Monthly Precipitation Forecasting in the Han River Basin, South Korea, Using Large-Scale Teleconnections and Multiple Regression Models

Chul-Gyum Kim *, Jeongwoo Lee, Jeong Eun Lee, Nam Won Kim and Hyeonjun Kim

Department of Land, Water and Environment Research, Korea Institute of Civil Engineering and Building Technology, 283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 10223, Korea; ljw2961@kict.re.kr (J.L.); jeus22@kict.re.kr (J.E.L.); nwkim@kict.re.kr (N.W.K.); hjkim@kict.re.kr (H.K.)

* Correspondence: cgkim@kict.re.kr; Tel.: +82-31-910-0545

Received: 28 April 2020; Accepted: 1 June 2020; Published: 3 June 2020

Abstract: In this study, long-term precipitation forecasting models capable of reflecting constantly changing climate characteristics and providing forecasts for up to 12 months in advance were developed using lagged correlations with global and local climate indices. These models were applied to predict monthly precipitation in the Han River basin, South Korea. Based on the lead month of forecast, 10 climate indices with high correlations were selected and combined to construct four-variable multiple regression models for monthly precipitation forecasting. The forecast results for the analytical period (2010–2019) showed that predictability was low for some summer seasons but satisfactory for other seasons and long periods. In the goodness-of-fit test results, the Nash–Sutcliffe efficiency (0.48–0.57) and the ratio of the root mean square error to the standard deviation of the observation (0.66–0.72) were evaluated to be satisfactory while the percent bias (9.4%–15.5%) was evaluated to be between very good and good. Due to the nature of the statistical models, however, the predictability is highly likely to be reduced if climate phenomena that are different from the statistical characteristics of the past appear in the forecast targets or predictors. The forecast results were also presented as tercile probability information (below normal, normal, above normal) through a comparison with the observation data of the past 30 years. The results are expected to be utilized as useful forecast information in practice if the predictability for some periods is improved.

Keywords: monthly precipitation forecasting; large-scale teleconnections; multiple regression models; climate indices

1. Introduction

Reliable prediction of precipitation is essential for stable operation and management of water resources. In particular, a long-term forecast ranging from one month to several months, which corresponds to seasonal forecasts, is very important for responding to disasters caused by extreme weather events, such as droughts, floods, and heat waves, and securing stable water resources. Forecasting models have been widely used to predict climate elements, such as precipitation and temperature. These models can be mainly divided into two types depending on their approach [1–3]. The first type uses dynamic models that obtain numerical solutions through various equations that describe the fluid phenomena in the atmosphere, ocean, and land. The general circulation models (GCMs) are widely used here. The second type performs forecasting through statistical relationships between the forecast targets and the atmosphere-ocean-land data.
The statistical approaches used included analog methods, time series models, multiple regression models, canonical correlation models, and artificial neural network models.

Some countries, including South Korea, Australia, the United States, the United Kingdom, and Japan, use numerical dynamic models to provide monthly and seasonal forecasting data on precipitation and temperature. With the development of modeling and computing technologies, there is an increase in the short-term (a few days or less) predictability through the reproduction of complex and nonlinear phenomena. However, the accuracy of long-term prediction of more than several months is still not sufficient when using numerical dynamic models, due to the uncertainty caused by initial conditions and the increasing error caused by the integration time of these models [1,4,5]. Statistical models, on the other hand, are relatively flexible in model construction and have a potential to improve the forecast period and predictability depending on the accumulation of observations and related data, although their predictability is unstable compared to dynamic models. Until 2013, the long-term weather forecast in Australia was performed using statistical techniques [6], and the runoff forecast data for major nationwide branches are still being generated using statistical techniques. Statistical techniques are also utilized in South Korea, the United States, and Japan, and forecasts based on various statistical approaches have been attempted by many studies due to the benefits of statistical models compared to dynamic models despite the fact that they cannot explain physical relationships among natural phenomena. A representative statistical approach in recent years is to derive monthly or seasonal forecast information by constructing forecasting models, such as multiple regression and artificial neural network models, through the teleconnections with large-scale climate indices and forecast targets (e.g., precipitation). Many statistical forecasts have been performed with the focus on precipitation and temperature, as well as on runoff [7–12], snowfall [13–15], crop yield [16–19], and brown planthopper infestation [20].

In South Korea, seasonal forecasting technology has been actively developed since the 2000s. Many studies on monthly and seasonal precipitation forecasting have also been conducted using regression models [9,21–27], canonical correlation models [28,29], and artificial neural network models [30–32].

There are two limitations in precipitation forecasting studies using teleconnections with climate indices, other than the predictability issue. One limitation is that it is difficult to judge predictability on future periods that were not used in the analysis even though the studies [21,25–27,33,34] generally present optimized results for calibration and validation periods by dividing the past analytical period. This means that if forecasting models are derived based on the correlation results for a certain period only, the utilization of the models can be difficult when teleconnections vary due to climate change in the future. The other limitation is that the lead time of the forecast is limited by the lead time of the predictors. For example, if the analysis results for correlations with climate data that occurred in the same month as the forecast target are presented [35–37], it is not possible to forecast more than one month using these results although statistical interactions can be identified. In other words, when n-month preceding data of climate indices are selected as predictors, only n months or less can be forecasted.

Therefore, this study was conducted with a focus on building more flexible models and providing forecast information to improve utilization for practical purposes. The monthly precipitation of the Han River basin in South Korea was set as the forecast target. The 39 global climate indices and 6 local climate indices in the target basin, including precipitation and temperature, were utilized as predictors. Statistical multiple regression models were used as forecasting technique, and forecasting models capable of predicting up to 12 months in advance were constructed by analyzing the lagged correlations with climate indices and selecting the optimal predictors according to the forecast lead time. The range of forecast values were derived from a number of forecasting models constructed for each target month, and tercile probability information was also presented for monthly forecasts through a comparison with the observation data of the past 30 years.
2. Materials and Methods

2.1. Study Area and Data Sets

The Han River basin, which is the focus area of this study, is located in the center of the Korean Peninsula as shown in Figure 1. It is a large basin that represents approximately 19% (41,947 km²) of the total area of the Korean Peninsula. This area includes Seoul, the capital of South Korea, with more than 27 million people. Table 1 shows the names and geographical locations of the automated synoptic observation system (ASOS) stations of the Korea Meteorological Administration (KMA).
Table 1. ASOS stations used in this study.

| ID | Station name       | Latitude (°N) | Longitude (°E) | Elevation (m a.s.l.) |
|----|--------------------|--------------|----------------|----------------------|
| 90 | Sokcho             | 38.25        | 128.56         | 18.06                |
| 93 | Bukchuncheon       | 37.95        | 127.75         | 95.61                |
| 95 | Cheorwon           | 38.15        | 127.30         | 155.48               |
| 98 | Dongducheon        | 37.90        | 127.06         | 115.62               |
| 99 | Paju               | 37.89        | 126.77         | 30.59                |
| 100| Daegwallyeong      | 37.68        | 128.72         | 772.57               |
| 101| Chuncheon          | 37.90        | 127.74         | 76.47                |
| 104| Bukgangneung       | 37.80        | 128.86         | 78.90                |
| 105| Gangneung          | 37.75        | 128.89         | 26.04                |
| 106| Donghae            | 37.51        | 129.12         | 39.91                |
| 108| Seoul              | 37.57        | 126.97         | 85.67                |
| 112| Incheon            | 37.48        | 126.62         | 68.99                |
| 114| Wonju              | 37.34        | 127.95         | 148.60               |
| 116| Gwanaksan          | 37.44        | 126.96         | 626.76               |
| 119| Suwon              | 37.27        | 126.99         | 34.84                |
| 121| Yeongwol           | 37.18        | 128.46         | 240.60               |
| 127| Chungju            | 36.97        | 127.95         | 116.30               |
| 129| Seosan             | 36.78        | 126.49         | 25.86                |
| 130| Uljin              | 36.99        | 129.41         | 50.00                |
| 131| Cheongju           | 36.64        | 127.44         | 58.70                |
| 201| Ganghwa            | 37.71        | 126.45         | 47.84                |
| 202| Yangpyeong         | 37.49        | 127.49         | 47.26                |
| 203| Icheon             | 37.26        | 127.48         | 80.09                |
| 211| Inje               | 38.06        | 128.17         | 200.16               |
| 212| Hongcheon          | 37.68        | 127.88         | 139.95               |
| 214| Samcheok           | 37.37        | 129.22         | 3.90                 |
| 216| Taebaek            | 37.17        | 128.99         | 712.82               |
| 217| Jeongseongun       | 37.38        | 128.65         | 307.40               |
| 221| Jecheon            | 37.16        | 128.19         | 259.80               |
| 226| Boeun              | 36.49        | 127.73         | 174.99               |
| 232| Cheonan            | 36.76        | 127.29         | 81.50                |
| 271| Bonghwa            | 36.94        | 128.91         | 324.30               |
| 272| Yeongju            | 36.87        | 128.52         | 210.79               |
| 273| Mungyeong          | 36.63        | 128.15         | 170.61               |

The monthly precipitation for the last 10 years (2010–2019) was selected as forecast targets. A total of 45 indices, including 39 global climate indices, such as the Antarctic oscillation (AAO) and the Atlantic meridional mode (AMM), provided by the National Oceanic and Atmospheric Administration (NOAA) and 6 local climate indices for the target basin were utilized as predictors. Table 2 summarizes the global and local climate indices that were used as predictors, where the global indices are monthly data provided by each institution.

To analyze the lagged correlations between precipitation and predictors, the data of the predictors from 1968 to 2019 were used, where the data of precipitation and other local climate indices were area-averaged using the monthly data from 34 ASOS stations of the KMA in and around the basin.
Table 2. Predictors used in this study.

| Predictor   | Description                                           | Provider |
|-------------|-------------------------------------------------------|----------|
| AAO         | Antarctic oscillation                                  | NOAA     |
| AMM         | Atlantic meridional mode                              | NOAA     |
| AMO         | Atlantic multidecadal oscillation                      | NOAA     |
| AO          | Arctic oscillation                                     | NOAA     |
| BEST        | Bivariate ENSO timeseries                             | NOAA     |
| CAR         | Caribbean SST index                                   | NOAA     |
| CPOLR       | Central Pacific outgoing long wave radiation index (170E–140W, 5S–5N) | NOAA     |
| EA          | East Atlantic pattern                                 | NOAA     |
| EAWR        | East Atlantic / Western Russia pattern                 | NOAA     |
| EPNP        | East Pacific / North Pacific oscillation               | NOAA     |
| GML         | Global mean land-ocean temperature index              | NASA     |
| MEI.v2      | Multivariate ENSO index version 2                     | NOAA     |
| NAO         | North Atlantic oscillation                            | NOAA     |
| NOI         | Northern oscillation index                            | NOAA     |
| NP          | North Pacific pattern                                 | NCAR 1   |
| NTA         | North tropical Atlantic SST index                     | NOAA     |
| ONI         | Oceanic Niño index                                    | NOAA     |
| PNA         | Pacific American index                                | NOAA     |
| POL         | Polar / Eurasia pattern                               | NOAA     |
| QBO         | Quasi-biennial oscillation                            | NOAA     |
| SCAND       | Scandinavia pattern                                  | NOAA     |
| SLP_DAR     | Darwin sea level press                                | NOAA     |
| SLP_EEP     | Equatorial eastern Pacific sea level press            | NOAA     |
| SLP_INDO    | Indonesia sea level press                             | NOAA     |
| SLP_TAH     | Tahiti sea level press                                | NOAA     |
| SOI         | Southern oscillation index                            | NOAA     |
| SOLAR       | Solar flux (10.7 cm)                                  | NOAA     |
| TNA         | Tropical northern Atlantic index                      | NOAA     |
| TNI         | Trans-Niño index                                      | NOAA     |
| TPI         | Tripole index for the interdecadal Pacific oscillation| NOAA     |
| TSA         | Tropical southern Atlantic index                      | NOAA     |
| WHWP        | Western hemisphere warm pool                          | NOAA     |
| WP          | Western Pacific index                                 | NOAA     |
| EASMI       | East Asian summer monsoon index                       | [38]     |
| PCP         | Monthly precipitation                                 | KMA      |
| TMP         | Monthly average temperature                           | KMA      |
| VapPr       | Monthly average vapor pressure                        | KMA      |
| HMD         | Monthly average relative humidity                     | KMA      |
| LocPr       | Monthly average local pressure                        | KMA      |
| HrDL        | Monthly sum of daylight hours                         | KMA      |

1 The National Center for Atmospheric Research.

2.2. Teleconnection Analysis

In order to determine predictors for constructing forecasting models, the correlation between the past precipitation data and 1–18 months of the preceding data for each predictor were analyzed.
In this instance, the delayed correlations between the forecast target (precipitation) and predictors were analyzed using the data of the past 40 years based on the target month considering the short-term variability of each index and their long-term variability due to climate change.

Figure 2 shows the correlations between the precipitations and the preceding data for 1–18 months for each index from the past 40 years (1970–2009) based on January and July 2010, and Figure 3 shows the results of analyzing the correlations between the monthly precipitation and each predictor over the past 40 years (1979–2018) based on January and July 2019. The interpretation of correlation coefficients differs depending on the field of study [39], and as suggested by Evans [40], in statistics, the absolute values above 0.4 are considered to be moderate. Therefore, in this study, when the absolute value of the correlation was 0.4 or higher, the numerical values were also indicated. Red indicates a positive correlation, and blue indicates a negative one. The absence of color means that there is no climate index available for that period.

As shown in Figure 2(a), the correlations between the precipitation in January from 1970 to 2009, and the predictors revealed the highest value, $-0.483$ in the EASMI data of the preceding 17 months, i.e., EASMI (17). Other significant correlations were found in TPI (7) ($-0.420$) and HrDL (9) ($0.407$). Meanwhile, as shown in Figure 2(b), TNA (12) ($0.539$) exhibited the highest correlation with the precipitation in July from 1970 to 2009, followed by NTA (11) ($0.520$) and NTA (12) ($0.517$). The correlation results were different for each month of each year. In general, significant positive correlations were observed for summer (July to September), and significant negative correlations were observed for winter.

In this study, forecasting models were constructed using 10 indices that exhibited high correlations for each target month. Figure 4 shows the correlations of the climate indices that were used for constructing monthly forecasting models from January to December 2010, issued in December 2009. Figure 5 shows the results for the highly correlated indices used in the monthly forecast models issued in December 2018. As shown in Figures 2 to 5, many predictors with negative correlations were selected for winter forecasting, while many predictors with positive correlations were utilized for summer.

![Figure 2](image-url)  
Figure 2. Results of the correlation analysis with predictors for forecasting the precipitation of (a) January and (b) July 2010.
Correlation coefficients between observations and climate indices for forecasting data in 2019-01

Correlation coefficients between observations and climate indices for forecasting data in 2019-07

Figure 3. Results of the correlation analysis with predictors for forecasting the precipitation of (a) January and (b) July 2019.

Correlation coefficients of climate indices used to forecast issued in 2009-12

Figure 4. Correlations of the climate indices used for monthly forecasting models from January to December 2010.
2.3. Forecasting Model Construction

In this study, a multiple regression model with a simple configuration was selected as a forecast model. In regression analysis, many researchers, including Altman [41], said that there should be at least 10 observations per variable. If too little variables are selected, the reliability of the regression model may decrease. In this study, we used data from the past 40 years (i.e., 40 monthly data) to derive the forecast models for each month, so a regression equation with four independent variables was used and constructed as follows:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon, \]  

where \( Y \) is the response (monthly precipitation); \( \beta_0, \beta_1, \beta_2, \beta_3, \beta_4 \) are the regression coefficients; \( X_1, X_2, X_3, X_4 \) are the highly correlated climate indices (predictors); and \( \epsilon \) is the residual term of the model.

As shown in Figures 4 and 5, 10 indices with high correlations were extracted based on the correlation analysis results for each target month, and multiple regression models were constructed by combining four indices among these. In this instance, the predictors that can be used were limited based on the lead month of forecast. For example, in the case of forecasting the precipitation for May 2010 in December 2009 (i.e., forecasting five months in advance), the climate data with the preceding periods of less than 5 months were excluded from predictors even if they had high correlations. In the course of deriving regression models consisting of 4 independent variables among 10 indices with high correlations, regression models that had multicollinearity between the variables or those that generated a negative value for the target month were excluded. In addition, leave-one-out cross-validation was performed for the data over a period of 40 years to improve the goodness of fit of the forecasting models; the forecasts were finally derived by selecting 2520 regression models for each month. Models for forecasting up to 12 months ahead from the time of the forecast were constructed. To evaluate the predictability of the models, forecasts for the period from January 2010
to December 2019 were derived, and the results were compared with the observed data in the corresponding period.

3. Results and Discussion

3.1. Application Frequency of Climate Indices

Figures 6 and 7 show the frequency of the climate indices that were used in the 2520 regression models, which were finally selected by combining 10 climate indices with high correlations for each month, based on the correlation analysis results in Figures 4 and 5, respectively. The high correlations did not necessarily result in a high application frequency, and most of the selected climate indices were used evenly.

**Figure 6.** Application frequency of the climate indices used for monthly precipitation forecasting in 2010.
Figure 7. Application frequency of the climate indices used for monthly precipitation forecasting in 2019.

Figure 8 shows the 10 indices with the highest application frequency for the entire period (2010–2019) by analyzing the application frequency of climate indices for each month (January to December). The results were collected for all lead months, and it was found that the climate indices with relatively longer lead times exhibited a high application frequency. For precipitation forecasting in January, PNA (16), which is the 16-month preceding data of the PNA index, SO_EQ (10), and EASM (17) were applied in many models. For July, CAR (13), WHWP (12), and TNA (12) were frequently applied.
When comparing the results of the previous studies that analyzed correlations between precipitation in the Korean Peninsula and global climate indices with the results presented in Figure 8, common results were found for some indices although the areas and periods were different. Cho et al. [9] used the indices of AO (5) in February, AO (8) in May, and NAO (6) in June for predicting monthly precipitation in the Andong Dam basin. Kim and Kim [24] analyzed correlations between the global climate indices and the monthly precipitation of 60 weather stations in South Korea and found that the indices of AO (5) and NAO (6) were highly correlated with the precipitation in February and June, respectively. Kim et al. [25] proposed a multiple regression model using AO (4) with a high correlation to forecast the precipitation in October at 12 weather stations in South Korea. Kim et al. [36] found that AO (5), NAO (6), and NOI (5) had a high correlation with the precipitation in February, June, and November, respectively, through a teleconnection analysis with global climate indices. The results of this study were therefore similar to those of the previous studies in forecasting the precipitation in South Korea for the months of February, May, June, October, and November.

3.2. Results of Monthly Precipitation Forecasting

Figure 9 shows the results of forecasting that was performed in December 2009, and a comparison of the forecasts with the observed values for the period from January to December in 2010 are illustrated. January 2010 represents the forecast result of one month ahead and December 2010 represents the forecast result of 12 months ahead. As mentioned above, 2520 forecasts were derived for each month, and the forecast ranges were drawn using box plots. These box plots describe the 25th, 50th (i.e., the median), and 75th percentiles of the forecasts, and the whiskers refer to the maximum and minimum values excluding outliers. In the same manner, Figure 10 shows the results of the forecasting from January to December 2019, which was forecasted in December 2018.
As shown in the above two examples, the forecast ranges for spring and winter, when precipitation is relatively low, were narrow and not significantly different from the observed values. However, the range was very wide in the summer when there was a large amount of precipitation, and the difference from the observed value was also large for some months.

Figures 11 and 12 show the 12-month forecast results predicted in June 2010 and in June 2018, respectively. Compared to the results in Figures 9 and 10, the forecast ranges were different depending on the lead time of forecasting, but the impact of the lead time was not significant.
As shown in Figures 9 to 12, it is possible to derive the forecast values for the next 12 months based on the forecast time. Figure 13 compares the average values of the monthly forecasts with the observed values for the entire analytical period (January 2010 to December 2019). As up to 12 average values for the forecasts for each month were generated depending on the lead month (LM) of the forecast, these were divided by shades. As shown in Figure 13, there was no significant difference in forecast depending on the lead month; however, there were clear differences from summer observation data for certain years. In July 2011 and 2013 when there was much precipitation, the models were not sufficient in reproducing the phenomena of heavy rain. Meanwhile, for 2014, 2015, and 2019, when the precipitation was low, the predictability of summertime was found to be relatively low. For the other periods, however, the tendencies of the observation data were reproduced relatively well.
Figure 13. Results of monthly precipitation forecasting according to the lead month of forecast.

Figure 14 shows the results of the goodness-of-fit test analyzed by the lead time for the forecast results presented in Figure 13. The percent bias (PBIAS) ranged from 9.4% to 15.5%, and the ratio of the root mean square error (RMSE) to the standard deviation of the observation (RSR) ranged from 0.66 to 0.72. The Nash–Sutcliffe efficiency (NSE) ranged from 0.48 to 0.57, and the Pearson correlation coefficient (r) ranged from 0.73 to 0.78. When compared with the evaluation grade of each index suggested by Moriasi et al. [42], NSE and RSR were evaluated to be satisfactory while PBIAS was rated to be between very good and good. In other words, the predictability for some specific periods was low, but the predictability over a long period was found to be satisfactory. There was no tendency according to the lead month of forecast (1–12 months of lead time), and some goodness-of-fit indices exhibited better results for the 12-month lead time. These appear to be the characteristics of statistical models.
3.3. Probabilities of Monthly Precipitation Forecasts

Some countries that perform long-term forecasts by government agencies, such as South Korea, Australia, the United States, the United Kingdom, and Japan, provide tercile probability information (below normal, normal, above normal) through a comparison with historical observations. In this study, the tercile probability was analyzed for the 2520 forecasts of each month compared to the observed precipitation data over the past 30 years.

Figure 15(a) compares the forecast results from January to December 2010 predicted in December 2009 with the observation data of the past 30 years for each month, using box plots, and Figure 15(b) also compares the forecasts for the next 12 months predicted in December 2018 and the observations over the past 30 years. From the figures, it is possible to qualitatively judge the levels of the forecast results compared to the past observation data. Figure 16 quantitatively shows the results in Figure 15 as a tercile probability information. Figure 16 shows that more precipitation than normal was forecasted for February and July in 2010, and a normal precipitation was highly probable for January and April 2019.
4. Conclusions

In this study, statistical precipitation forecasting models capable of reflecting long-term climate change and providing forecasts up to 12 months in advance were developed through the continuous update of the past climate data and the use of selected predictors according to the forecast lead time.

As a test, the monthly precipitation of the Han River basin was set as the forecast target, and 1–18 months of lagged correlations were analyzed for 45 global and local climate indices. Ten climate indices with high correlations were selected for each target month according to the lead month of
forecast. These were combined to construct four-variable multiple regression models for monthly precipitation forecasting. In this process, many indices with positive correlations were selected for summer, and indices with negative correlations were selected for winter. In the course of deriving regression models, the models that had multicollinearity between the independent variables or derived values smaller than zero for the target month were excluded. In addition, leave-one-out cross-validation was performed for the data of the past 40 years to derive 2520 forecasting models with good predictability for each month.

When examining the climate indices used in the monthly forecasting models for the entire period (2010–2019), we found that similar indices were also used in previous studies for the analysis of precipitation in South Korea for the months of February, May, June, October, and November. The results of comparing monthly forecasts (forecast ranges) with observation data for the period between 2010 and 2019 revealed that predictability was relatively low for the summer of 2011 and 2013 when the precipitation was high due to sudden heavy rain, and for the summer of 2014, 2015, and 2019 when the precipitation was significantly lower than normal. For the other periods, however, the tendencies of the observation data were reproduced relatively well. In the goodness-of-fit analysis results, NSE and RSR were evaluated to be satisfactory while PBIAS was evaluated to be between very good and good. In other words, predictability for some specific periods was low, but predictability over a long period was found to be satisfactory. The monthly forecast results were also presented as tercile probability information (below normal, normal, above normal) through a comparison with the observation data of the past 30 years to be used for practical purposes in the future.

The novelty of this study lies in the fact that, unlike previous studies, it developed a method that allows forecasts for future periods (up to 12 months), by updating the relationships between the forecast targets and predictors of each month based on the statistical characteristics of the past data. However, the predictability of some summer periods is very low. This is attributed to irregular rainfall characteristics (such as heavy rains caused by typhoons and monsoonal front) in summer in northeast Asia, including Korea. In particular, the prediction of summer precipitation is a very important issue in water management related to the response to flooding. Due to the limitations of statistical models, however, predictability is bound to decline if precipitation or climate indices that are different from the past statistical characteristics occur. That is, for now, there is a problem of poor predictability due to insufficient past available data required to find a statistical relationship, but in the future, if the data continues to accumulate and some statistical relationship is explained from the accumulated data, the predictability for a specific period may improve. In future studies, it will be important to track predictors that can reflect abnormal weather conditions (e.g., sudden typhoons, heavy rain, and temporary droughts) and to improve forecasting models that consider such predictors.

5. Discussions

The investigation of teleconnections with climate signals, as performed by many researchers, is the first process to predict precipitation statistically using the relationships with climate indices. Many studies have been successful in deriving meaningful correlations between precipitation and climate indices [2,21,24–31,33,37], but using these correlations to obtain reliable long-term precipitation estimates is still difficult. As shown in this study, the results show high accuracy for regions or periods with relatively stationary climate characteristics, but the prediction accuracy is still low for irregular climatic conditions, making it difficult to use in practice. In particular, the predictability of the heavy rains caused by typhoons or monsoonal fronts in northeast Asia is significantly lower. The reasons for this may be a failure to find a signal of a climate index related to such a heavy rain, a failure to properly understand the signal of precipitation, or a lack of a significant teleconnection between the precipitation and climate indices in the historical data. Since the relationship between precipitation and climate indices is very complicated, and the signals of precipitation and climate indices have regular or irregular variability depending on temporal scales, various studies have been conducted to analyze these variabilities. He and Guan [43] developed a
wavelet-based method to solve these problems, presenting three difficulties (interdependent difficulty of explanatory variables, multitemporal component difficulty, nonuniform time-lag difficulty) in the process of correlation and variability analysis.

For the interdependent difficult, a multicollinearity validation was performed to avoid intercorrelations between the climate indices in this study. The variation inflation factor (VIF) was used as a criterion for multicollinearity. Various recommendations for acceptable levels of VIF have been published in the literature, and this study applied a value of 10, which has been mostly recommended as the maximum level of VIF [44–47].

Multitemporal component analysis is considered to be a very useful way to examine several limited climate indices. However, this study tried to include as many climate indices as possible. As a result, a total of 45 indices (39 global and 6 local) were used as candidates as predictors. Using many climate indices does not necessarily increase predictability, but it can increase the utility in constructing statistically flexible models. A direct statistical relationship was analyzed for the original time series between precipitation and climate indices. As He and Guan [43] mentioned, there is a limitation that a direct comparison of the original time series can only reveal part of the relationship, and if various correlations for different temporal scales are found, it can be very useful for prediction. In the future, more detailed analyses of the correlations at different temporal scales will be performed for the climate indices that have been found to be highly correlated in this study through other studies.

For the non-uniform time-lag difficulty, as in the examples of Figures 2 and 3, since the correlation was analyzed using data from the past 40 years according to the forecast month, the results of the correlation analysis differ from month to month. This is the difference from previous studies. Most of the existing studies [2,21,23–25,27,31,33,34] have constructed a “fixed” model using these climate indices after investigating the lag time (month) and the climate indices that showed the maximum (or higher) correlation among all available historical data. If the signal characteristics of the climate indices differ from the past (i.e., the statistical characteristics vary with time), there will be restrictions on the continued use of the “fixed” model in the future. However, in this study, the models were constructed by flexibly analyzing the corresponding past data according to the forecast month, and even if the climate characteristics changed, the statistical relationship was newly derived from the moved historical data to update the forecast models. These “flexible” models have the advantage of being able to actively respond to long-term climate change.

Author Contributions: All authors substantially contributed in conceiving and designing of the research and realizing this manuscript. Conceptualization, C.-G.K. and N.W.K.; Data curation, J.L. and J.E.L.; Formal analysis, C.-G.K. and J.L.; Funding acquisition, H.K.; Investigation, C.-G.K. and J.E.L.; Methodology, C.-G.K., J.L. and N.W.K.; Project administration, N.W.K. and H.K.; Resources, J.E.L.; Supervision, N.W.K. and H.K.; Validation, J.L. and H.K.; Visualization, C.-G.K. and J.E.L.; Writing—original draft, C.-G.K. All authors have agreed to the published version of the manuscript.

Funding: This research was funded by the Korea Institute of Civil Engineering and Building Technology (grant number 20200041-001) and the APC was funded by the Korea Institute of Civil Engineering and Building Technology.

Acknowledgments: This research was supported by a grant from a Strategic Research Project (Developing technology for water scarcity risk assessment and securing water resources of small and medium sized catchments against abnormal climate and extreme drought) funded by the Korea Institute of Civil Engineering and Building Technology. Authors appreciate the editors of the journal and the reviewers for their valuable comments and suggestions for improvements.

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Doblas-Reyes, F.J.; Garcia-Serrano, J.; Lienert, F.; Biescas, A.P.; Rodrigues, L.R.L. Seasonal climate predictability and forecasting: status and prospects. WIREs Clim. Chan. 2013, 4, 245–268. doi: 10.1002/wcc.217.
2. Navid, M.A.; Niloy, N.H. Multiple regressions for predicting rainfall for Bangladesh. Communications 2018, 6, 1–4.
3. Schepen, A.; Wang, Q.J.; Robertson, D.E. Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. J. Geophys. Res. 2012, 117, D20107, doi:10.1029/2012JD018011.
4. Rodwell, M.J.; Palmer, T.N. Using numerical weather prediction to assess climate models. Q. J. Roy. Meteor. Soc. 2014, 133, 129–146.
5. Woldemeskel, F.M.; Sharma, A.; Sivakumar, B.; Mehrotra, R. A framework to quantify GCM uncertainties for use in impact assessment studies. J. Hydrol. 2014, 519, 1453–1465.
6. Charles, A.N.; Duell, R.E.; Wang, X.; Watkins, A.B. Seasonal forecasting for Australia using a dynamic model: Improvements in forecast skill over the operational statistical model. Aust. Meteorol. Ocean. 2015, 65, 356–375.
7. Azimi, M.; Golpaygani, F.; Tajrishy, M.; Abrishamchi, A. Seasonal prediction of Karoon streamflow using large-scale climate indices. In proceedings of World Environmental and Water Resources Congress, Palm Springs, California, CA, USA, 22–26 May, 2011, pp. 1184–1193
8. Chen, C.-J.; Lee, T.-Y. On the relationship between teleconnections and Taiwan’s streamflow: Evidence of climate regime shift and implications for seasonal forecasting. Hydrol. Earth Syst. Sc. 2016, https://doi.org/10.5194/hess-2016–216.
9. Cho, J.; Jung, I.W.; Kim, C.G.; Kim, T.G. One-month lead dam inflow forecast using climate indices based on tele-connection. J. Korea Water Resour. Assoc. 2016, 49, 361–372. (in Korean)
10. Fleming, S.W.; Dahlke, H.E. Parabolic northern-hemisphere river flow teleconnections to El Niño-Southern Oscillation and the Arctic Oscillation. Environ. Res. Lett. 2014, 9, 104007.
11. Lee, J.H.; Julien, P.Y.; Thornton, C.; Lee. C.H. Large-scale climate teleconnections with South Korean streamflow variability. Hydrolog. Sci. J. 2020, 65, 57–70.
12. Meидani, Е.; Araghinejad, S. Long-lead streamflow forecasting in the Southwest of Iran by sea surface temperature on the Mediterranean Sea. J. Hydrol. Eng. 2014, 19, 05014005.
13. Baltaci, H.; Arslan, H.; Akkoyunlu, B.O.; Gomes, H.B. Long-term variability and trends of extended winter snowfall in Turkey and the role of teleconnection patterns. Meteorol. Appl. 2020, 27, e1891; doi:10.1002/met.1891.
14. Kim, Y.; Kang, N.; Kim, S.; Kim, H. Evaluation for snowfall depth forecasting using neural network and multiple regression models. J. Korean Soc. Hazard Mitig. 2013, 13, 269–280. (in Korean)
15. Martin, J.-P.; Germain, D. Large-scale teleconnection patterns and synoptic climatology of major snow-avalanche winters in the Presidential Range (New Hampshire, USA). Int. J. Climatol. 2017, 37, doi:10.1002/joc.4985.
16. Gonsamo, A.; Chen, J.M.; Lombardozzi, D. Global vegetation productivity response to climatic oscillations during the satellite era. Glob. Chan. Biol. 2016, 22, 3414–3426.
17. Heino, M.; Guillaume, J.H.A.; Mülller, C.; Iizumi, T.; Kummu, M. A multi-model analysis of teleconnected crop yield variability in a range of cropping systems. Earth Syst. Dynam. 2020, 11, 113–128.
18. Martinez, C.J.; Baigorria, G.A.; Jones, J.W. Use of climate indices to predict corn yields in southeast USA. *Int. J. Climatol.* 2009, 29, 1680–1691.

19. Quiring, S.M.; Blair, D. The utility of global teleconnection indices for long-range crop forecasting on the Canadian Prairies. In *Prairie Perspectives: Geographical Essays*, Eds.; Romanowski, J.I., University of Manitoba: Winnipeg, Manitoba, Canada, 2000, Volume 3, pp. 23–61.

20. Kim, K.-H.; Cho, J.; Lee, Y.-H. Forecasting brown planthopper infestation in Korea using statistical models based on climatic tele-connections. *Korean J. Appl. Entomol.* 2016, 55, 139–148. (in Korean)

21. Jo, S.; Ahn, J.-B. Statistical forecast of early spring precipitation over South Korea using multiple linear regression. *Clim. Res.* 2017, 12, 53–71. (in Korean)

22. Kim, H. Seasonal prediction of regional precipitation in Korea using the lag-correlated large-scale predictors. *Atmosphere* 2002, 12, 133–136. (in Korean)

23. Kim, J.-Y.; Seo, K.-H.; Son, J.-H.; Ha, K.-J. Development of statistical prediction models for Changma precipitation: An ensemble approach. *Asia-Pac. J. Atmos. Sci.* 2017, 53, 207–216.

24. Kim, M.-K.; Kim, Y.-H. Seasonal prediction of monthly precipitation in China using large-scale climate indices. *Adv. Atmos. Sci.* 2010, 27, 47–59.

25. Kim, M.-K.; Kim, Y.-H.; Lee, W.-S. Seasonal prediction of Korean regional climate from preceding large-scale climate indices. *Int. J. Climatol.* 2007, 27, 925–934.

26. Kwon, M. Diagnosis of Northeast Asian summer precipitation suing the Western North Pacific subtropical high index. *J. Korean Earth Sci. Soc.* 2013, 34, 102–106. (in Korean)

27. Lee, S.-E.; Seo, K.-H. The development of a statistical forecast model for Changma. *Weather and Forecast.* 2013, 28, 1304–1321.

28. Kwon, M.; Lee, K.-J. A prediction of Northeast Asian summer precipitation using the NCEP climate forecast system and canonical correlation analysis. *J. Korean Earth Sci. Soc.* 2014, 35, 88–94. (in Korean)

29. Lee, K.-J.; Kwon, M. A prediction of precipitation over East Asia for June using simultaneous and lagged teleconnection. *Atmosphere* 2016, 26, 711–716. (in Korean)

30. Kim, T. Monthly precipitation forecasting with large-scale climate indicators using ensemble-based neural network approaches. Ph.D. Thesis, Yonsei University, Korea, June 2018.

31. Lee, J.; Kim, C.-G.; Lee, J.E.; Kim, N.W.; Kim, H. Application of artificial neural networks to rainfall forecasting in the Geum River basin, Korea. *Water* 2018, 10, 1448, doi:10.3390/w10101448.

32. Lee, S.; Kim, G.; Yoon, S.; An, H. Improvement of precipitation forecasting skill of ECMWF data using multi-layer perceptron technique. *J. Korea Water Resour. Assoc.* 2019, 52, 475–482. (in Korean)

33. Kumar, D.N.; Reddy, M.J.; Maity, R. Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. *J. Intell. Syst.* 2007, 16, 307–322.

34. Mekanik, F.; Imteaz, M.A.; Gato-Trinidad, S.; Elmahdi, A. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *J. Hydrol.* 2007, 334, 11–21.

35. Kim, J.Y.; Park, H.J. Impact of northern hemisphere teleconnection patterns on precipitation in Korea. *Mag. Korea Water Resour. Assoc.* 2010, 43, 57–61. (in Korean)

36. Kim, Y.-H.; Kim, M.-K.; Lee, W.-S. An investigation of large-scale climate indices with the influence on temperature and precipitation variation in Korea. *Atmosphere* 2008, 18, 83–95. (in Korean)

37. Tan, X.; Shao, D. Precipitation trends and teleconnections identified using quantile regressions over Xinjiang, China. *Int. J. Climatol.* 2017, 37, 1510–1525.

38. Li, J.; Zeng, Q. A unified monsoon index. *Geophys. Res. Lett.* 2002, 29, 1274, 10.1029/2001GL013874.

39. Akoglu, H. User’s guide to correlation coefficients. *Turk. J. Emerg. Med.* 2018, 18, 91–93.
40. Evans, J.D. *Straightforward Statistics for the Behavioral Sciences*; Brooks/Cole Publishing; Pacific Grove: California, CA, USA, 1996.

41. Altman, D.G. *Practical statistics for medical research*; Chapman and Hall: London, United Kingdom, 1991.

42. Moriasi, D.N.; Arnold, J.G.; Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 2007, 50, 885–900.

43. He, X.; Guan, H. Multiresolution analysis of precipitation teleconnections with large-scale climate signals: A case study in South Australia. *Wat. Resour. Res.* 2013, 49, 6995–7008.

44. Hair, J.F.; Anderson, R.E.; Tatham, R.L.; Black, W.C. *Multivariate Data Analysis*, 3rd ed.; Macmillan: New York, NY, USA, 1995.

45. Kennedy, P. *A Guide to Econometrics*; Blackwell: Oxford, United Kingdom, 1992.

46. Marquardt, D.W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 1970, 12, 591–612.

47. Neter, J.; Wasserman, W.; Kutner, M.H. *Applied Linear Regression Models*; Richard D. Irwin: Homewood, IL, USA, 1989.