Research Article

An Application of ANN Ensemble for Estimating of Precipitation Using Regional Climate Models

Dongwoo Jang

Department of Civil & Environmental Engineering, Incheon National University, Incheon 22012, Republic of Korea

Correspondence should be addressed to Dongwoo Jang; nightray@paran.com

Received 7 February 2020; Revised 15 January 2021; Accepted 9 February 2021; Published 9 March 2021

1. Introduction

Construction of future climate scenarios under global warming conditions, at a local or regional scale, is necessary for the assessment of climate change impacts on economic activities such as agriculture or energy production [1]. General circulation models (GCMs) represent the most satisfactory approach to predicting future climate changes [2, 3], but their present low spatial resolution measuring a few hundred kilometers makes their output problematic for use in impact studies [4, 5]. Various developed methods have been applied to downscale from GCM coarse resolution output to finer spatial scale (e.g., local stations or river catchment). Physically based numerical models such as regional dynamical models that work at a finer spatial scale are nested within a GCM [6]. Current applications of this method have revealed its capacity to reproduce fine-scale features of regional climates [7–9]. In spite of the substantial improvement in computer abilities in recent decades, this numerical model method remains computationally very demanding [10].

In recent years, a number of reports and papers within the meteorological community have adopted neural network as a method to downscale from large-scale atmospheric circulation to regional climate variables [11, 12]. Several application methods are developed for the purposes of constructing climate change scenarios [3, 10]. Other authors have applied ANN (artificial neural network) models to calculate problems on a seasonal scale [13–15] and to short-term (0 to 6 hourly) predictions [16]. Pattern recognition and Atlantic blocking detection using ANN are becoming gradually important [17, 18]. And there are various researches about advanced ANN such as M5 model tree (M5Tree), multivariable adaptive regression spline (MARS), least-square support vector regression (LSSVR), and gene-expression programming (GEP) models and various applications of ANN [19–25].

ANN models can be trained to establish the closest mathematical relationship between atmospheric circulation and the local climate, without predefined limitations. Thus, this method is able to capture some of the nonlinear relationships between the local climate and large-scale
circulation [1]. In particular, it has been shown that there is always a certain configuration of the multilayer perceptron that can arbitrarily approximate any continuous function [26, 27]. In this sense, it is helpful to regard the multilayer perceptron as a very powerful multiple regression technique [28]. Gardner and Dorling (1998) have shown that most atmospheric sciences ANN applications have employed standard feed-forward configuration of the multilayer perceptron [1, 29].

Downscaling methods are broadly classified as dynamic or statistical. Statistical downscaling is the development of statistical relationships between local climate variables and large-scale predictors. Dynamic downscaling is based upon nesting a finer scale regional climate model (RCM) of up to 10 km horizontal resolution within GCMs. Downscaling methods can be used to improve the spatial resolution of GCM output in order to overcome the challenge of assessing climate-related impacts on water resources at the catchment scale [30–32].

For the prediction of future climate conditions, global climate models were used. The grid resolution of current GCMs is in the order of hundreds of kilometers. However, the resolution required by environmental impact models is less than 10 km.

Therefore, as Tsanis et al. indicate, changes in precipitation and/or temperature predicted by GCMs for the Special Report on Emissions Scenarios (SRES) in small Mediterranean river basins cannot be used to assess extreme hydrological impacts, without previously applying a downscaling methodology. Two main downscaling techniques are available to represent this subgrid variability: (a) a dynamic approach using regional climate models (RCMs) and (b) a statistical approach [33].

Compared with RCMs, statistical methods include simpler computational procedures as well as the ability to produce results that can be directly compared with the regional observed data [34]. Therefore, statistical downscaling techniques have been widely used for climate impact studies [32, 35–39].

A method exists in which it is possible to apply an ensemble of several models in such a way as to reduce uncertainty in climate change scenarios [40]. This ensemble method uses a single-model ensemble (SME) that combines various initial conditions and a multimodel ensemble (MME) that combines multiple model results.

These include an ensemble method that averages the values of climate models, a Reliability Ensemble Average method that combines weights according to the accuracy of global models, and a method of using a median for each model. A method of synthesizing the values is also used [41–43].

In this study, observed precipitation data were compared with estimated precipitation data using the RCM. Via multilayer perceptron of the artificial neural network, it also helps to improve the accuracy of precipitation prediction by using an ANN ensemble for precipitation data generated from the RCM. For this purpose, data on RCM presented by the Korea Meteorological Administration and Coordinated Regional Climate Downscaling Experiment (CORDEX) were collected and applied to artificial neural networks.

Data for the artificial neural networks comprised the monthly precipitation of five RCMs used in Korea from 2006 to 2015. During that period, simulated precipitation by RCM at the same point as the observed rainfall was compared through the Pearson correlation coefficient in four major cities in Korea.

Through the coefficient of determination, the model closest to the actual precipitation was estimated, further estimating new precipitation by using the RCM precipitation data as an independent variable of the artificial neural network.

Based on the Pearson correlation coefficient, the optimal RCM for each city was selected. An ANN model was then built showing precipitation by the selected RCM as an independent variable. ANN models using five RCMs of four observatories were simulated, and ANN simulation cases were determined by the RCM’s combination in order of lowest determination coefficients. Accuracy assessment was performed in order to compare the precipitation data generated by the ANN model with the data actually observed.

2. Theoretical Background

Artificial neural network (ANN) was used as a statistical analysis method to find correlations between measured and simulated values of precipitation by the RCM. In addition, accuracy assessment was performed to compare the ANN-derived simulation results with measured data. Section 2 covers the theories of the statistical methods used.

2.1. ANN Analysis for Determining Precipitation from RCM.

The analysis procedure of the artificial neural network consists of input, hidden, and output layers, as shown in Figure 1. Learning and output are done through back-propagation algorithms.

According to Haykin (1994), a neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge that can then be made available for use. The neural network resembles the human brain in two respects: knowledge is acquired by the network through a learning process, and interneuron connection strengths (known as synaptic weights) are used to store the knowledge. The ANN procedure used is a feed-forward network type with input, hidden, and output layers [44].

Neurons in the input layer simply act as a buffer. Those in different layers are interconnected by means of weights. The neurons in the hidden and output layers are called the activation function, which when used here is a sigmoidal activation function. The input to each neuron \( j \) in the hidden layer is the sum of the weighted input signal \( x_i \) \( (\sum w_{ij}x_i = \text{net}_j, \text{in which } w_{ij} \text{ is the interconnecting weight between neuron } j \text{ in the hidden layer and neuron } i \text{ in the input layer}) \). Output \( y_j \) from the neuron is given by [44, 45]
2.2. Accuracy Assessment for Selecting Optimal ANN Model. In evaluating the accuracy of the leakage ratio using the ANN developed during research, an error ratio analysis was performed to evaluate the difference between the actual and model values. Accuracy assessment can be estimated by comparing the actual measured value with that generated by model simulation.

The mean absolute error (MAE), mean square error (MSE), BIAS, and G value were used to evaluate the accuracy of the estimation results. The calculation method of each equation is shown in the following equations:

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |z(x_i) - \bar{z}(x_i)|, \quad \text{(2)}
\]

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} [z(x_i) - \bar{z}(x_i)]^2, \quad \text{(3)}
\]

\[
\text{BIAS} = \frac{1}{n} \sum_{i=1}^{n} [z(x_i) - \bar{z}(x_i)], \quad \text{(4)}
\]

\[
G = \left(1 - \frac{\sum_{i=1}^{n} [z(x_i) - \bar{z}(x_i)]^2}{\sum_{i=1}^{n} [z(x_i) - \bar{z}]^2}\right) \times 100, \quad \text{(5)}
\]

where \(\bar{z}(x_i)\) is the estimated value at \(i\) and \(\bar{z}\) is the mean value of total usage data.

If the MAE and MSE are smaller, the estimated value indicates higher accuracy. The nearer the BIAS is to zero, the less biased the estimation result is. A G value of 100 indicates a perfect estimation. A negative G value indicates it as being less reliable than using an average of data values as a predictor [45].

2.3. Pearson Coefficient for Correlation Analysis. In statistics, the Pearson correlation coefficient (PCC), also referred to as Pearson’s \(r\), is a measure of the linear correlation between the variables \(X\) and \(Y\). It has a value between +1 and −1, where 1 is a total positive linear correlation, 0 is no linear correlation, and −1 is a total negative linear correlation. Widely used in the sciences, PCC was developed by Karl Pearson in the 1880s from a related idea introduced by Francis Galton [46–48].

Correlation analysis is a method of analyzing the linear relationship between two variables in probability theory and statistics. The two variables can be correlated with each other from an independent relationship, and the strength of the relationship between them is the correlation coefficient, as defined in

\[
r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \quad \text{(6)}
\]

where \(\bar{x}\) and \(\bar{y}\) are mean values of \(x\) and \(y\).

The PCC (\(r\)) is also used to determine the relationship between two variables with a coefficient of determination (\(r^2\)). The PCC is used to explain the linear relationship between the two variables. However, it may prove insufficient to determine the accuracy between the target and simulated values.

Though the results of the two variables are positively correlated, the linearity of the linear equation cannot be accurately estimated according to its slope and constant value, so an error rate analysis is performed when evaluating accuracy. If the linear relationship between the comparison variables is shown, it is concluded that the relationship between the two variables is closely related to the tendency of the trend line and the degree of convergence of the values [45].

3. Data Collection of Climate Change Scenarios and Model Building

3.1. Precipitation Data Collection Generated by RCMs

3.1.1. Provision of RCM for Generating Precipitation Data. The RCP scenarios in the IPCC’s fifth assessment report are input into climate models and are provided through the IPCC’s data distribution center. The development of global climate change scenarios (CMIP5) produces global data based on RCP scenarios. Based on global data, RCM regionally generated climate change scenarios are available in the local climate specification experiment (CORDEX) [49].

CORDEX calculates regional scenarios of climate change by dividing the world into 14 regions, including North America, Europe, Africa, and East Asia, at a resolution of
0.44° (about 50 km). With regard to Korea, data on climate change forecasts by the RCM can be used in CORDEX-EA for the East Asia region by HadGEM3-RA, RegCM4, SNU-MM5, SNU-WRF, and YSU-RSM [49].

Each website providing climate change scenarios provides data through various models, data periods, variables, and resolutions. This paper collected data from the CORDEX-EA and the websites of the Korea Global Atmosphere Watch Center (KGAWC) [50], which provides RCM results of several variables in RCP climate change scenarios.

The World Climate Research Program (WCRP) initiated the CORDEX framework in order to improve worldwide regional climate change projections and provide a framework for better coordination of regional climate downscaling [51].

The CORDEX experiments provided an opportunity to evaluate the relative and absolute performances of various RCMs over predefined regions. As one of 14 branch domains within the CORDEX framework, CORDEX-East Asia (CORDEX-EA) covers a large area of East Asia with a horizontal resolution of approximately 50 km [52].

In the framework of the CORDEX project, five RCMs driven by ERA-Interim reanalysis were evaluated in terms of their ability to simulate the climatology of the East Asia domain (CORDEX-EA) in comparison with the Aphrodite observational data set [53]. The models are affected by a temperature bias ranging from 1.1 (RegCM4) to 0.2°C (SNU-MM5) from June to August and from 5.5 (SNU-MM5) to 0.9°C (RegCM4) from December to February [53].

3.1.2. Collection of Precipitation Data from RCM. To compare results of the RCM and actual data in order to determine climate model forecasting accuracy, data were selected for use based on the status of the climate change scenario. The climate change scenarios collected are shown in Table 1.

Data for the same period are required to compare climate change scenarios and observations. Thus, 50 km resolution precipitation data for 2006–15 overlapping with the previous period’s data were collected from the HadGEM3-RA, RegCM4, SNU-WRF, SNU-MM5, and CORDEX-EA YSU-RS prediction data. Only monthly SUN-WRF data were available from 2006 to 2010. Daily data of SNU-WRF were thus collected and then converted into monthly data from 2011 to 2015. So monthly data by all RCMs were collected for ANN simulation.

The collected RCM data were interpolated using Universal Kriging, the optimal spatial interpolation method available to collect data at the same location as the actual observatory site [49]. The interpolated precipitation data are used in the study as shown in Section 3.3.

3.2. Observation Data Collection for Comparison with RCM. Precipitation data are available through the KMA website and Korea Global Atmosphere Water Center [54]. Ground observation data are classified into the automated synoptic observation system (ASOS), automatic weather system (AWS), and automated agricultural observing system (AAOS). Data are provided in CSV and PDF formats.

ASOS comprises ground observations performed at the same time on all observatories in order to determine the atmospheric conditions at any given time. Air pressure, temperature, wind direction, wind speed, relative humidity, precipitation, solar radiation, daylight time, ground temperature, vertical temperature, and ground temperature are all automatically observed. Each observation station is measuring the precipitation by one-minute time unit. It is currently being observed at more than 80 sites in Korea [54].

Observed precipitation data for 2006–2015, consistent with the RCM data’s period, were collected [46]. Observation points were selected from the synoptic meteorological network, where no data were missing. No movement of precipitation observation equipment was seen over the measurement period.

Location information on 4 selected observation points of precipitation data collected through the Korea Global Atmosphere Water Center is shown in Figure 2.

3.3. Status of Collected Data. In all four cities, monthly precipitation by RCMs and observed precipitation patterns were similar. In the case of Seoul, observed precipitation in the summer season was higher than the RCM simulation. In the case of Busan, RCM simulated precipitation was frequently higher than the observed precipitation.

Park (2017) analyzed five RCMs in Korea and found a similar correlation between observed and RCM precipitation, whether located inland or in northern parts [49]. Figure 3 shows that precipitation forecasting patterns for Seoul and Daejeon, both situated further north than Daegu and Busan, are more stable. As shown in the precipitation comparison graph, the monthly precipitation of five RCMs demonstrates a differing pattern.

Table 2 shows the Pearson coefficients for the actual precipitation and the RCM simulated precipitation in four major Korean cities. As can be seen in the table, all have a positive correlation of 0.2–0.6 for each model.

In particular, the Busan area has a low Pearson coefficient except for the HadGEM3-RA model. Except for the HadGEM3-RA model, in Busan, it is found that the correlation with observed precipitation is low. Correlation between the three cities also varies depending on the model used. Of the five RCMs, the HadGEM3-RA model generally shows a high correlation, while the YSU-RSM shows a high correlation in Daejeon and Daegu.

4. ANN Ensemble Analysis

4.1. Simulation Model Construction. ANN simulation cases were selected to produce ensemble precipitation forecasting using five RCMs in four major cities in Korea. Observed precipitation was used as a dependent variable while 5 RCM precipitation data were used as independent variables.

In constructing an ANN simulation, 80% of total data are trained and 20% verified to simulate the constructed neural network. The hyperbolic tangent function is used for the
Table 1: Data collection from RCM [49].

| Classification                        | Contents                                                                 |
|---------------------------------------|--------------------------------------------------------------------------|
| Data source                           | CORDEX-EA                                                                |
| Climate change scenario               | RCP 8.5                                                                  |
| Regional climate model                | HadGEM3-RA, SNU-WRF, RegCM4, SNU-MM5, and YSU-RSM                      |
| Spatial range                         | East Asia                                                                |
| Time range (year)                     | 2006–2015                                                                |
| Spatial and time resolution           | 50 km/monthly                                                            |
| Variable                              | Precipitation                                                            |
| Data format                           | NetCDF(.nc)                                                              |

Figure 2: Locations of 4 major cities in Korea.

Figure 3: Continued.
activation function and the conjugate gradient method is used as the optimization algorithm. The conjugate gradient method is an algorithm for finding the nearest local minimum of a function of $n$ variables which presupposes that the gradient of the function can be computed [55]. The minimum variation of the learning error is 0.0001 in the shelter.
multilayer perceptron of ANN, and the minimum relative variation of the learning error is 0.001. The results are repeatedly calculated until the error is minimized.

ANN simulation cases were selected according to the coefficient of determination calculated in Table 2, using five RCMs. The RCM independent variables were removed one by one in descending order of the coefficient of determination. Five cases were then selected as shown in Table 3 for each city.

The number of neurons in the ANN model was set to twice the independent variable, that is, the optimal stimulation condition, while the number of independent variables was taken into account [44, 45].

### 4.2. ANN Ensemble Results Using RCMs

#### 4.2.1. Accuracy Assessment Results.

Using ANN, the most reliable case for estimating the precipitation of 4 major cities was selected. Accuracy assessment was determined by using the MAE, MSE, BIAS, and G value (goodness of prediction). Figure 4 shows the accuracy assessment results for all simulation cases. Where the calculated value of the MAE and MSE is small and the G value close to 100, the simulated value is therefore corrected to the measured value.

The results of the ANN ensemble are more accurate than the results of RCM estimated precipitation. In all cities, the precipitation calculated by ANN is lower in the MAE and MSE values than the RCM results.

Additionally, since ANN simulation results show that the G value is closer to 100 than the conventional RCM, the ANN can be considered to be effective in precipitation prediction.

However, it is difficult to use as an estimation model if the G value is negative. All ANN simulation results show positive values, while some RCMs show negative values.

In the case of BIAS, RCMs demonstrate bias according to location and model, whereas the ANN ensemble model is relatively unbiased. Detailed results of the ANN ensemble accuracy assessment are shown in Table 4. Pearson co. (avr.) is averaged PCC for selected RCMs.

As shown in Figure 4, HadGEM3-RA for Seoul and Busan, YSU-RSM for Daejeon, and RegCM4 for Daegu were selected as the most accurate RCMs. Therefore, various RCMs are required for accurate model selection in each region. Pearson coefficient by ANN is higher than that by averaged precipitation by RCMs in all cases.

In Table 4, the most accurate ANN simulation cases were selected by analyzing MAE, MSE, and G values, which are the comparison indices of measured and predicted values. When analyzing ANN using 4 or 5 RCMs, results using ANN showed the highest accuracy.

In the case of Seoul, accuracy was high except when using the YSU-RSM model, which had the lowest Pearson coefficient of 0.31 in Table 2. In the case of Daejeon, accuracy was high except when using the SNU-MM5 model, and in the case of Busan, accuracy was high except when using the RegCM4 model which had the lowest Pearson coefficient. In the case of Daegu, cases using all RCMs were the most accurate.

#### 4.2.2. Graph Analysis.

Figure 5 shows ANN ensemble simulation results for the optimal cases selected in Table 4. And, Table 5 is a basic statistical analysis for precipitations from Figure 5. ANN ensemble was determined through the results such as MAE, MSE, and G values. It was compared with a case where the correlation was high among the 5 RCMs.

In Seoul, Case 1-2 is the most accurate. It uses 4 RCMs, not including the YSU-RSM model. Certain monthly estimated precipitation in Seoul is higher than 1,000 mm. The amount of precipitation corresponding to annual rainfall is simulated. The HadGEM3-RA had the highest accuracy among the five RCMs used. However, it was not reliable in predicting extreme precipitation. Instead, predictions of monthly rainfall patterns proved better than in other cities.
Table 3: ANN simulation cases.

| Location | RCMs | Number of neurons | Modeling case |
|----------|------|-------------------|---------------|
| Seoul    | HadGEM3-RA, RegCM4, SNU-WRF, SNU-MM5, and YSU-RSM | 10 | Case 1-1 |
|          | HadGEM3-RA, RegCM4, SNU-WRF, and SNU-MM5 | 8  | Case 1-2 |
|          | HadGEM3-RA, RegCM4, and YSU-WRF | 6  | Case 1-3 |
|          | HadGEM3-RA and RegCM4 | 4  | Case 1-4 |
|          | HadGEM3-RA | 2  | Case 1-5 |
| Daejeon  | HadGEM3-RA, RegCM4, SNU-WRF, SNU-MM5, and YSU-RSM | 10 | Case 2-1 |
|          | HadGEM3-RA, RegCM4, SNU-WRF, and YSU-RSM | 8  | Case 2-2 |
|          | HadGEM3-RA, SNU-WRF, and YSU-RSM | 6  | Case 2-3 |
|          | HadGEM3-RA and YSU-RSM | 4  | Case 2-4 |
|          | YSU-RSM | 2  | Case 2-5 |
| Daegu    | HadGEM3-RA, RegCM4, SNU-WRF, SNU-MM5, and YSU-RSM | 10 | Case 3-1 |
|          | RegCM4, SNU-MM5, YSU-RSM | 6  | Case 3-2 |
|          | RegCM4, and YSU-RSM | 4  | Case 3-3 |
|          | YSU-RSM | 2  | Case 3-4 |
| Busan    | HadGEM3-RA, RegCM4, SNU-WRF, SNU-MM5, and YSU-RSM | 10 | Case 4-1 |
|          | HadGEM3-RA, SNU-WRF, SNU-MM5, and YSU-RSM | 8  | Case 4-2 |
|          | HadGEM3-RA, SNU-WRF, and YSU-RSM | 6  | Case 4-3 |
|          | HadGEM3-RA and SNU-WRF | 4  | Case 4-4 |
|          | HadGEM3-RA | 2  | Case 4-5 |

Figure 4: Continued.
Figure 4: Results of accuracy assessment between ANN simulation and RCMs: (a) MAE, (b) MSE, (c) G value, and (d) BIAS.

Table 4: Accuracy assessment results by ANN simulation.

| Location | Selected RCMs                        | Cases       | MAE  | MSE  | BIAS | G     | Pearson co. (ANN) | Pearson co. (avr.) | Optimal model |
|----------|--------------------------------------|-------------|------|------|------|-------|-------------------|-------------------|---------------|
| Seoul    | HadGEM3-RA, RegCM4, SNU-WRF, SNU-MM5, and YSU-RSM | Case 1-1    | 81.7 | 22,373 | 11.2 | 35.8  | 0.60              | 0.36              |               |
|          | HadGEM3-RA, RegCM4, SNU-WRF, and SNU-MM5 | Case 1-2    | 70.1 | 17,288 | -0.1 | 50.4  | 0.72              | 0.35              | O             |
|          | HadGEM3-RA, RegCM4, and SNU-WRF | Case 1-3    | 79.5 | 21,290 | 1.1  | 38.9  | 0.62              | 0.38              |               |
|          | HadGEM3-RA and RegCM4 | Case 1-4    | 76.1 | 26,456 | 34.5 | 24.1  | 0.57              | 0.46              |               |
|          | HadGEM3-RA | Case 1-5    | 85.2 | 24,656 | 13.8 | 29.3  | 0.55              | 0.56              |               |
| Daejeon  | HadGEM3-RA, SNU-WRF, and YSU-RSM | Case 2-1    | 57.2 | 8,048  | 6.3  | 45.5  | 0.68              | 0.39              |               |
|          | HadGEM3-RA, RegCM4, SNU-WRF, and YSU-RSM | Case 2-2    | 58.7 | 7,871  | -2.6 | 46.7  | 0.69              | 0.41              | O             |
|          | HadGEM3-RA, SNU-WRF, and YSU-RSM | Case 2-3    | 61.9 | 8,579  | -2.4 | 41.9  | 0.65              | 0.44              |               |
|          | HadGEM3-RA and YSU-RSM | Case 2-4    | 60.7 | 8,172  | -1.9 | 44.7  | 0.67              | 0.51              |               |
|          | YSU-RSM | Case 2-5    | 63.0 | 9,160  | 1.2  | 38.0  | 0.62              | 0.45              |               |
Table 4: Continued.

| Location | Selected RCMs | Cases | MAE   | MSE   | BIAS  | G     | Pearson co. (ANN) | Pearson co. (avr.) | Optimal model |
|----------|---------------|-------|-------|-------|-------|-------|-------------------|--------------------|---------------|
| Daegu    | RegCM4, SNU-MM5, and YSU-RSM | Case 3-1 | 52.0  | 5,527 | 0.9   | 46.8  | 0.69              | 0.41               | O             |
|          |               | Case 3-2 | 53.4  | 6,845 | −2.1  | 34.2  | 0.59              | 0.41               |               |
|          | RegCM4 and YSU-RSM | Case 3-3 | 52.2  | 6,933 | 4.6   | 33.3  | 0.58              | 0.37               |               |
|          |               | Case 3-4 | 53.5  | 6,948 | 1.1   | 33.2  | 0.58              | 0.36               |               |
|          |               | Case 3-5 | 56.6  | 7,150 | 2.9   | 31.2  | 0.56              | 0.35               |               |

| Busan    | HadGEM3-RA, SNU-WRF, SNU-MM5, and YSU-RSM | Case 4-1 | 66.4  | 12,038 | 6.8   | 29.3  | 0.54              | 0.48               | O             |
|          | HadGEM3-RA, SNU-WRF, SNU-MM5, and YSU-RSM | Case 4-2 | 65.4  | 10,360 | 7.8   | 39.1  | 0.63              | 0.49               |               |
|          | HadGEM3-RA and SNU-WRF | Case 4-3 | 64.7  | 10,446 | 3.5   | 38.6  | 0.63              | 0.49               |               |
|          |              | Case 4-4 | 69.9  | 12,197 | −5.0  | 28.3  | 0.53              | 0.51               |               |
|          |       | Case 4-5 | 71.2  | 13,701 | 0.8   | 19.5  | 0.44              | 0.49               |               |

Figure 5: Continued.
As a result of the ANN ensemble, there was a pattern similar to the HadGEM3-RA precipitation model which was the optimal RCM for Seoul. In the case of Daejeon and Daegu, the YSU-RSM model achieved the results closest to the actually observed precipitation. In certain periods, precipitation by RCM was overestimated. The model selected through the ANN ensemble was adjusted to be close to the observed precipitation.

**Figure 5**: Precipitation comparison between ANN simulation and RCM with observed value: (a) Seoul, (b) Daejeon, (c) Daegu, and (d) Busan.

**Table 5**: Basic precipitation comparison (unit: mm).

| Location | Selected RCMs    | Max.  | Min.  | Ave.   |
|----------|------------------|-------|-------|--------|
| Seoul    | Observation      | 1,131.0 | 0.8   | 121.2  |
|          | HadGEM3-RA       | 457.1 | 2.3   | 85.7   |
|          | ANN ensemble     | 530.8 | 0.0   | 121.3  |
| Daejeon  | Observation      | 587.3 | 2.5   | 107.6  |
|          | HadGEM3-RA       | 527.5 | 2.9   | 78.3   |
|          | ANN ensemble     | 391.6 | 30.6  | 110.1  |
| Daegu    | Observation      | 568.1 | 0     | 87.5   |
|          | HadGEM3-RA       | 359.8 | 0.4   | 81.7   |
|          | ANN ensemble     | 379.8 | 7.0   | 86.6   |
| Busan    | Observation      | 886.1 | 0     | 123.9  |
|          | HadGEM3-RA       | 452.5 | 1.6   | 97.8   |
|          | ANN ensemble     | 316.9 | 10.4  | 116.1  |
In Busan, the ANN ensemble shows a similar pattern to the optimal RCM. However, it was overestimated in certain periods. The average monthly precipitation in Busan is 124 mm, and the average monthly precipitation by HadGEM3-RA is 98 mm, which is underestimated compared with observed data in Table 5. The ANN ensemble is 116 mm, which is closer to the observed rainfall than HadGEM3-RA. However, HadGEM3-RA was 452 mm and ANN ensemble was 316 mm compared with the maximum monthly rainfall of 886 mm. Even during periods when no precipitation was observed, a certain amount of precipitation was predicted in the model.

The results showed that the ANN ensemble predicted precipitation results closer to the observed precipitation than the single used RCM. In particular, the ANN ensemble produced more reliable patterns than overestimating RCM results in inland cities such as Daejeon and Daegu.

5. **Comparison between ANN and Multiple Regression Analysis**

In Section 5, the multiple regression analysis (MRA) was calculated and compared with the ANN results. MRA is one of the most widely used statistical techniques. The ANN simulation results of target cities were compared using the various statistical analysis methods.

5.1. **Generation of Equation by Multiple Regression Analysis**

Multiple regression analysis (MRA) is used for identifying relationships between variables with high correlations and predict the value of variables. In the regression model, one independent variable is called simple regression analysis, and one independent variable does not sufficiently explain the dependent variable.

The multiple linear regression model with independent variables is expressed as [56, 57]

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k,
\]

where \( x_i (i = 1, \ldots, k) \) is the independent variable (precipitation by RCMs), \( y \) is the dependent variable (observed precipitation), \( \beta_i (i = 1, \ldots, k) \) is the regression coefficient, and \( \beta_0 \) is the intercept of \( y \).

MRA was performed on selected appropriate ANN simulation cases of each city. Selected cases are Case 1-2, Case 2-2, Case 3-1, and Case 4-2. Using the 4 cases, generated MRA equations are shown in equations (8)–(11). The precipitation unit is mm per month for 4 cities.

\[
\begin{align*}
\text{Prec}_{Seoul} &= -30.6 + 0.94 \text{HadGEM} + 1.26 \text{RegCM4} + 0.26 \text{SNUWRF} - 0.54 \text{SNUMM5}, \\
\text{Prec}_{Daegon} &= 11.2 + 0.39 \text{HadGEM} + 0.44 \text{RegCM4} - 0.27 \text{SNUWRF} + 0.50 \text{YSURSM}, \\
\text{Prec}_{Daegu} &= 5.24 + 0.23 \text{HadGEM} + 0.65 \text{RegCM4} - 0.41 \text{SNUWRF} + 0.002 \text{SNUMM5} + 0.36 \text{YSURSM}, \\
\text{Prec}_{Busan} &= 50.3 + 0.53 \text{HadGEM} - 0.21 \text{SNUWRF} + 0.21 \text{SNUMM5} + 0.17 \text{YSURSM}.
\end{align*}
\]

5.2. **Comparison between ANN and Multiple Regression Analysis**

Multiple regression equations from (8) to (11) were used to generate precipitation; then these results were compared with observed data in Table 6 through the accuracy assessment indices.

As shown in Table 6, the prediction estimation accuracy of ANN was higher than MRA in all analyzed cities. MAE and MSE values are low and the G value is close to 100, which is the result of judging that the estimation accuracy is high.

In the case of MRA, the BIAS value is 0 in all cities, and the estimated value is not biased compared to ANN. As a result of comparing ANN and MRA with MAE, MSE, and G values, ANN ensembles in all cities showed appropriate results to observations in all cities.

Figure 6 is a graph comparing ANN and MRA to actual precipitation. Estimated precipitation was underestimated in all four cities. Overall ANN and MRA precipitation distributions were similar. The results show that the accuracy of precipitation estimation using ANN and MRA is higher than that of RCM’s precipitation, but the accuracy is still lower than the observed precipitation in all analyzed cities.
Table 6: Accuracy assessment results between ANN and MRA.

| Classification | Seoul | Daejeon | Daegu | Busan |
|----------------|-------|---------|-------|-------|
|                | MRA   | ANN     | MRA   | ANN   | MRA   | ANN   | MRA   | ANN   |
| BIAS           | 0.0   | -0.1    | 0.0   | -2.6  | 0.0   | 0.9   | 0.0   | 6.8   |
| MAE            | 79.5  | 70.1    | 58.5  | 58.7  | 53.2  | 52.0  | 71.7  | 66.4  |
| MSE            | 21,863| 17,288  | 8165.6| 7,871 | 5,971 | 5,527 | 13,528| 12,038|
| G              | 37.3  | 50.4    | 44.7  | 46.7  | 42.6  | 46.8  | 20.5  | 29.3  |
| Pearson co.    | 0.61  | 0.72    | 0.67  | 0.69  | 0.70  | 0.69  | 0.50  | 0.54  |
| Optimal model  | O     | O       | O     | O     | O     | O     | O     | O     |

Figure 6: Precipitation comparison between ANN and MRA with observed value: (a) Seoul, (b) Daejeon, (c) Daegu, and (d) Busan.
6. Conclusions

This study is an application of ANN in order to improve the accuracy of estimated precipitation according to climate change scenarios in Korea. For this purpose, RCM precipitation data were collected from CORDEX and the Korea Meteorological Administration. ANN was applied to major cities in Korea and the following conclusions were drawn.

The five RCMs provided in Korea were compared with the observed precipitation, and the ANN ensemble model was applied to combine the advantages of each model. The results show that using the ANN ensemble simulation estimates observed precipitation appropriately compared to a single RCM. In addition, irregular precipitation patterns in single RCM results have steadily changed by using the ensemble model.

In applying ANN to predicting the precipitation of major cities in Korea, it is most accurate to apply 4 or 5 RCMs as independent variables. When ANN is applied, optimal weighting factors should be adjusted for low and high correlation models.

In the time series analysis, the ANN ensembles are closer to the observed precipitation pattern than the single used RCM. In particular, it is shown that precipitation pattern forecasting is more effective in inland areas such as Daejeon and Daegu than coastal areas such as Busan, where the precipitation pattern is more irregular.

ANN and MRA were compared by the RCM ensemble. As a result, ANN was selected as the accurate model in representative cities in Korea. However, it is necessary to improve the accuracy in quantitatively evaluating the precipitation.

An RCM of 50 km resolution was extracted by the spatial interpolation method and used for this study. Estimated precipitation may vary according to the statistical downscaling method using RCM, so that different optimal ANN models may be selected by different downscaling results. A more detailed analysis will be possible if applied to other regions in the future, and the applicability of the ANN model will see growth following the collection of more observation data. To increase model accuracy, M5Tree, MARS, LSSVR, and GEP models can be applied in future research. Also, this study was conducted on four major cities and needs to be supplemented through research in various domestic or overseas regions with different city characteristics in the future.

Since it is not possible to predict optimal precipitation using all RCMs, an additional RCM selection methodology is required for ANN application. The study classifies simulation cases using Pearson correlation coefficients, but further studies are needed to select the optimal ANN cases.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the Incheon National University (International Cooperative) Research Grant in 2019.

References

[1] R. Trigo and J. Palutikof, "Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach," Climate Research, vol. 13, pp. 45–59, 1999.
[2] T. M. L. Wigley, P. D. Jones, K. R. Briffa, and G. Smith, "Obtaining sub-grid-scale information from coarse-resolution general circulation model output," Journal of Geophysical Research, vol. 95, no. D2, pp. 1943–1953, 1990.
[3] R. G. Crane and B. C. Hewitson, "Doubled CO2 precipitation changes for the Susquehanna Basin: down-scaling from the Genesis general circulation model," International Journal of Climatology, vol. 18, no. 1, pp. 65–76, 1998.
[4] T. R. Karl and W.-C. Wang, M. E. Schlesinger, M. E. Schlesinger, R. W. Knight, and D. Portman, "A method of relating general circulation model simulated climate to the observed local climate. Part I: seasonal statistics," Journal of Climate, vol. 3, no. 10, pp. 1053–1079, 1990.
[5] J. A. Winkler, J. P. Palutikof, J. A. Andresen, and C. M. Goodess, "The simulation of daily temperature time series from GCM output. Part II: sensitivity analysis of an empirical transfer function methodology," Journal of Climate, vol. 10, no. 10, pp. 2514–2532, 1997.
[6] F. Giorgi, "Simulation of regional climate using a limited area model nested in a general circulation model," Journal of Climate, vol. 3, no. 9, pp. 941–963, 1990.
[7] R. G. Jones, J. M. Murphy, and M. Noguer, "Simulation of climate change over Europe using a nested regional-climate model. I: assessment of control climate, including sensitivity to location of lateral boundaries," Quarterly Journal of the
Advances in Civil Engineering

Royal Meteorological Society, vol. 121, no. 526, pp. 1413–1449, 1995.

[8] L. Mearns, F. Giorgi, L. McDaniel, and C. Shields, "Analysis of variability and diurnal range of daily temperature in a nested regional climate model: comparison with observations and doubled CO2 results," Climate Dynamics, vol. 11, no. 4, pp. 193–209, 1995.

[9] J. W. Kidson and C. S. Thompson, "A comparison of statistical and model-based downscaling techniques for estimating local climate variations," Journal of Climate, vol. 11, no. 4, pp. 735–753, 1998.

[10] B. Hewitson and R. Crane, "Climate downscaling: techniques and application," Climate Research, vol. 7, pp. 85–95, 1996.

[11] B. C. Hewitson and R. G. Crane, "Large-scale Atmospheric controls on local precipitation in tropical Mexico," Geophysical Research Letters, vol. 19, no. 18, pp. 1835–1838, 1992.

[12] T. Cavazos, "Downscaling large-scale circulation to local water rainfall in north-eastern Mexico," International Journal of Climatology, vol. 17, no. 10, pp. 1069–1082, 1997.

[13] H. Navone and H. Cuccato, "Predicting Indian monsoon rainfall: a neural network approach," Climate Dynamics, vol. 10, no. 6–7, pp. 305–312, 1994.

[14] B. Tang, G. M. Flato, and G. Holloway, "A study of arctic sea ice and sea-level pressure using POP and neural network methods," Atmosphere-Ocean, vol. 32, no. 3, pp. 507–529, 1994.

[15] F. T. Tangang, B. Tang, A. H. Monahan, and W. W. Hsieh, "Forecasting ENSO events: a neural network-extended EOF approach," Journal of Climate, vol. 11, no. 1, pp. 29–41, 1998.

[16] R. J. Kuligowski and A. P. Barros, "Experiments in short-term precipitation forecasting using artificial neural networks," Monthly Weather Review, vol. 126, no. 2, pp. 470–482, 1998.

[17] J. E. Peak and P. M. Tang, "Towards automated interpolation of satellite imagery for nvy shipboard applications," Bull Am Meteoro! Soc, vol. 73, pp. 955–1008, 1992.

[18] M. Verdecchia, G. Visconti, F. D’Andrea, and S. Tibaldi, "A neural network approach for blocking recognition," Geophysical Research Letters, vol. 23, no. 16, pp. 2081–2084, 1996.

[19] Y. Xiang, L. Gou, L. He, S. Xia, and W. Wang, "A SVR-ANN combined model based on ensemble EMD for rainfall prediction," Applied Soft Computing, vol. 73, pp. 874–883, 2018.

[20] R. Mirabbasi, O. Kisi, H. Sanikhani, and S. Gajbhiye Meshram, "Monthly long-term rainfall estimation in Central India using M5Tree, MARS, LSSVR, ANN and GEP models," Neural Computing and Applications, vol. 31, no. 10, pp. 6843–6862, 2019.

[21] M. Rezaie-Balf, S. Kim, H. Fallah, and S. Alaghmand, "Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: application on the perennial rivers in Iran and South Korea," Journal of Hydrology, vol. 572, pp. 470–485, 2019.

[22] S. O. Sulaiman, J. Shiri, H. Shiralizadeh, O. Kisi, and Z. M. Yaseen, "Precipitation pattern modeling using cross-station perception: regional investigation," Environmental Earth Science, vol. 77, p. 709, 2018.

[23] M. Nazafzadeh, A. Ghaemi, and S. Emamgholizadeh, "Prediction of water quality parameters using evolutionary computing-based formulations," International Journal of Environmental Sciences and Technology, vol. 16, pp. 6377–6396, 2019.

[24] M. Zeinolabedini and M. Najafzadeh, "Comparative study of different wavelet-based neural network models to predict sewage sludge quantity in wastewater treatment plant," Environmental Monitoring and Assessment, vol. 191, p. 163, 2019.

[25] O. Kisi, S. Mohsenzadeh Karimi, J. Shiri, and A. Keshavarzi, "Modelling long term monthly rainfall using geographical inputs: assessing heuristic and geostatistical models," Meteorological Applications, vol. 26, no. 4, pp. 698–710, 2019.

[26] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Networks, vol. 2, no. 5, pp. 359–369, 1988.

[27] M. T. Hagan, H. B. Demuth, and M. H. Beale, Neural Networks Design, PWS Publishing Company, Boston, MA, USA, 1996.

[28] W. S. Sarle, "Neural networks and statistical models," in Proceedings of the 19th Annual SAS Users Group International Conference, SAS Institute, Cary, NC, USA, April 1994.

[29] M. Gardner and S. R. Dorling, "Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London," Atmospheric Environment, vol. 33, no. 5, pp. 709–719, 1999.

[30] S. Mandal, R. K. Srivastav, and S. P. Simonovic, "Use of beta regression for statistical downscaling of precipitation in the campbell river basin, British columbia, Canada," Journal of Hydrology, vol. 538, pp. 49–62, 2016.

[31] A. W. Wood, L. R. Leung, V. Sridhar, and D. P. Lettenmaier, "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs," Climatic Change, vol. 62, no. 1-3, pp. 189–216, 2004.

[32] N. N. Kourgialas, Z. Dokou, and G. P. Karatzas, "Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: the example of a small Mediterranean agro-watershed," Journal of Environmental Management, vol. 154, pp. 86–101, 2015.

[33] I. K. Tsisanis, A. G. Koutoulis, I. N. Daliakopoulos, and D. Jacob, "Severe climate-induced water shortage and extremes in Crete," Climatic Change, vol. 106, no. 4, pp. 667–677, 2011.

[34] L. Vasiliadis, A. Loukas, and G. Patsonas, "Evaluation of a statistical downscaling procedure for the estimation of climate change impacts on droughts," Natural Hazards and Earth System Sciences, vol. 9, no. 3, pp. 879–894, 2009.

[35] A. Bardossy, I. Bogardi, and I. Matyasovszky, "Fuzzy rule-based downscaling of precipitation," Theoretical and Applied Climatology, vol. 82, no. 1-2, pp. 119–129, 2005.

[36] O. Varis, T. Kajander, and R. Lemmeli, "Climate and water: from climate models to water resources management and vice versa," Climatic Change, vol. 66, no. 3, pp. 321–344, 2004.

[37] H. J. Fowler, S. Blenkinsop, and C. Tebaldi, "Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling," International Journal of Climatology, vol. 27, no. 12, pp. 1547–1578, 2007.

[38] A. Loukas, L. Vasiliadis, and J. Tzabiras, "Climate change effects on drought severity," Advances in Geosciences, vol. 17, pp. 23–29, 2008.

[39] B. Timbal, E. Fernandez, and Z. Li, "Generalization of a statistical downscaling model to provide local climate change projections for Australia," Environmental Modelling & Software, vol. 24, no. 3, pp. 341–358, 2009.

[40] S. A. Lee, S.-H. Lee, S.-Y. Ji, and J. Choi, "Predicting change of suitable plantation of schisandra chinensis with ensemble of climate change scenario," Journal of Environmental Impact Assessment, vol. 25, no. 1, pp. 77–87, 2016, In Korean.

[41] Y. H. Byun, J. K. Park, and H. S. Jung, "A comparison of ensemble method between models in prediction of season," Journal of Meteorological Society, vol. 12, pp. 1–3, 2002, In Korean.
[42] Y. D. Kim, "New trend of research in reduction of uncertainty in effect assessment of climate change," *Journal of Water Resources Association*, vol. 43, pp. 32–36, 2010, In Korean.

[43] Korea Meteorological Administration, *Catchable Forecasting Technology*, Korea Meteorological Administration, Seoul, South Korea, 2012, In Korean.

[44] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan, New York, NY, USA, 1994.

[45] D. W. Jang, *Estimation of non-Revenue Water ratio using PCA and ANN in water distribution systems*, Ph.D. Thesis, Incheon National University, Incheon, Korea, 2017.

[46] F. Galton, "The British association: section II, anthropology: opening address by Francis Galton, F.R.S., etc., president of the anthropological institute, president of the section," *Nature*, vol. 32, pp. 507–510, 1885.

[47] K. Pearson, "Notes on regression and inheritance in the case of two parents," in *Proceedings of the Royal Society of London, Royal Society*, vol. 58, pp. 240–242, 1895.

[48] S. M. Stigler, "Francis Galton’s account of the invention of correlation," *Statistical Science*, Institute of Mathematical Statistics, vol. 4, no. 2, pp. 73–79, 1989.

[49] H. S. Park, *Improving the applicability of precipitation forecasting considering climate change in Korea*, Ph.D. Thesis, Incheon National University, Incheon, Korea, 2017.

[50] Korea Global Atmosphere Watch Center, *KGAWC*, 2016, http://www.climate.go.kr/home/Eng/htmls/kgawc/sub1.html.

[51] F. Giorgi, J. Jones, and G. R. Asrar, "Addressing climate information needs at the regional level: the CORDEX framework," *WMO Bull.*, vol. 58, pp. 175–183, 2009.

[52] W. Zhou, J. Tang, X. Wang, S. Wang, X. Niu, and Y. Wang, "Evaluation of regional climate simulations over the CORDEX-EA-II domain using the COSMO-CLM model," *Asia-Pacific Journal of Atmospheric Sciences*, vol. 52, no. 2, pp. 107–127, 2016.

[53] E. Bucchignani, M. Montesarchio, L. Cattaneo, M. P. Manzi, and P. Mercogliano, "Regional climate modeling over China with COSMO-CLM: performance assessment and climate projections," *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 21, p. 151, 2014.

[54] Korea Meteorological Administration, 2016.

[55] P. Sonneveld, "CGS, a fast lanczos-type solver for nonsymmetric linear systems," *SIAM Journal on Scientific and Statistical Computing*, vol. 10, no. 1, pp. 36–52, 1989.

[56] D. W. Jang and G. W. Choi, "Estimation of non-revenue water ratio using MRA and ANN in water distribution networks," *Water*, vol. 10, pp. 1–13, 2017.

[57] J. M. Gwak, *Research and Statistical Analysis*, Informa, London, UK, 2013.

[58] E. H. Park, E. S. Im, W. T. Kwon, and E. J. Lee, "Uncertainties estimation of AOGCM-based climate scenarios for impact assessment on water resources," in *Proceedings of the Korea Water Resources Association Conference*, pp. 138–142, Daejeon, Korea, May, 2006.