of the 14th AMS Conference on Applied Climatology*, pp. 13–16, Seattle, WA, USA, January 2004.
- [15] 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, pp. 13–22, 2009.
- [16] G. P. Cressman, "An operational objective analysis system," *Monthly Weather Review*, vol. 87, no. 10, pp. 367–374, 1959.
- [17] 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.
- [18] J.-H. Seo, Y. H. Lee, and Y.-H. Kim, "Feature selection for very short-term heavy rainfall prediction using evolutionary computation," *Advances in Meteorology*, vol. 2014, Article ID 203545, 15 pages, 2014.
- [19] M. Jarraud, *Guide to Meteorological Instruments and Methods of Observation (wmo-no. 8)*, World Meteorological Organisation, Geneva, Switzerland, 2008.
- [20] V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.
- [21] H. Drucker, C. J. Burges, L. Kaufman et al., "Support vector regression machines," *Advances in Neural Information Processing Systems*, vol. 9, pp. 155–161, 1997.
- [22] T. Joachims, "SVMlight: support vector machine," 1999, <http://svmlight.joachims.org>.
- [23] P. J. Clark and F. C. Evans, "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, no. 4, pp. 445–453, 1954.
- [24] J. H. Holland, *Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, Cambridge, MA, USA, 1992.
- [25] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.
- [26] C. M. Fonseca and P. J. Fleming, "Multiobjective optimization and multiple constraint handling with evolutionary algorithms. I. A unified formulation," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 28, no. 1, pp. 26–37, 1998.
- [27] C. A. Coello, "An updated survey of GA-based multiobjective optimization techniques," *ACM Computing Surveys (CSUR)*, vol. 32, no. 2, pp. 109–143, 2000.
- [28] L. Xiujuan and S. Zhongke, "Overview of multi-objective optimization methods," *Journal of Systems Engineering and Electronics*, vol. 15, no. 2, pp. 142–146, 2004.
- [29] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: a tutorial," *Reliability Engineering and System Safety*, vol. 91, no. 9, pp. 992–1007, 2006.
- [30] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [31] T. Murata and H. Ishibuchi, "MOGA: Multi-objective genetic algorithms," in *Proceedings of the 1995 IEEE International Conference on Evolutionary Computation*, vol. 1, p. 289, Perth, Australia, 1995.
- [32] C. M. Fonseca and P. J. Fleming, "Multiobjective genetic algorithms," in *Proceedings of the IEE Colloquium on Genetic Algorithms for Control Systems Engineering (Digest No. 1993/130)*, London, UK, 1993.
- [33] H. Ishibuchi and T. Murata, "Multi-objective genetic local search algorithm," in *Proceedings of the IEEE International Conference on Evolutionary Computation*, pp. 119–124, Anchorage, AK, USA, 1996.
- [34] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [35] P. Allerup, H. Madsen, and F. Vejen, "A comprehensive model for correcting point precipitation," *Hydrology Research*, vol. 28, no. 1, pp. 1–20, 1997.



Hindawi

Submit your manuscripts at
www.hindawi.com

