Kim Taereem (Orcid ID: 0000-0001-8689-1112)
Shin Ju-Young (Orcid ID: 0000-0002-1520-3965)
Heo Jun-Haeng (Orcid ID: 0000-0003-4727-3893)

Ensemble-based Neural Network Modeling for Hydrologic Forecasts: Addressing Uncertainty in the Model Structure and Input Variable Selection

Taereem Kim[1], Ju-Young Shin[2], Hanbeen Kim[3], and Jun-Haeng Heo[3]

[1] School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, USA.

[2] National Institute of Meteorological Science, Seogwipo, Jeju, South Korea.

[3] School of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea.

Corresponding author: Jun-Haeng Heo (jhheo@yonsei.ac.kr)

Key Points:

- We develop a new artificial neural network (ANN) ensemble model (ANN-ENS) to deal with uncertainty in model structure and input variable selection
- This is based on a modified backward elimination method that improves preliminary input variables for the ANN-ENS
- The resulting ANN-ENS demonstrated remarkable performance in terms of forecasting accuracy, stability, and ensemble reliability

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019WR026262

Abstract

Artificial neural networks (ANNs) have been extensively used to forecast monthly precipitation for water resources management over the past few decades. Efforts to produce more accurate and stable forecasts face ongoing challenges as the so-called single ANN (S-ANN) approach has several limitations, particularly regarding uncertainty. Many attempts have been made to deal with different types of uncertainties by applying ensemble approaches. Here, we propose a new ANN ensemble model (ANN-ENS) dealing with uncertainty in model structure and input variable selection to provide a more accurate and stable forecasting performance. This model is structured by generating various input layers, considering all the candidate input variables (i.e., large-scale climate indices and lagged precipitation). We developed a modified backward elimination method to select the preliminary input variables from all the candidate input variables. Then, we tested and validated the proposed ANN-ENS using observed monthly precipitation from 10 meteorological stations in the Han River basin, South Korea. Our results demonstrated that the ANN-ENS enhanced the forecasting performance in terms of both accuracy and stability. Although a significant uncertainty was introduced by using all the candidate input variables, the forecasting result outperformed S-ANNs for all employed stations. Additionally, the ANN-ENS provided a more stable forecasting performance in comparison with S-ANNs, which are highly sensitive. Moreover, the generated ensemble members were slightly biased at some stations, but were generally reliable.

1 Introduction

Precipitation forecasting is a fundamental challenge in agricultural, industrial, environmental, and water resources management and planning. Monthly precipitation forecasting is particularly important for hydrology and climatology because it captures seasonal variations (such as monsoons). Many attempts at monthly precipitation forecasting have focused on applying suitable techniques and models to provide accurate forecasts. Efficient management requires such forecasts to be both accurate and stable given the increasing complexity of weather conditions (Parasuraman and Elshorbagy, 2008; Tiwari and Adamowski, 2013; Tiwari and Chatterjee, 2010).

Artificial neural networks (ANNs), the most popular type of data-driven model, have been employed in many fields since the 1990s. ANNs have been widely used in hydrology-related fields, where they have shown outstanding performance in rainfall-runoff models, streamflow and precipitation forecasts, groundwater models, water-quality models, hydrologic time-series predictions, reservoir operations, and water management policy development (ASCE, 2000; Campolo et al., 1999; French et al., 1992; Hsu et al., 2010; Hung et al., 2009; Koizumi, 1999; Maier and Dandy, 1996; Rogers and Dowla, 2010). One advantage of ANNs is their superior predictability, based on their computational efficiency without requiring of explicit interpretation of physical processes occurring between input and output data. This enables ANNs to simulate complex geophysical processes considering both the linear and nonlinear relationships between input and output data. Despite these advantages, the ANNs still suffer from uncertainties in model structures, because they are extremely sensitive to model parameters and settings. Wang et al. (2006) argued that selecting a single best ANN model can no longer provide the best forecasting result owing to its sensitivity; therefore, many researchers have proposed the use of ANN ensemble models (ANN-ENSs) rather than single-result ANN models (S-ANNs).

The most widely used ensemble approach generates a number of ensemble members depending on the model structures (Adhikari and Agrawal, 2012; Elshorbagy et al., 2010;

Kasiviswanathanathan et al., 2013; Parasuraman et al., 2007; Parasuraman and Elshorbagy, 2008). Some examples of ANN-ENSs have already been presented along with evaluations of their effectiveness in hydrological modeling. Sharkey (1999) first suggested using arcing, boosting, and interior structural alterations to generate a set of S-ANNs by varying the initial weight parameters and the number of nodes in the hidden layer with an unchanged training set, further suggesting that a set of S-ANNs could be generated by altering the training algorithm. Cannon and Whitfield (2002) applied the bootstrap aggregation ensemble method to an ANN-ENS by considering the number of training iterations and the number of hidden nodes, with large-scale atmospheric conditions as model inputs, recommending an ANN-ENS for downscaling streamflow applications. Anctil and Nicolas (2004) compared five generalization approaches for reducing the bias and variance errors in a daily streamflow model, and found that a generalization approach performed better than an S-ANN in estimating the mean and median, and recommended a combination of ensemble members to improve and stabilize model performance. Shu and Burn (2004) examined six approaches for generating ensembles and applied an ANN-ENS to pooled flood frequency analysis for estimating the index flood and 10-year flood quantiles, concluding that the ANN-ENS was less sensitive, and provided better flood estimates than the S-ANN. Boucher et al. (2009) generated a set of ensemble members by perturbing the initial conditions with different weights and biases to assess the uncertainty inherent in the ANN-ENS process, demonstrating that a randomly initialized ensemble method outperformed an S-ANN in streamflow forecasting. However, they concluded that their ensemble generation method could not cover all uncertainties, and a more comprehensive methodology was needed to solve the uncertainty problem. Zaier et al. (2010) generated ensembles for an ANN-ENS using randomization, bagging, and boosting, and combined them using averaging and stacking techniques, concluding that their approach exhibited better performance than an S-ANN, and that a combination of ensemble members was effective in estimating ice thickness in 17 lakes in Canada. Araghinejad et al. (2011) proposed a procedure for forecasting seasonal streamflow and peak discharge using an ANN-ENS with different training algorithms having diverse performance functions. Tiwari et al. (2013) suggested a new hybrid wavelet bootstrap neural network model using an ANN-ENS provided 100 forecasts from a set of 100 weight parameters through the bootstrap resamples of a wavelet subtime series, concluding that the use of bootstrapping ensured model robustness and reliability. Kim and Seo (2015) applied an ANN-ENS to predict water quality variables while considering the influence of initial weight parameters in the training dataset, concluding that 100 ensemble members were adequate to reduce the dependence on the initial weight parameters in comparison with an S-ANN. Lee and Kang (2016) also applied an ANN-ENS to simulate daily discharge using lagged rainfall and discharge as input variables, concluding that the ensemble's performance and uncertainty depended on the length of the training data and the nodes of the hidden layer, while the activation functions may not have been important.

Although the importance of input variables for ANNs has been demonstrated (Bowden et al., 2005a,b; Hung et al., 2009), no explicit procedure has been defined for selecting input variables. Additionally, little interest has been shown in the uncertainty arising from input variable selection in comparison with other model structures, despite the fact that input variables are of primary importance (Pachepsky and Rawls, 2004). Researchers have often used input variables that are known to be influential and fixed them to the input layer without much consideration of the uncertainty arising from input variable selection. To deal with the importance of input variables and their selection, herein, we propose a new ANN-ENS that generates a set of input layers that account for such uncertainty in input variable selection. Our approach also uses a number of candidate input variables (i.e., large-scale climate indices and lagged precipitation) to improve the forecasting performance of monthly

precipitation. The large-scale climate indices represent climate variability as a simple diagnostic quantity that can be used to characterize an aspect of the geophysical system (NCAR, 2015). Many studies have attempted to identify large-scale climate signals (i.e., atmosphere-ocean teleconnection patterns) using a correlation between climate indices and precipitation, as climate indices are considered to be potential indicators of the global climate system under complex weather conditions (Fallah and Sodoudi, 2015; Giannini et al., 2003; Hartmann, 2016; Hastenrath and Greischar, 1993; Kumar et al., 2006). Kim et al. (2018) provided evidence of the influence of climate indices on monthly precipitation, identifying three key climate indices (the four- and ten-month lagged NINO1+2, along with the Atlantic Multidecadal Oscillation (AMO)) for monthly precipitation forecasting in South Korea, suggesting that these should be preferentially considered as predictive indicators. Here, we used general climate indices (all those excluding the above-mentioned key climate indices) to take advantage of the ANN-ENS approach's flexible consideration of both linear and nonlinear relationships between input and output data.

Thus, this study aimed to develop a new ANN-ENS capable of providing accurate and stable monthly precipitation forecasts by generating a number of input layers to address uncertainties related to input variable selection. The candidate input variables included large-scale climate indices and precipitation to take into account atmospheric and oceanic impacts. Additionally, we developed a modified backward elimination method for selecting preliminary input variables when generating ensemble members.

The rest of this paper is organized as follows. In section 2, we present the development methodology for the proposed ANN-ENS; section 3 provides details regarding the data and processing steps, section 4 describes the application of this ANN-ENS to 10 target stations in the Han River basin, South Korea, section 5 provides further discussion, and section 6 presents our conclusions.

2 Methodology

2.1 Basic ANN Structure

ANNs are complex nonlinear systems that mimic the working principles of the human brain, and are based on the relationship between neurons in different layers (Araghinejad, 2014). The basic S-ANN structure comprises three layers (input, hidden, and output) with input variables ($x_1, x_2, x_3, \dots, x_n$) and output ($y$; Figure 1).

An S-ANN uses a training algorithm to determine the best model structure for providing optimal output. A backpropagation algorithm is most commonly used to effectively train a neural network, and it is necessary to set the model parameters and settings appropriately for each layer. For the input layer, appropriate input variables should be assigned because they can guarantee good model performance, whereas inappropriate input variables can significantly degrade the model performance. For the hidden layer, the optimal number of nodes should be determined through trial-and-error because there is no definitive method available (Şahin et al., 2012; Valipour et al., 2013). However, uncertainty is bound to occur in this process, which is sensitive to model parameters and settings. For example, S-ANN are extremely sensitive to input variables, which directly affect the forecasting accuracy and produce different outputs even when the model settings are unchanged because of the randomly generated initial weight parameters (Kolen and Pollack, 1991; Santos-García, 2004). Our proposed ANN-ENS addresses this as follows.

2.2 Input Variable Selection

We developed a modified backward elimination method to select the preliminary input variables prior to generating ensemble members. Starting with all the candidate variables (a full model), the least significant variables are eliminated one by one until the model contains only a single variable. At each step, the remaining candidate variable having the smallest contribution to the model is eliminated. This approach is commonly used for linear regression models, in which p-values are used as a selection criterion. However, these p-value are too inefficient to be used as a selection criterion for ANNs because they can include input variables having both linear and nonlinear relationships with the outputs. To overcome this problem, we modified the backward elimination method to select preliminary input variables; the mean of 100 correlation coefficients (hereafter referred to as the mean correlation coefficient) is obtained from 100 S-ANNs after omitting a single variable one by one. After calculating the mean correlation coefficients for all variables, the least significant variable is eliminated according to the selection criterion (the mean correlation coefficient in the test period). This procedure can be easily conducted using an iterative approach, resulting in the selection of appropriate input variables considering both linear and nonlinear relationships between the input and output variables. The modified backward elimination procedure is outlined below and in Figure 2.

1) Set up a full model with all candidate variables, $x_1, x_2, \dots, x_c$. (c is the number of candidate variables).
2) Construct 100 S-ANNs with a set of input variables, each with a single variable omitted.
3) Calculate the mean correlation coefficients in the test period obtained by running 100 S-ANNs ($r_i = mean\ (r_i^1, r_i^2, \dots, r_i^{100})$, $i = 1, 2, \dots, c$), and repeat this step for all variables (x_i, $i = 1, 2, \dots, c$).
4) Compare the mean correlation coefficients for the omitted single variables (r_i, $i = 1, 2, \dots, c$).
5) Eliminate the least significant variable (e.g., if r_2 has the highest value, eliminate x_2)
6) Set the remaining variables as the new candidates, and then repeat steps 2–5 until only one variable remains.

2.3 Consideration of Uncertainty in Input Variable Selection

Our proposed ANN-ENS deals with uncertainty in the input variable selection by generating a number of input layers (Figure 3). Once the preliminary input variables are selected (section 2.2), different input layers are generated by altering (one by one) the variables that are selected (n) and unselected (m), producing a total of ($n \times m$) input layers. Based on this, the total outputs ($n \times m$) are then produced and defined as the ensemble members. Consequently, it is possible to generate ensemble members that integrate the information provided by all candidate variables. The median of the ensemble members is the result of the ANN-ENS.

The optimal number of nodes (K) in the hidden layer is determined by generating 100 S-ANNs for each node to consider the uncertainty in the randomly generated initial weight parameters. The preliminary input variables selected in section 2.2 are assigned to the input layer, and 100 S-ANNs are generated for each node (from a minimum of 1 to a maximum of N) in the hidden layer. After producing 100 outputs in the output layer for each node in the hidden layer, the median of the 100 outputs for each node is considered to be the forecasting

result. The optimal number of nodes in the hidden layer can be determined by comparing the *N* forecasting results produced for each node.

2.4 Model Evaluation Statistics

We applied three statistical criteria to estimate ANN-ENS performance: correlation coefficient (*r*), root mean square error (RMSE), and Nash-Sutcliffe model efficiency coefficient (NSE). The *r*-value ranges from -1 to 1, indicating either a negative or positive linear relationship between the observed and forecasted data series. The RMSE is used to estimate the magnitude of error between observed and forecasted data series. The NSE ranges from -∞ to 1, and is used to describe the hydrological model accuracy; a range from 0.5 to 0.65 has been suggested as sufficient for quality results (Moriasi et al., 2007). These criteria are defined as follows.

$$r = \frac{\sum_{t=1}^{T}(x(t)-\bar{x}(t))(y(t)-\bar{y}(t))}{\sqrt{\sum_{t=1}^{T}(x(t)-\bar{x}(t))^2}\sqrt{\sum_{t=1}^{T}(y(t)-\bar{y}(t))^2}} \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y(t)-x(t))^2} \quad (2)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T}(y(t)-x(t))^2}{\sum_{t=1}^{T}(x(t)-\bar{x}(t))^2} \quad (3)$$

where T is the data length, $x(t)$ and $y(t)$ are the observed and forecasted data series, and $\bar{x}(t)$ and $\bar{y}(t)$ are the means of the observed and forecasted data series, respectively.

2.5 Ensemble Evaluation Approaches

We employed three evaluation approaches to examine the reliability of the ensemble members: continuous ranked probability score (CRPS) (Hersbach, 2000; Matheson and Winkler, 1976), rank histogram (RH) (Talagrand et al., 1999), and confidence interval reliability diagram (RD) (Wilks 1995). The CRPS is a numerical criterion for assessing the quality of ensemble-based forecasting (Boucher et al., 2009), defined as:

$$CRPS(F_t, x_t) = \int_{-\infty}^{\infty}(F_t(y) - H\{y \geq x_t\})^2\, dy \quad (4)$$

where y is the forecasted variable, F_t is the cumulative distribution function of the forecasts at time *t*, and x_t is the corresponding observations. $H\{y \geq x_t\}$ indicates the Heaviside function that equals 0 for $y < x_t$ and 1 for $y \geq x_t$. The mean CRPS ($\overline{CRPS}$) can be calculated by averaging CRPS over the data length (T) as:

$$\overline{CRPS} = \frac{1}{T}\sum_{t=1}^{T} CRPS(F_t, x_t) \quad (5)$$

$\overline{CRPS}$ can be interpreted as a general form of the mean absolute square error (MAE) because it is reduced to the MAE for a point forecast (Gneiting and Raftery, 2007).

The RH and RD are widely used as graphical diagnostic tools to evaluate the spread of the ensemble members (Boucher et al., 2009; Boucher et al., 2010; Hwang and Carbone, 2009; Velázquez et al., 2009; Xu et al., 2019). The RH is obtained by calculating the frequency of plotting position (pp) from a set of ensemble members and observed data ($x(t)$). The plotting position of the observed data is calculated as:

$$pp(t) = \frac{\mathrm{rank}[x(t)]}{E+2}, t = 1, \ldots, T \quad (6)$$

where *E* is the total number of ensemble members. A flat RH is an ideally reliable case in which the ensemble members are uniformly distributed and all ranks have an equal relative

frequency. A U-shaped RH represents a distribution in which the ensemble members are under-dispersed and many observations fall outside the ensemble. A dome-shaped RH represents a distribution in which the ensemble members are over-dispersed and most observations fall near the center of the ensemble (Boucher et al., 2009; Boucher et al., 2010).

The RD is a plot of the observed relative frequency against the forecast probability represented by the confidence level (Velázquez et al., 2009; Xu et al., 2019). Confidence intervals for ensemble forecasts are calculated with confidence levels from 10% to 90% with an increment of 10%. For each confidence interval, the total number of observations included in the confidence interval is counted over the forecasting data length (T). Division by T produces the observed relative frequency. Therefore, the RD shows how well the forecast probabilities for the predicting period match with their observed relative frequencies. The closer the plotted points to the 1:1 diagonal, the more reliable the ensemble forecasts; plotted points above or below this line indicate under- or over-forecasting, respectively.

3 Data and Processing

3.1 Study Region and Data Sources

South Korean's Han River basin supplies water to Seoul, and other cities, and contains 13 multi-purpose and hydropower dams and weirs; therefore, accurate and stable monthly precipitation forecasting is important for the effective management and planning within the basin. We used monthly observed precipitation data collected by the Korean Meteorological Administration over a period of more than 40 years from 10 meteorological stations (Figure 4, Table 1).

3.2 Model Inputs

We used three categories of input variables (key climate indices, general climate indices, and the lagged precipitation) to consider the broad set of atmospheric and oceanic impacts having possible linear and nonlinear relationships with monthly precipitation. ANNs can include a number of potential input variables that behave synergistically without imposing any fixed relationships between the input and output data.

The key climate indices and lagged precipitation are used to consider the linear relationship with monthly precipitation. The key climate indices (4-month lagged NINO1+2, the 10-month lagged NINO1+2, and non-lagged AMO) were previously identified by Kim et al. (2018) and shown to have a statistically significant relationship with monthly precipitation in South Korea. To focus on real-time ANN-ENS forecasts, the time lag of the AMO index was set to 1- month (instead of no lag) because it was not possible to provide this index for the same month. A partial autocorrelation function (PACF) was used to determine the time lag in considering the autoregressive characteristics of monthly precipitation. The PACF plots of monthly precipitation demonstrated prominent spikes at 1, 12, 24, and 36 months for most stations (Figure 5). Therefore, the 1- month, 12- month, 24- month, and 36- month lagged precipitation (LAG1, LAG2, LAG3, and LAG4, respectively) were included in the candidate input variables at all stations.

The general climate indices were used to consider the nonlinear relationship with monthly precipitation. These are potentially useful predictors for monthly precipitation forecasting because they have already been shown to have a concurrent relationship (Schepen et al., 2012). Furthermore, there is a possibility of sporadic and complex interactions among the general climate indices when applying an ANN-ENS. Considering the nonlinear

relationships existing among the variables, the time lag of these general climate indices was set to 1- month to reduce complexity in the real-time forecasts. The large-scale climate indices used in this study (Table 2) were obtained from the National Oceanic & Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) and screened for data availability for real-time forecasts in South Korea. All 27 variables, including key climate indices, general climate indices, and lagged precipitation were set as candidate input variables at all stations (Table 3).

3.3 Model Setting

Our proposed ANN-ENS was built using MATLAB (Mathworks, 2017b) software. A multi-layer perceptron using the logistic sigmoid function was adopted as the superior option (Valipour et al., 2013). Resilient backpropagation, which is well known for fast training and robustness, was used for the training algorithm. The detailed conditions of the model were determined by the basic MATLAB parameters. The maximum number of epochs was set to 1,000 and the mean square error was used as the error function. The maximum number of nodes in the hidden layer was set to 30 since the number of input variables was less than 30.

In general, S-ANNs use a training period to train the neural network by fitting the weight parameters to minimize error, a validation period to estimate performance by tuning the structure, and a test period to assess the overall performance (Kumar et al., 2002). However, tuning the network structures during the validation period is not necessary for ANN-ENSs because the ensemble approach can itself respond to uncertainty in the network structures. Therefore, we split the monthly precipitation data into two sets: a training from the start of the data record to fDecember 2014, and a test period of the most recent three years (January 2015 to December 2017) to validate the availability of real-time forecasts from the ANN-ENS for extreme weather conditions, such as severe droughts, that occurred during the test period.

In the input layer, the ensemble members were generated based on the number of preliminarily selected input variables (m) and unselected input variables ($n = 27 - m$) at each station. Therefore, different ensemble members were generated at each station according to the input variable selection results. The optimal number of nodes (K) in the hidden layer was determined by generating 100 ensemble members for each node from the minimum to the maximum (1 to 30) with the preliminarily selected input variables, considering randomly generated initial weight parameters. The medians of the 100 ensemble members were produced for each node in the hidden layer, and the optimal number of nodes (K) was determined by comparing the forecasting performance of these medians based on the r values in the test period. Table 4 presents the optimal number of nodes (K) in the hidden layer based on the highest r value in the test period at all employed stations. Finally, a total of ($n \times m$) ensemble members were generated in the output layer.

4 Results

4.1 Preliminary Selected Input Variables

We used the modified backward elimination method (section 2.2) to select input variables from the candidate variables; the key climate indices were fixed and thus could not be eliminated during this process.

The *r* values of the training and test periods gradually changed as variables were eliminated (Figure 6); the *r* value in the training period gradually decreased, whereas the *r* value in the test period gradually increased and then decreased. These results demonstrate that the input variables were properly selected by the modified backward elimination method. In the training period, the highest *r* value was obtained with all the input variables because a large number of input variables is considerably powerful for ANN training, such that performance decreases as variables are eliminated. However, there was an inflection point in the test period, indicating a change in performance according to the input variables. The *r* value initially improved as variables were eliminated; this can be seen as an effect of reducing the overfitting problem. However, after the inflection point, the performance slightly decreased as the information available for forecasting decreased. Therefore, the variables remaining after the inflection point were the preliminary selected input variables (Table 5).

Among these, the 12 month lagged precipitation (LAG2) was included at all stations except no. 202 because this reflects the seasonality of monthly precipitation in South Korea. In contrast, different sets of general climate indices were selected at each station.

4.2 Forecasting Accuracy

The ANN-ENS forecasting results in the training and test periods demonstrated that different ensemble members were generated for each station, according to the input variable selection (Figure 7). For example, at station no. 108, a total of 180 (15×12) ensemble members were generated because $m = 15$ and $n = 12$. The ANN-ENS consistently exhibited a narrow ensemble range in the training period, with the median closely matching the observations, demonstrating that all candidate input variables are potential predictors and can provide sufficient information for monthly precipitation forecasting. In contrast, the ANN-ENS showed relatively broader ensemble ranges in the test period, with the median generally consistent with the pattern of observations at all stations. This suggests that the ANN-ENS can provide stable forecasting performance by considering the median of the ensemble members even though the input variable selection of the ANN model structure can introduce significant associated uncertainty.

The median in the test period exhibited good agreement with the observed precipitation. Specifically, extreme weather conditions, such as the severe drought in 2015 and the restored climate conditions in the following years, were well-captured by the ANN-ENS (Figure 7). To compare these results in detail, the statistical measurements examine the forecasting accuracy for the medians of the ensemble members. The forecasting accuracy measured by the averaged value of *r*, RMSE, and NSE was generally good, with a strong positive linear relationship and predictive power (Table 6). Station no. 211 exhibited the most accurate forecasts in the test period ($r = 0.9378$, RMSE = 44.85, and NSE = 0.8694). These results suggest that our proposed ANN-ENS is capable of providing accurate forecasting performance, although the input variables are the greatest contributors to uncertainty.

4.3 Forecasting Stability

To examine the stability of the ANN-ENS, we compared its forecasting performance (*r* and RMSE) to that of 100 S-ANNs built using the preliminarily selected input variables and the optimal number of nodes (K) in the hidden layer (sections 2.2 and 2.3). Box plots of the 100 *r* (left) and RMSE (right) values for the 100 S-ANNs and 100 ANN-ENSs at all

stations (Figure 8) showed that the ANN-ENSs generally produced narrower bands and more accurate performances than the S-ANNs in both the training and test periods. In the test periods, the *r* and RMSE values of the S-ANNs were much more dispersed than those of the ANN-ENSs, demonstrating lower sensitivity and superior forecasting ability of the latter along with the greater uncertainty of the former.

4.4 Ensemble Reliability

For all stations, the CRPS values were lower than the mean MAE values in both the training and test periods (Table 7), indicating that the ANN-ENS exhibited a more reliable performance than the S-ANNs.

Although the graphical methods are not sufficient for this purpose, they have the advantage of simply identifying the tendency of the generated ensemble members. The RH of the ensemble members for the ANN-ENS in the training and test periods (Figure 9), indicated that each station had a different number of ensemble members depending on the input variable selection results. The RHs in the training periods showed different shapes by station, such as domed-shape (no. 101), slightly skewed and domed-shape in the lower quantile (nos. 108, 127, 211, and 221), and slightly skewed U-shape (nos. 100, 114, 202, 203, and 212). In contrast, the RHs in the test periods generally showed a flat shape (nos. 127, 211, and 221) or a flattened U-shape with one spike at the highest relative frequency in the first rank of the RH (No. 100, 101, 108, 114, 202, 203, 212). We considered a flat RH as a good indicator for reliable ensemble members, whereas one spike at the highest relative frequency in the first rank meant the observation was usually lower than all ensemble members. Consequently, the ensemble members generally appeared to have a small bias in the forecasts at all stations.

The RD of the ensemble members for the ANN-ENS (Figure 10) in the training period was generally well-matched with the 1:1 diagonal line (nos. 108, 127, 211, and 221), whereas some were slightly below (nos. 100, 114, 202, 203, and 212) or above (no. 101). Additionally, the RD in the test period confirmed the ideally dispersed ensemble members with plotted points similar to the 1:1 line (nos. 202, 203, 212, and 221) or slightly above (nos. 101, 108, 127, and 211) or below (nos. 100 and 114). This agrees with the RH results, indicating that the ensemble members generated for the ANN-ENS provided reliable ensemble members with small biases, although they did not cover all possible uncertainties.

5 Discussion

Input variable selection is crucial to the development of data-driven models, particularly for water resources modeling (Maier and Dandy, 2000, Galelli et al., 2014, Quilty et al., 2016, Sahoo et al., 2017); however, uncertainty remains as there is no explicit procedure for accomplishing this. Therefore, we developed a new ANN-ENS to address this problem by selecting preliminary input variables using modified backward elimination and generating a number of input layers by altering the selected and unselected input variables one by one. This approach enables the consideration of all possible information that candidate input variables can provide and avoid problems (e.g. overfitting) that may occur when all the candidate input variables are used at once.

Our candidate input variables included large-scale climate indices and lagged precipitation, based on the Abbot and Marohasy (2014), who reported on the usefulness of such climate indices in precipitation forecasts. The use of multiple candidate input variables is an attempt to flexibly consider both linear and nonlinear interactions between monthly precipitation and large-scale climate indices based on the advantage of ANNs. This was

based on two principles: (1) ANNs can handle a number of potential input variables that can behave synergistically without imposing any fixed relationships between input and output data, and (2) the circulation of the atmosphere and ocean is a great communicator for precipitation.

These principles represent a key point of this study that has not been generally applied in previous hydrologic forecasting models. Thus far, many researchers have used known influential input variables because insignificant input variables do not guarantee good model performance, particularly in physically- based hydrologic forecasting models. However, it is becoming increasingly difficult to determine the dominant factors influencing climate conditions for a certain region; thus, using multiple input variables can be an alternative to the difficulty in selecting certain input variables. Additionally, generating an ensemble with large-scale climate indices would help us to make decisions under complex weather situations. Our results demonstrate that the proposed ANN-ENS provides satisfactory forecasting performance and prove the viability of a number of large-scale climate indices as candidate input variables. This leads to enhanced flexibility when using general climate indices and enables accurate and stable monthly precipitation forecasting by considering all the candidate input variables.

Our method produces two consequences with regard to the forecasting accuracy of the ANN-ENS. First, the importance of input variable selection is highlighted by the spread of ensemble members (Figure 7), which represents the uncertainty obtained by a set of input layers generated by altering the preliminarily selected and unselected input variables. These have a high sensitivity, indicating that each input variable has disparate climatic information and there is a possibility that forecasting accuracy can be highly altered by only one input variable. Second, the forecasting accuracy exhibits outstanding performance despite the broad ensemble range caused by the high sensitivity of the input variables (Table 6). It is remarkable that the proposed ANN-ENS can provide a satisfactory forecasting accuracy despite the significant uncertainty being introduced by the input variable selection.

We also examined the forecasting performance in terms of stability, as this is directly related to model availability. If forecasting accuracy is the only indicator that requires consideration for forecasting performance, it may be obtained from one of the 100 S-ANNs. However, the S-ANNs exhibited higher sensitivity than the ANN-ENS (Figure 8), whereas the ANN-ENS exhibited a more stable forecasting performance with reduced sensitivity. These results agree with the findings of Jeong and Kim (2005, 2009), Zaier et al. (2010), and Zou and Yang (2004), all of whom suggested the use of ensemble techniques in hydrologic forecasting because ANN-ENSs are less sensitive and provide better performance than S-ANNs. Furthermore, Kim and Seo (2015) reported that the performance of S-ANNs is significantly affected by the selection of data sets and that ANN-ENSs can reduce modeling errors.

It is necessary to verify the reliability of the ensemble itself, and not just the forecasting accuracy and stability to quantitatively understand the uncertainty represented by the ensemble members. In the present study, we applied three tools (CRPS, RH, and RD) to evaluate the ensemble reliability. The ANN-ENS's superior performance was shown by the much lower CRPS value (in comparison with the mean MAE) at all stations (Table 7). Additionally, the RHs differed by stations between domed-, slightly skewed and domed-, and slightly skewed U- shapes indicating the possibility of bias in the generated ensemble members (Figure 9). Most RHs in the training period appeared to have small biases and there were some high frequencies in the higher quantile at approximately half of the employed stations (nos. 100, 108, 202, 203,.212, and 221). However, there were no extremely skewed shapes, implying that the generated ensemble members were within a reliable range. Consequently, the RHs in the training periods confirmed that the ensemble members were

properly generated without extreme bias, although there was a small bias. However, the RHs in the test periods indicate that underestimation frequently occurred. The RDs (Figure 10) showed that the plotted points generally lined up along the 1:1 line with some slightly above or below, indicating that the ensemble members were generally well-dispersed. Together, these three measures demonstrated that the ensemble members were properly generated from multiple input layers with minor small bias. However, further research is required to investigate the dispersion of the generated ensemble members with regard to this bias.

6 Conclusions

We developed an ANN-ENS capable of dealing with uncertainty caused by input variable selection. Using candidate input variables consisting of large-scale climate indices and lagged precipitation, we developed a modified backward elimination method to select preliminary input variables prior to generating a number of input layers. We then tested this method using data from 10 meteorological stations in the Han River basin, South Korea and examined the forecasting performances with respect to accuracy, stability, and ensemble reliability with the following findings:

(1) Our proposed ANN-ENS proved the capable of using multiple candidate input variables by taking advantage of potential information from large-scale climate indices. Although input variables can introduce a significant uncertainty into modeling, our model demonstrated satisfactory performance in monthly precipitation forecasting at all stations, with an average r and NSE of 0.8727 and 0.6433, respectively, during the test period. Satisfactory performance occurred even during severe drought conditions.

(2) This method also exhibits improved performance in terms of forecasting stability; comparison of 100 S-ANNs and ANN-ENSs showed that the latter's r and RMSE values had much narrower box-plot ranges than the former's, particularly during the test period. This stability demonstrates the suitability and reliability of using the proposed ANN-ENS to forecast monthly precipitation and can benefit water resources management.

(3) The proposed ANN-ENS generated ensemble members that were well-dispersed but had a small bias at most stations. Although the generated members could not cover all sources of uncertainty, we considered their performance acceptable as they showed no sign of extreme bias. Although the generated ensemble members from the ANN-ENS were not ideally dispersed, it was remarkable that this method provided satisfactory forecasting performance in terms of both accuracy and stability. However, there are still limitations regarding the ability to enhance the spread of these ensembles, as mentioned Boucher et al. (2009).

In conclusion, our proposed ANN-ENS's ability to combine climate information, ANN analysis, and ensemble techniques has a significantly strong potential for practical applications in monthly precipitation forecasting for long-term water resources planning and management, such as dam and reservoir operations for hydropower and flood control, water supply, and integrated hydrological management.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2019R1A2C2010854). Monthly precipitation data used in this study are available in the supporting information and can be obtained from the Korea Meteorological Administration (http://www.weather.go.kr/weather/climate/past_table.jsp). Large-scale climate indices used

in this study are obtained from the NOAA/ESRL (https://www.esrl.noaa.gov/psd/data/climateindices/list/).

References

Adhikari, R., and Agrawal, R. K. (2012). Combining Multiple Time Series Models through a Robust Weighted Mechanism. Proceedings of the 1st International Conference on Recent Advances in Information Technology, 455–60.

Aksoy, H., and Dahamsheh, A. (2009). Artificial Neural Network Models for Forecasting Monthly Precipitation in Jordan. Stochastic Environmental Research and Risk Assessment, 23 (7): 917–31.

Anctil, F., and Nicolas, L. (2004). Generalisation for Neural Networks through Data Sampling and Training Procedures, with Applications to Streamflow Predictions. Hydrology and Earth System Sciences, 8 (5): 940–58.

Araghinejad, S., Azmi, M., and Kholghi, M. (2011). Application of Artificial Neural Network Ensembles in Probabilistic Hydrological Forecasting. Journal of Hydrology, 407: 94–104.

Araghinejad, S. (2014). Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering. Water Science and Technology Library. Springer.

ASCE Task Committee (2000). Artificial Neural Networks in Hydrology: I.Preliminary Concepts. Journal of Hydrologic Engineering, 5 (1): 115–23.

Boucher, M. -A., Perreault, L., and Anctil, F. (2009). Tools for the Assessment of Hydrological Ensemble Forecasts Obtained by Neural Networks. Journal of Hydroinformatics, 11.3-4: 297–307.

Bowden, G. J., Dandy, G. C., and Maier, H. R. (2004a). Input Determination for Neural Network Models in Water Resources Applications. Part 1 – Background and Methodology. Journal of Hydrology, 301(1–4): 75–92.

Bowden, G. J., Dandy, G. C., and Maier, H. R. (2004b). Input Determination for Neural Network Models in Water Resources Applications. Part 2. Case study: Forecasting Salinity in a River. Journal of Hydrology, 301(1–4): 93–107.

Campolo, M., Soldati, A., and Andreussi, P. (1999). Forecasting River Flow Rate during Low-Pow Periods Using Neural Networks. Water Resources Research, 35 (11): 3547–52.

Cannon, A. J., and Whitfield, P. H. (2002). Downscaling Recent Streamflow Conditions in British Columbia, Canada Using Ensemble Neural Network Models. Journal of Hydrology, 259: 136–51.

Delleur, J. W., and Kavvas, M. L. (1978). Stochastic Models for Monthly Rainfall Forecasting and Synthetic Generation. Journal of Applied Meteorology, 17: 1528–36.

Elshorbagy, A., Corzo, G., Srinivasulu, S., and Solomatine, D. P. (2010). Experimental Investigation of the Predictive Capabilities of Data Driven Modeling Techniques in Hydrology – Part 2: Application. Hydrology and Earth System Sciences, 14: 1943–61.

Fallah, B., and Sodoudi, S. (2015). Bimodality and Regime Behavior in Atmosphere–ocean Interactions during the Recent Climate Change. Dynamics of Atmospheres and Oceans, 70: 1–11.

French, M. N., Krajewski, W. F., and Cuykendall, R. R. (1992). Rainfall Forecasting in Space and Time Using a Neural Network. Journal of Hydrology, 137 (1): 1–31.

Galelli, S., Humphrey, G. B., Maier, H. R., Castelletti, A., Dandy, G. C., Gibbs, M. S. (2014). An Evaluation Framework for Input Variable Selection Algorithms for Environmental Data-Driven Models. Environmental Modelling and Software, 62: 33-51.

Giannini, A, Saravanan, R., and Chang, P. 2003. Oceanic Forcing of Sahel Rainfall on Interannual to Interdecadal Time Scales. Science, 302 (5647): 1027–30.

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R., and Cane, M. A. (2001). Current Approaches to Seasonal-to-Interannual Climate Predictions. International Journal of Climatology, 21 (9): 1111–52.

Hartmann, D. (2016). Global Physical Climatology. 2nd Edition. Springer.

Hartmann, H., Snow, J. A., Su, B., and Jiang, T. (2016). Seasonal Predictions of Precipitation in the Aksu-Tarim River Basin for Improved Water Resources Management. Global and Planetary Change, 147: 86–96.

Hastenrath, S., and Greischar, L. (1993). Circulation Mechanisms Related to Northeast Brazil Rainfall Anomalies. Journal of Geophysical Research: Atmospheres, 98: 5093–5102.

Hawthorne, S., Wang, Q. J., Schepen, A., and Robertson, D. (2013). Effective Use of General Circulation Model Outputs for Forecasting Monthly Rainfalls to Long Lead Times. Water Resources Research, 49 (9): 5427–36.

Hsu, K. –L., Gupta, H. V., and Sorooshian, S. (2010). Artificial Neural Network Modeling of the Rainfall-Runoff Process. Water Resources Research, 31: 2517–30.

Hung, N. Q., Babel, M. S., Weesakul, S., and Tripathi, N. K. (2009). An Artificial Neural Network Model for Rainfall Forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences, 13 (8): 1413–25.

Hwang, Y., and Carbone, G. J. (2009). Ensemble Forecasts of Drought Indices Using a Conditional Residual Resampling Technique. Journal of Applied Meteorology and Climatology, 48 (7): 1289–1301.

Kasiviswanathan, K. S., Cibin, R., Sudheer, K. P., and Chaubey, I. (2013). Constructing Prediction Interval for Artificial Neural Network Rainfall Runoff Models Based on Ensemble Simulations. Journal of Hydrology, 499: 275–88.

Khalil, B., Ouarda, T. B. M. J., and St-Hilaire, A. (2011). Estimation of Water Quality Characteristics at Ungauged Sites Using Artificial Neural Networks and Canonical Correlation Analysis. Journal of Hydrology, 405: 277–87.

Kim, S. E., and Seo, I. W. (2015). Artificial Neural Network Ensemble Modeling with Conjunctive Data Clustering for Water Quality Prediction in Rivers. Journal of Hydro-Environment Research, 9 (3): 325–39.

Kim, T., Shin, J. –Y., Kim, S., and Heo, J. –H. (2018). Identification of Relationships between Climate Indices and Long-Term Precipitation in South Korea Using Ensemble Empirical Mode Decomposition. Journal of Hydrology, 557: 726–39.

Koizumi, K. (1999). An Objective Method to Modify Numerical Model Forecasts with Newly Given Weather Data Using an Artificial Neural Network. Weather and Forecasting, 14 (1): 109–18.

Kolen, J. F., and Pollack, J. B. (1991). Back Propagation Is Sensitive to Initial Conditions. Advances in Neural Information Processing Systems 3, 860–67.

Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G., and Cane, M. (2006). Unraveling the Mystery of Indian Monsoon Failure During El Niño. Science, 314 (5796): 115–19.

Kumar, M., Raghuwanshi, N., Singh, R., Wallender, W., and Pruitt, W. (2002). Estimating Evapotranspiration Using Artificial Neural Network. Journal of Irrigation and Drainage Engineering, 128 (4): 224–33.

Lang, X., and Wang, H. (2010). Improving Extraseasonal Summer Rainfall Prediction by Merging Information from GCMs and Observations. Weather and Forecasting, 25 (4): 1263–74.

Lee, D.-H., and Kang, D.-S. (2016). The Application of the Artificial Neural Network Ensemble Model for Simulating Streamflow. Procedia Engineering, 154: 1217–24.

Maier, H. R., and Dandy, G. C. (1996). The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. Water Resources Research, 32: 1013–22.

Maier, H. R., and Dandy, G. C. (2000). Neural Networks for the Precipitation and Forecasting of Water Resources Variables: A Review of Modelling Issues and Applications. Environmental Modelling and Software, 15: 101-124.

Mahsin, M., Akhter, Y., and Begum, M. (2012). Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series. Journal of Mathematical Modelling and Application, 1 (5): 67–73.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodness, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I. (2010). Precipitation Downscaling under Climate Change: Recent Developements to Bridge the Gap between Dynamical Models and the End User. Reviews of Geophysics, 48: 1–38.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. American Society of Agricultural and Biological Engineers, 50 (3): 885-900.

National Center from Atmospheric Research Staff (Eds). Last modified 28 May 2015. The Climate Data Guide: Overview: Climate Indices.

Parasuraman, K., Elshorbagy, A., and Si, B. C. (2007). Estimating Saturated Hydraulic Conductivity Using Genetic Programming, Soil Science Society of America Journal, 71(6): 1676–84.

Parasuraman, K., and Elshorbagy, A. (2008). Toward Improving the Reliability of Hydrologic Prediction: Model Structure Uncertainty and Its Quantification Using Ensemble-based Genetic Programming Framework. Water Resources Research, 44, W12406.

Papalaskaris, T., Panagiotidis, T., and Pantrakis, A. (2016). Stochastic Monthly Rainfall Time Series Analysis, Modeling and Forecasting in Kavala City, Greece, North-Eastern Mediterranean Basin. Procedia Engineering, 162: 254–63.

Quilty, J., Adamowski, J., Khalil, B., and Rathinasamy, M. (2016). Bootstrap Rank-ordered Conditional Mutual Information (broCMI): A Nonlinear Input Variable Selection Method for Water Resources Modeling. Water Resources Research, 52: 2299-2326.

Rogers, L. L., and Dowla, F. U. (2010). Optimization of Groundwater Remediation Using Artificial Neural Networks with Parallel Solute Transport Modeling. Water Resources Research, 30: 457–81.

Sachindra, D. A., and Perera, B. J. C. (2016). Statistical Downscaling of General Circulation Model Outputs to Precipitation Accounting for Non-Stationarities in Predictor-Predictand Relationships. PLoS ONE, 11 (12): e0168701.

Şahin, C. Ş., Gundry, S., and Uyar, M. Ü. (2012). Markov Chain Analysis of Self-Organizing Mobile Nodes. Journal of Intelligent & Robotic Systems, 67 (2): 133–53.

Sahoo, S., Russo, T. A., Elliott, J., and Foster, I. (2017). Machine Learning Algorithms for Modeling Groundwater Level Changes in Agricultural Regions of the U.S., Water Resources Research, 53: 3878-95.

Santos-García, G., Varela, G., Novoa, N., and Jiménez, M. F. (2004). Prediction of Postoperative Morbidity after Lung Resection Using an Artificial Neural Network Ensemble. Artificial Intelligence in Medicine, 30: 61–69.

Schepen, A., Wang, Q. J., and Robertson, D. (2012). Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rainfall. Journal of Climate, 25: 1230–46.

Schepen, A., and Wang, Q. J. (2014). Ensemble Forecasts of Monthly Catchment Rainfall out to Long Lead Times by Post-Processing Coupled General Circulation Model Output. Journal of Hydrology, 519: 2920–31.

Sharkey, A. J. C. (1999). Combining Artificial Neural Nets. Springer.

Shu, C., and Burn, D. H. (2004). Artificial Neural Network Ensembles and Their Application in Pooled Flood Frequency Analysis. Water Resources Research, 40: 1–10.

Soltani, S., Modarres, R., and Eslamian, S. S. (2007). The Use of Time Series Modeling for the Determination of Rainfall Climates of Iran. International Journal of Climatology, 27: 819–29.

Tiwari, M. K., and Chatterjee, C. (2010). Uncertainty Assessment and Ensemble Flood Forecasting Using Bootstrap based Artificial Neural Networks (BANNs). Journal of Hydrology, 382: 20–33.

Tiwari, M. K., and Adamowski, J. (2013). Urban Water Demand Forecasting and Uncertainty Assessment Using Ensemble Wavelet-Bootstrap-Neural Network Models. Water Resources Research, 49: 6486–507.

Valipour, M., Banihabib, M. E., and Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the Autoregressive Artificial Neural Network Models in Forecasting the Monthly Inflow of Dez Dam Reservoir. Journal of Hydrology, 476: 433–41.

Wang, S., Feng, J. and Liu, G. (2013). Application of Seasonal Time Series Model in the Precipitation Forecast. Mathematical and Computer Modelling, 58: 677–83.

Wang, W., Van Gelder, P., Vrijling, J.K., and Ma, J. (2006). Forecasting daily streamflow using hybrid ANN models. Journal of Hydrology, 324:383–99.

Xu, J., Ye, A., Duan, Q., Ma, F., and Zhou, Z. (2017). Improvement of Rank Histograms for Verifying the Reliability of Extreme Event Ensemble Forecasts. Environmental Modelling and Software, 92: 152–62.

Zaier, I., Shu, C., Ouarda, T. B. M. J., Seidou, O., and Chebana, F. (2010). Estimation of Ice Thickness on Lakes Using Artificial Neural Network Ensembles. Journal of Hydrology, 383: 330–40.

Zou, H., and Yang, Y. (2004). Combining Time Series Models for Forecasting. International Journal of Forecasting, 20: 69–84.

Table 1. Meteorological stations providing data for this study.

Station No.	Station name	Longitude (°)	Latitude (°)	Altitude (El.m)	Data period
100	Daegwallyeong	128.72	37.68	772.40	Jan 1972-Dec 2017
101	Chuncheon	127.74	37.90	76.80	Jan 1966-Dec 2017
108	Seoul	126.97	37.57	85.50	Jan 1960-Dec 2017
114	Wonju	127.95	37.34	150.70	Jan 1973-Dec 2017
127	Chungju	127.95	36.97	113.70	Jan 1973-Dec 2017
202	Yangpyeong	127.49	37.49	47.40	Jan 1973-Dec 2017
203	Icheon	127.48	37.26	90.00	Jan 1973-Dec 2017
211	Inje	128.17	38.06	198.70	Jan 1973-Dec 2017
212	Hongcheon	127.88	37.68	146.20	Jan 1973-Dec 2017
221	Jecheon	128.19	37.16	263.10	Jan 1973-Dec 2017

Table 2. Large-scale climate indices used in this study.

Group	Climate index	Abbreviation	Classification	Data period
Sea Surface Temperature	Atlantic Meridional Mode	AMM	SST:Atlantic	Jan 1948-Dec 2017
	Atlantic Multidecadal Oscillation	AMO	SST:Atlantic	Jan 1948-Dec 2017
	Tropical Northern Atlantic Index	TNA	SST:Atlantic	Jan 1948-Dec 2017
	Tropical Southern Atlantic Index	TSA	SST:Atlantic	Jan 1948-Dec 2017
	Oceanic Nino Index	ONI	SST:Pacific	Jan 1950-Dec 2017
	Trans-Nino Index	TNI	SST:Pacific	Jan 1948-Dec 2017
	Western Hemisphere Warm Pool	WHWP	SST:Pacific	Jan 1948-Dec 2017
	NINO 1+2	NINO12	ENSO/SST: Pacific	Jan 1950-Dec 2017
	NINO 3.4	NINO34	ENSO/SST: Pacific	Jan 1950-Dec 2017
	NINO 3	NINO3	ENSO/SST: Pacific	Jan 1950-Dec 2017
	NINO 4	NINO4	ENSO/SST: Pacific	Jan 1950-Dec 2017
	Pacific Decadal Oscillation	PDO	Teleconnections	Jan 1948-Dec 2017
Multiple Index	Bivariate ENSO Timeseries	BEST	ENSO	Jan 1948-Dec 2017
	Multivariate ENSO Index	MEI	ENSO	Jan 1950-Dec 2017
Atmosphere Circulation	Artic Oscillation	AO	Atmosphere	Jan 1950-Dec 2017
	Quasi-Biennial Oscillation	QBO	Atmosphere	Jan 1948-Dec 2017
	Southern Oscillation Index	SOI	Atmosphere	Jan 1951-Dec 2017
	Eastern Atlantic/Western Russia	EAWR	Teleconnections	Jan 1950-Dec 2017
	North Atlantic Oscillation	NAO	Teleconnections	Jan 1950-Dec 2017
	Northern Oscillation Index	NOI	Teleconnections	Jan 1948-Dec 2017
	Western Pacific Index	WP	Teleconnections	Jan 1950-Dec 2017
	Pacific North American Index	PNA	Teleconnections	Jan 1950-Dec 2017

Table 3. Candidate input variables for all stations.

Category	Candidate input variables (Time lag)
Lagged precipitation	LAG1 (1), LAG2 (12), LAG3 (24), LAG4 (36)
Key climate indices	NINO12 (4), NINO12 (10), AMO (1)
General climate indices	AMM (1), TNA (1), TSA (1), ONI (1), TNI (1), WHWP (1), NINO34 (1), NINO3 (1), NINO4 (1), PDO (1), BEST (1), MEI (1), AO (1), QBO (1), SOI (1), EAWR (1), NAO (1), NOI (1), WP (1), PNA (1)

Table 4. Optimal number of nodes (*K*) in the hidden layer at all stations.

Station No.	No. of node (*K*)	Station No.	No. of node (*K*)
100	21	202	12
101	30	203	8
108	10	211	10
114	6	212	30
127	25	221	13

Table 5. Input variables (including fixed and selected) determined through the modified backward elimination method at all stations.

Station No.	No. of input variables	Preliminary selected input variables	
		Fixed variables	Selected variables by modified backward elimination
100	5	NINO12(-4), NINO12(-10), AMO(-1)	AO, LAG2
101	11		LAG1, WHWP, NINO3, TSA, LAG2, AMM, LAG4, NAO
108	15		TSA, LAG2, TNA, AMM, WP, NAO, QBO, NINO3, NINO34, PNA, MEI, LAG3
114	15		ONI, LAG1, NINO34, NINO3, AMM, TSA, LAG4, QBO, WP, PNA, AO, LAG2
127	6		TSA, LAG2, NAO
202	4		NAO
203	10		NINO34, NINO3, MEI, WP, LAG3, LAG2, NAO
211	12		WHWP, BEST, NINO34, NINO3, LAG2, NOI, LAG4, WP, AO
212	4		LAG2
221	10		WP, ONI, NINO3, NINO34, WHWP, NAO, LAG2

Table 6. Statistical measurements (*r*, RMSE, and NSE) for the ANN-ENS performance in the training and test periods at all stations.

Station	*r*		RMSE		NSE	
No.	Training	Test	Training	Test	Training	Test
100	0.8369	0.8615	89.58	68.32	0.6722	0.5798
101	0.9778	0.9177	32.35	53.50	0.9511	0.8415
108	0.9447	0.9336	55.95	62.59	0.8799	0.7235
114	0.9423	0.8743	49.23	57.72	0.8748	0.6914
127	0.9268	0.8935	49.45	51.17	0.8365	0.7205
202	0.8101	0.7383	96.48	86.49	0.6265	0.4353
203	0.9276	0.8945	55.34	53.27	0.8449	0.6799
211	0.9401	0.9378	47.10	44.85	0.8677	0.8694
212	0.8632	0.8219	80.30	71.32	0.7159	0.6478
221	0.9315	0.8539	54.05	89.27	0.8555	0.2441
Average	0.9101	0.8727	60.98	63.85	0.8125	0.6433

Table 7. The CRPS and MAE in the training and test periods at all stations.

Station	CRPS		Mean MAE	
No.	Training	Test	Training	Test
100	41.92	44.78	66.20	75.38
101	15.10	30.57	30.67	64.00
108	25.89	37.44	49.04	73.54
114	23.56	34.49	42.38	62.16
127	22.39	30.97	41.27	60.38
202	42.52	48.83	62.98	90.17
203	26.09	32.25	44.45	58.61
211	21.26	30.69	40.04	63.40
212	34.71	37.28	55.71	56.07
221	25.16	48.02	44.94	95.87
Average	27.86	37.53	47.77	69.96

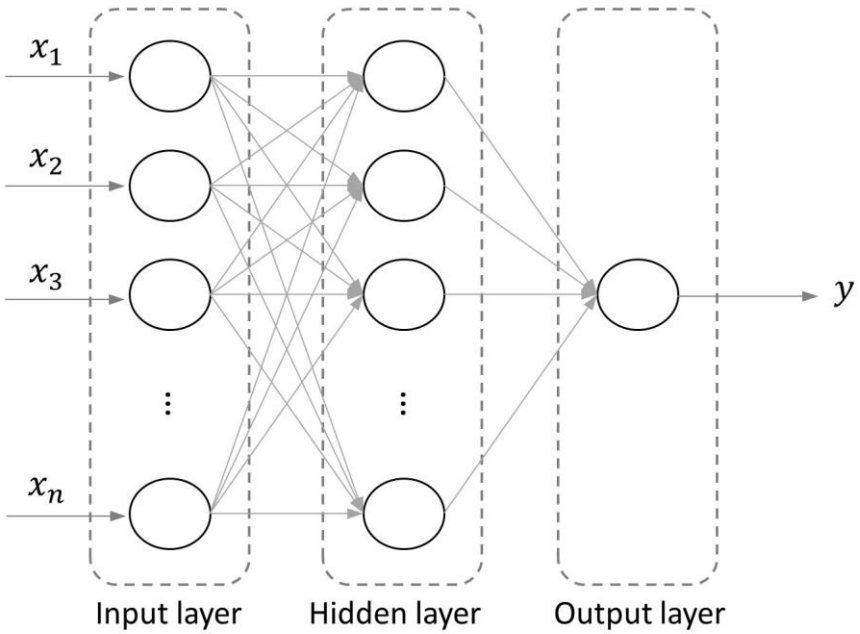

Figure 1. Basic structure of a single ANN model comprising three layers (input, hidden, and output).

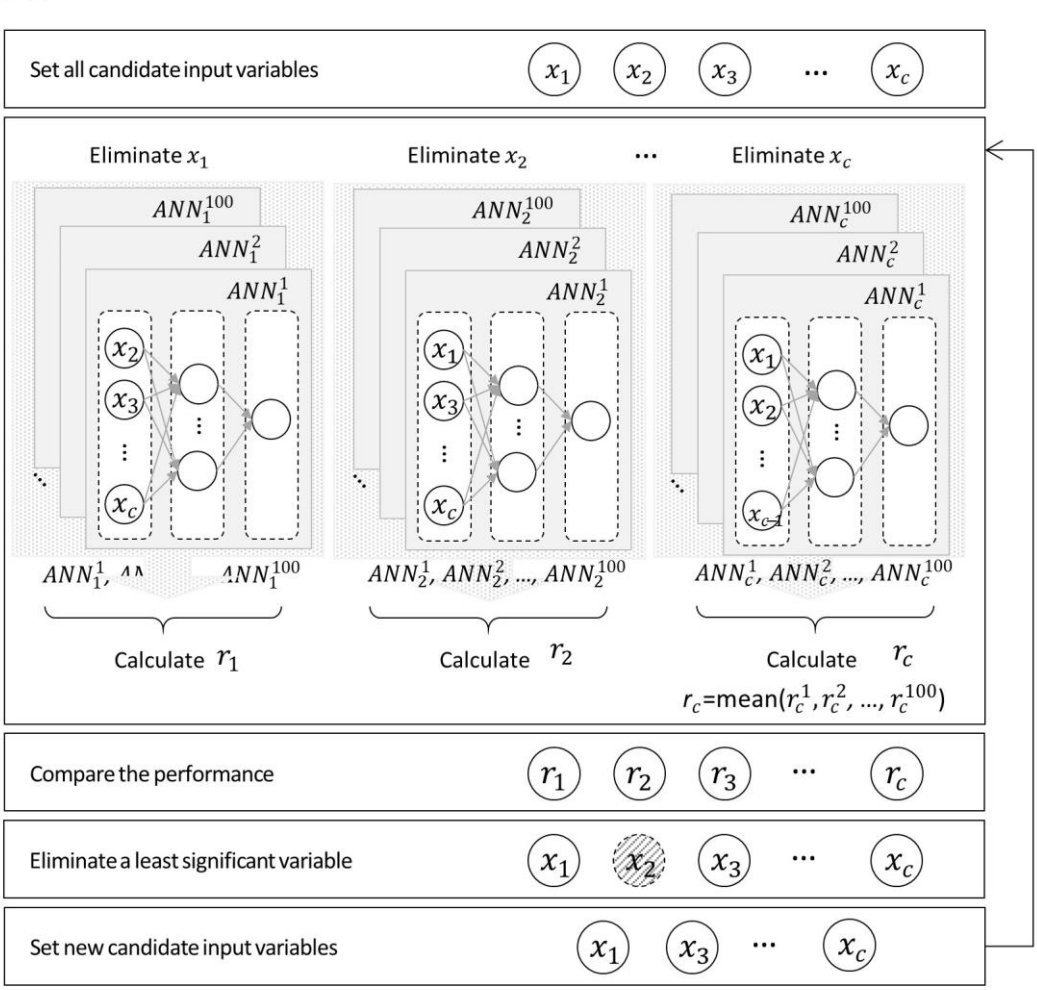

Figure 2. Proposed modified backward elimination procedure for the ANN model.

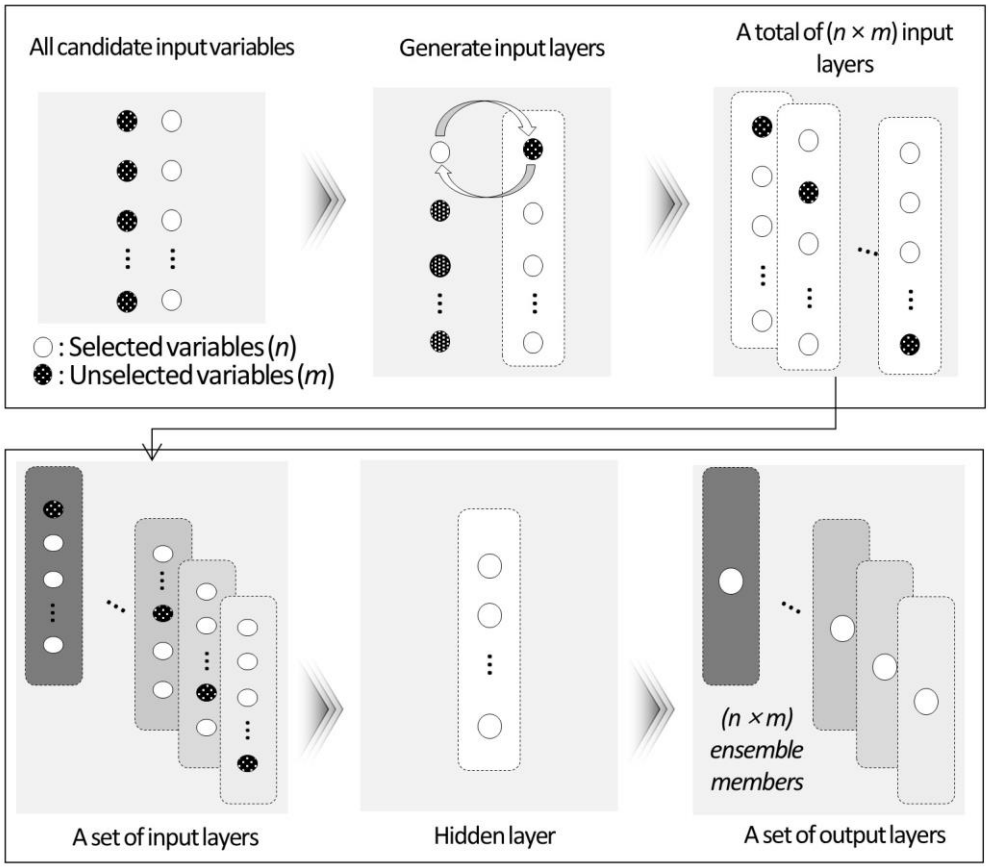

Figure 3. Consideration of uncertainty in input variable selection by generating a set of input layers.

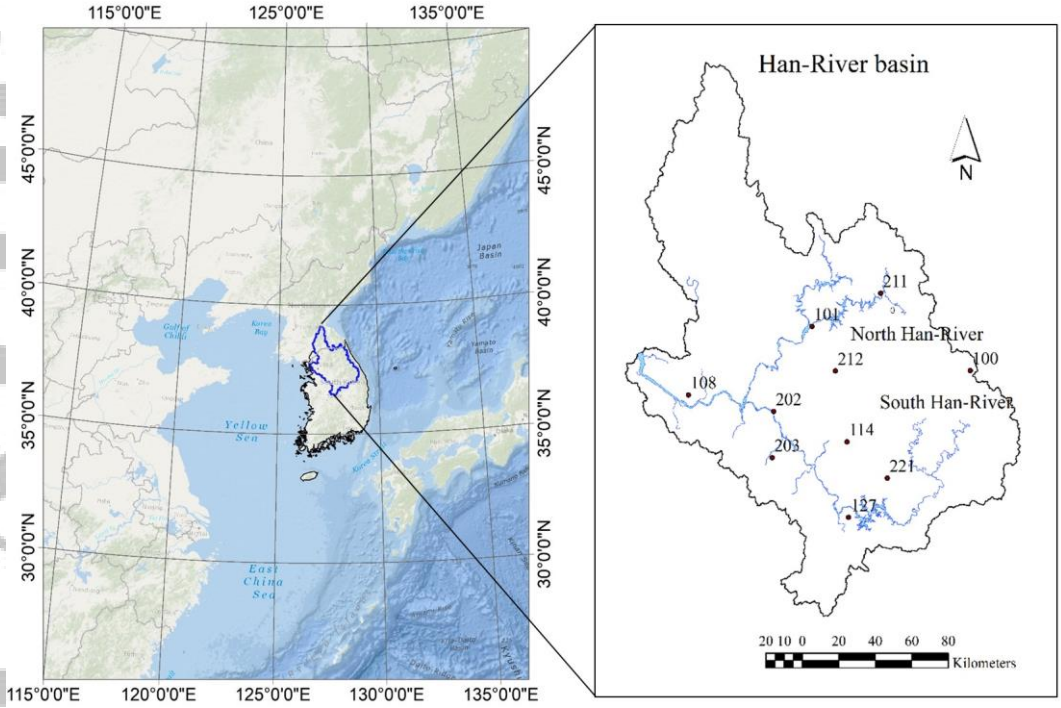

Figure 4. Location of the 10 meteorological stations used in this study.

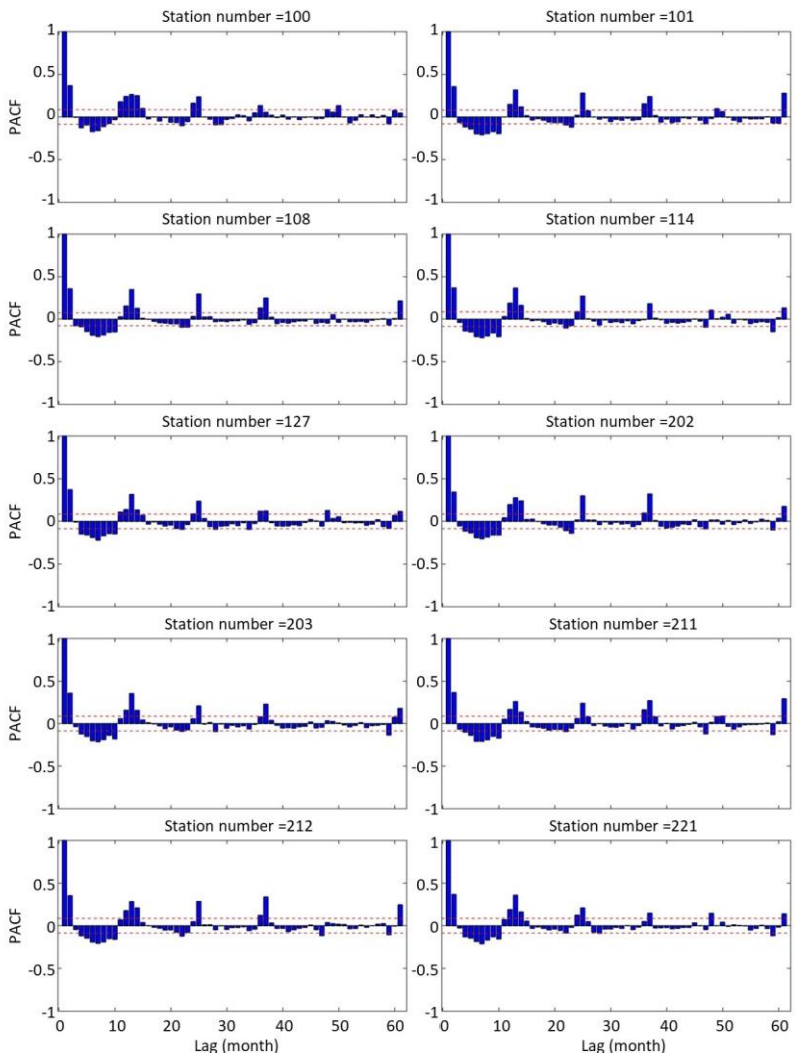

Figure 5. Partial autocorrelation function plots of monthly precipitation at all meteorological stations.

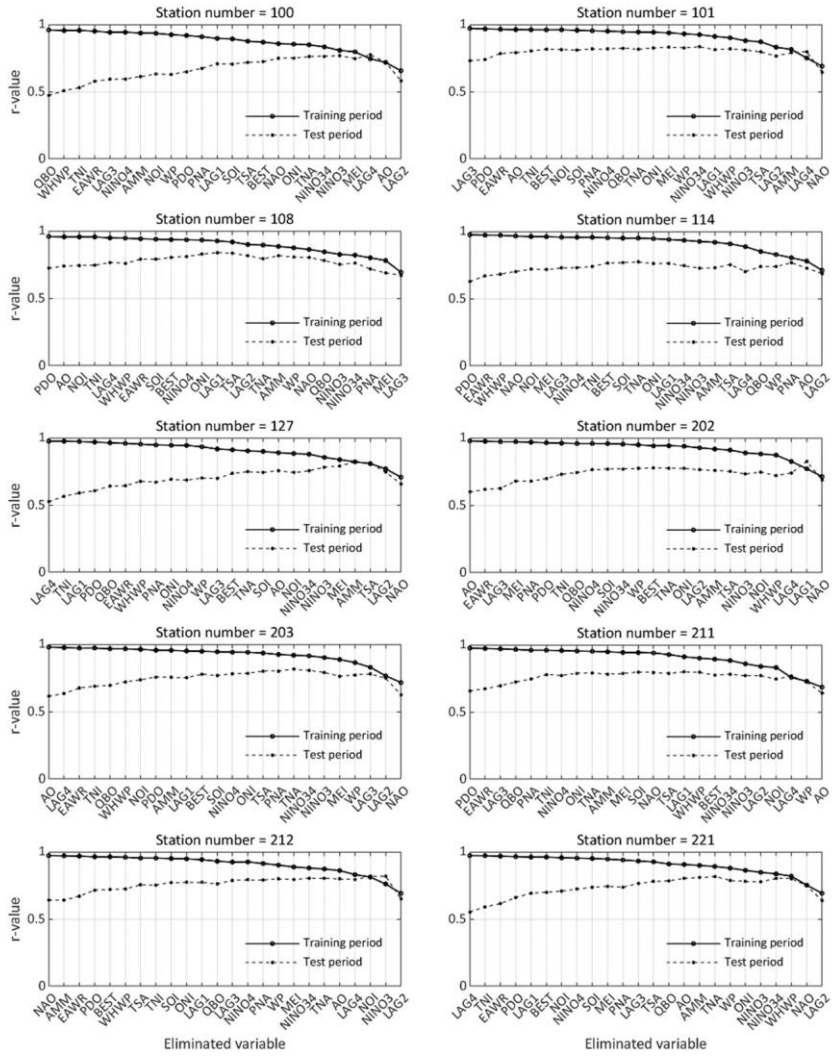

Figure 6. *r* values determined by modified backward elimination for all stations. y axis shows the mean of *r* values in the training and test periods from 100 S-ANNs, x axis shows the variables eliminated at each step through the modified backward elimination process, and *r* values at the end of the plots indicate the mean from 100 S-ANNs with only the three fixed key climate indices as input variables.

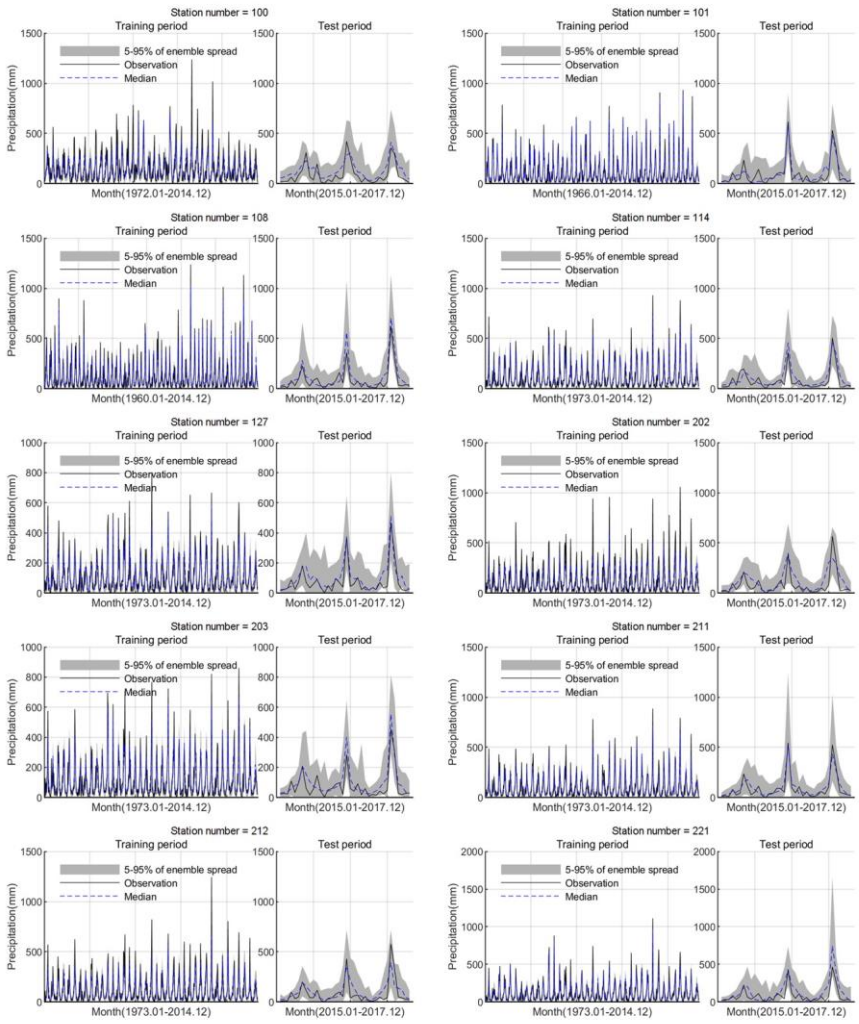

Figure 7. ANN-ENS results from the training and test periods at all stations. Black solid lines are the observed monthly precipitation and gray shading covers the 5th to 95th percentiles of the ensemble members, representing the uncertainty in input variable selection. A broader ensemble range indicates higher uncertainty included in the input variables. Blue dotted lines are the median of the ensemble members representing the forecasting result.

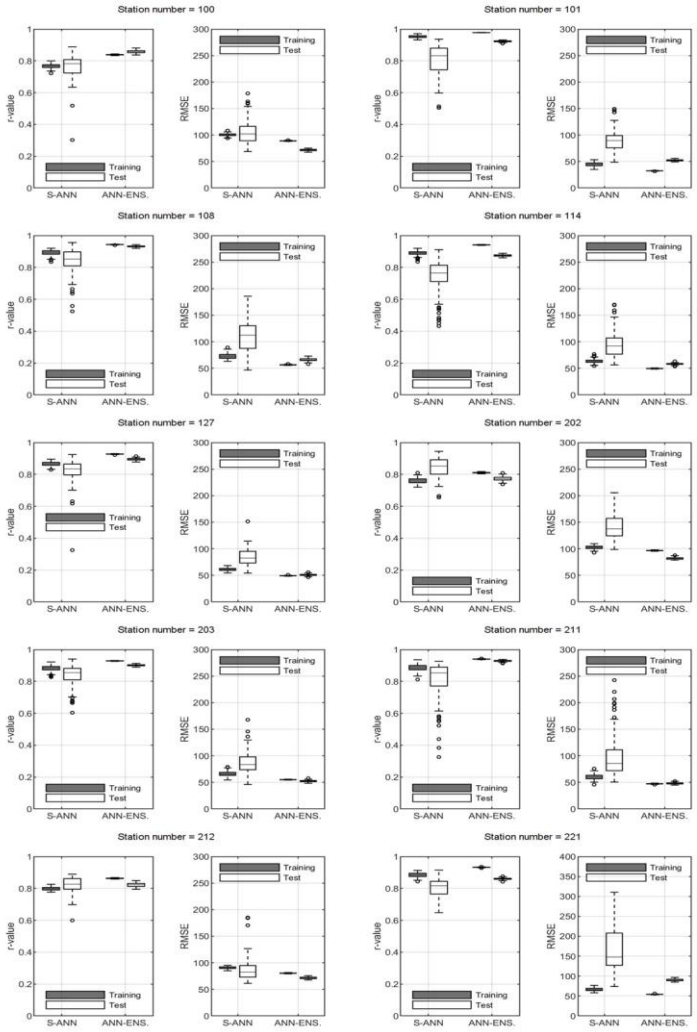

Figure 8. Box plots of the 100 *r* (left) and RMSE (right) values for the S-ANN and ANN-ENS at all stations.

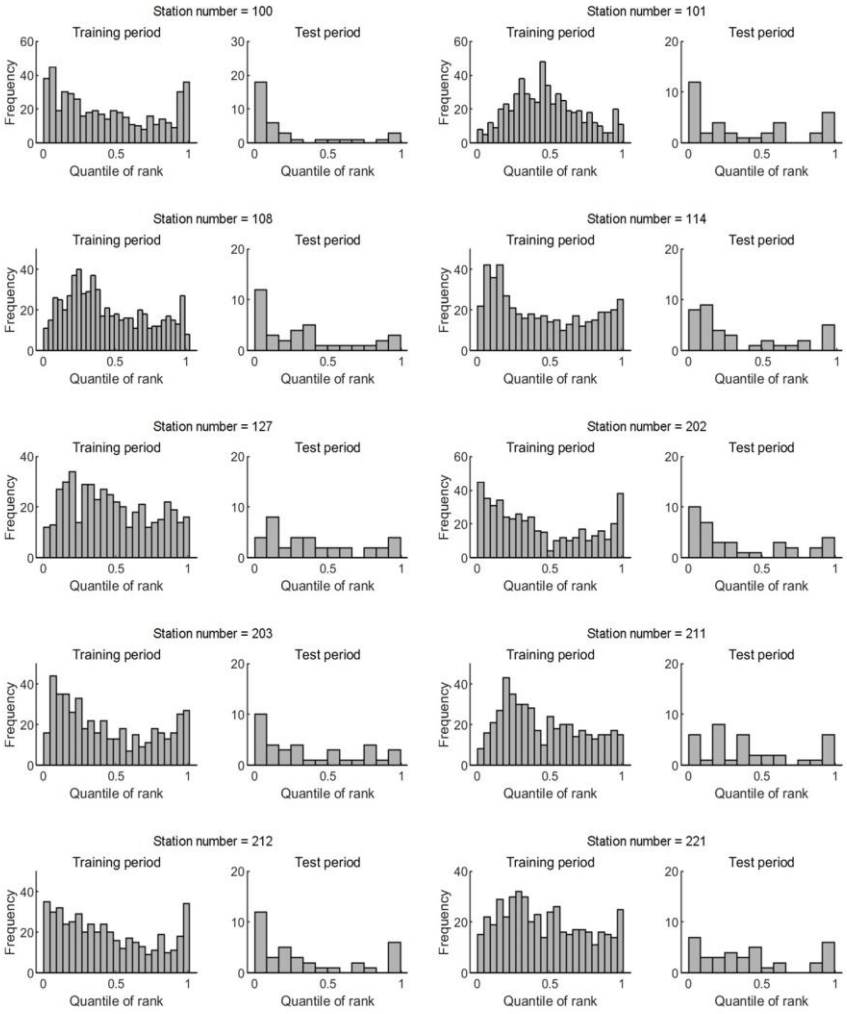

Figure 9. Rank histograms of the ANN-ENS ensemble members in the training and test periods at all stations.

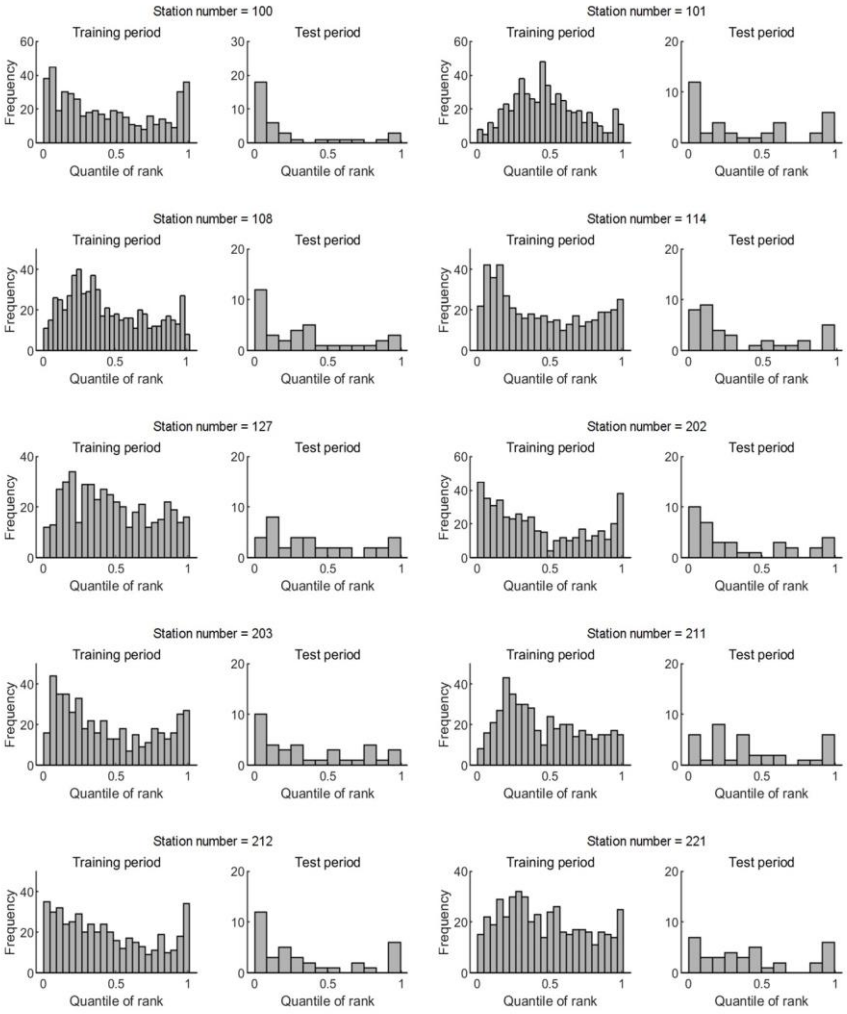

Figure 10. Reliability diagrams of the ANN-ENS ensemble members in the training and test periods at all stations.