

*Improvement of daily precipitation estimations using PRISM with inverse-distance weighting*

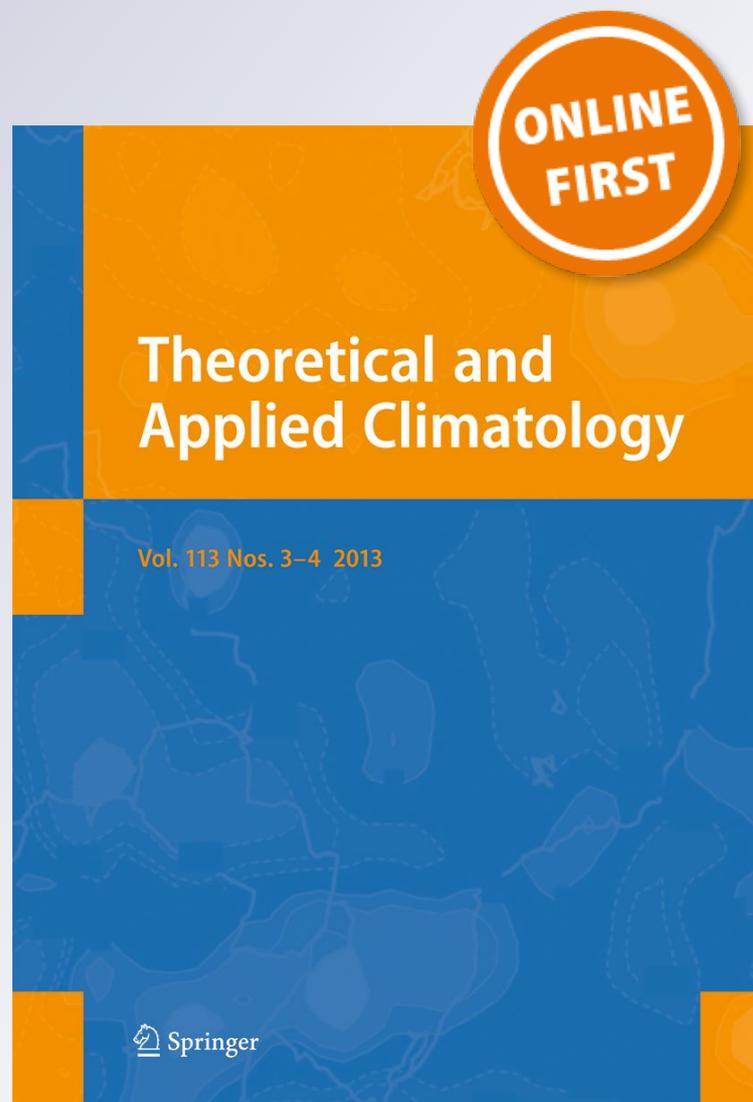
**Ha-Gyu Jeong, Joong-Bae Ahn, Joonlee Lee, Kyo-Moon Shim & Myung-Pyo Jung**

**Theoretical and Applied Climatology**

ISSN 0177-798X

Theor Appl Climatol

DOI 10.1007/s00704-019-03012-6



**Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.**



# Improvement of daily precipitation estimations using PRISM with inverse-distance weighting

Ha-Gyu Jeong<sup>1</sup> · Joong-Bae Ahn<sup>1</sup> · Joonlee Lee<sup>2</sup> · Kyo-Moon Shim<sup>3</sup> · Myung-Pyo Jung<sup>3</sup>

Received: 5 September 2018 / Accepted: 20 September 2019  
© The Author(s) 2019

## Abstract

Improved daily precipitation estimations were attempted using the parameter-elevation regressions on a parameter-elevation regression on independent slopes model (PRISM) with inverse-distance weighting (IDW) and a precipitation-masking algorithm for precipitation areas. The PRISM (PRISM\_ORG) suffers two overestimation problems when the daily precipitation is estimated: overestimation of the precipitation intensity in mountainous regions and overestimation of the local precipitation areas. In order to solve the problem of overestimating the precipitation intensity, we used the IDW technique that employs the same input stations as those used in the PRISM regression (PRISM\_IDW). A precipitation-masking algorithm that selectively masks the precipitation estimation grid points was additionally applied to the PRISM\_IDW results (PRISM\_MSK). For 6 months from March to August 2012, daily precipitation data were produced in a horizontal resolution of 1 km based on the above two experiments and PRISM\_ORG. Afterwards, each experiment was evaluated for improvements. The monthly root mean squared errors (RMSEs) of PRISM\_IDW and PRISM\_MSK were reduced by 0.83 mm/day and 0.86 mm/day, respectively, compared to PRISM\_ORG.

**Keywords** Statistical downscaling · Daily precipitation · PRISM · IDW · High resolution

## 1 Introduction

As computation performance improves, the production of grid-type meteorological data of higher resolution has become possible. In recent years, high-resolution meteorological data has been widely utilized not only in meteorological areas but also in the fields of agriculture, hydrology, engineering, and the economy (Daly et al. 2002; Daly et al. 2003). Despite this high utilization, it remains challenging to derive detailed meteorological data from observational data due to spatial and temporal heterogeneities. In particular, although around 70% of the land in South Korea consists of mountainous terrain, most observation stations are concentrated in urban regions

due to economic and technical limitations (Hong et al. 2007; Im and Ahn 2011).

In order to overcome the limitations of observational data and allow for easier utilization, several studies have investigated how to produce detailed grid-type data and information (Ahn et al. 2012; Gerelchuluun and Ahn 2014). The methods that can produce detailed grid-type meteorological information can be divided into dynamical and statistical methods. A dynamical method that uses a regional climate model has the advantage of producing physically and dynamically balanced data. However, the method has some limitations such as the fact that a regional climate model requires considerable integration time and significant storage to produce high-resolution data (Chen et al. 2012), and that the model outputs may include systematic errors (Ahn et al. 2012; Jo and Ahn 2014; Ahn and Lee 2016; Lee and Ahn 2018). Due to these drawbacks, a statistical method is also often used as an efficient method that produces meteorological and climate data at a high resolution (e.g., Daly 2006; Brunetti et al. 2013).

For statistical methods that produce high-resolution grid-point data from observations, some studies have been conducted on objective analysis (e.g., Barnes 1964; Cressman 1959). A statistical objective analysis assumes that the

✉ Joong-Bae Ahn  
jbahn@pusan.ac.kr

<sup>1</sup> Division of Earth Environmental System, Pusan National University, Jangjeon 2-dong, Geumjeong-gu, Busan 609-735, South Korea

<sup>2</sup> School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea

<sup>3</sup> Division of Climate Change and Agroecology, National Academy of Agricultural Science, Wanju County, South Korea

homogeneity of climate variables decreases over distance and that the estimated value at each grid point is inversely proportional to the square of the distance of the adjacent observational data. A Kriging (Krige 1951) technique that considers not only the distances but also the other variables that have a linear relationship with the estimation variables is also frequently used to interpolate meteorological data. Particularly, surface data are significantly affected by topographical features (Johnson et al. 2000; Perry and Konrad 2006). A hypsometric technique (Ahrens 2003) and a parameter-elevation regression on independent slopes model (PRISM) (Daly et al. 1994, 2002; Daly 2006) are two of the typical objective analysis methods that consider geographical features. In particular, since PRISM has a relatively low dependence on the adjacent meteorological data compared to other methods, it may be usefully employed to estimate high-resolution grid data from observational data that are distributed spatially inhomogeneously and coarsely (Ahn et al. 2014).

PRISM is a high-resolution model for grid-type climate data estimations developed by Oregon State University. The model not only applies distance weighting but also additional geomorphological factors such as the topographic altitude, topographic facet, and coastal proximity. Hong et al. (2007) proposed Korean PRISM (K-PRISM) in which regression variables and weights suitable for climate data estimations in South Korea were employed based on improvements to PRISM. They estimated the temperature data at a 5-km  $\times$  5-km resolution in South Korea using K-PRISM. Shin et al. (2008) also conducted a study on a grid-type estimation of precipitation in South Korea with the same resolution as Hong et al. (2007) using K-PRISM. Kim et al. (2012) estimated the daily precipitation with a 1-km  $\times$  1-km resolution in South Korea using MK-PRISM (modified K-PRISM) and improved the precipitation area estimations of K-PRISM.

Although these studies improved high-resolution grid-type climate data estimations, uncertainties remain. In particular, it is more difficult to estimate the spatial distribution and the intensity of precipitation compared to other variables such as temperature, due to the regional seasonal and topographic characteristics. In our study, experiments are conducted to improve the daily precipitation estimations produced from the PRISM (PRISM\_ORG) by correcting the overestimation of precipitation intensity and areas, and the results were verified.

## 2 Data and methods

### 2.1 Data

The data used in this study consisted of observational and geographic information system (GIS) data. For the observational data, the automated synoptic observation system

(ASOS) and the automatic weather station (AWS) were employed. For the GIS data, the ASTER Global Digital Elevation Model version 2 (ASTGTM2) (Tachikawa et al. 2011) was used.

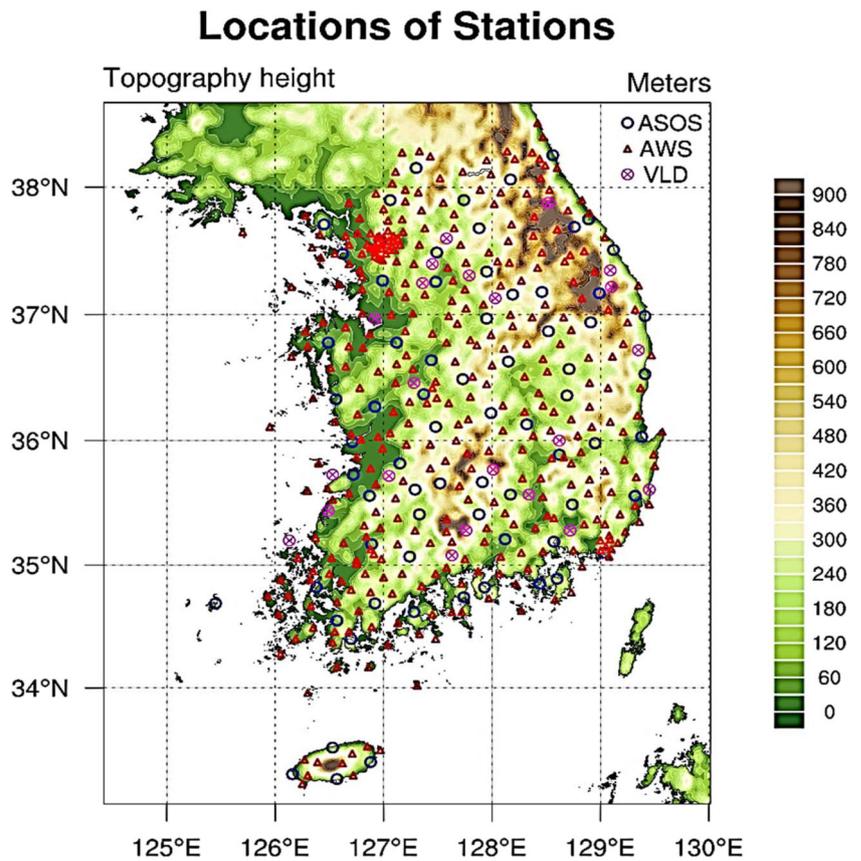
Observations provided by the Korean Meteorological Administration were used as the input and verification data in PRISM. Although 89 ASOS stations and 599 AWS sites collected observation data, the actual PRISM input data were selectively derived from 72 of the ASOS stations and 352 of the AWS sites (Fig. 1). The selection criteria for the stations were as follows. Among the ASOS stations, 72 stations operating continuously since 1973 were selected by considering the homogeneity of the input data. From the AWS sites, 352 were chosen by excluding sites where significant temporal gaps occurred in the measurement data or where the observation station locations were changed. The location of the points used for verification is also an important factor to consider. In this study, the 22 stations in the AWS used for validation were selected where PRISM input data were not used (Fig. 1). The selection criteria for the validation stations included the requirement that the stations must have had 10 or fewer days without missing measurements and that the minimum distance to the other stations used for the input data was more than 10 km.

For the digital elevation map, ASTER Global Digital Elevation Map Version2 (ASTGTM2) data developed and produced jointly by the USA and Japan as the deliverable of the Global Earth Observation System project was used. A 30-m resolution digital elevation map was converted to a 1-km resolution map. The topographic facets and coastal proximities calculated from this map were used as the PRISM input data. Sixteen directions (N, NNE, NE, NEE, E, SEE, SE, SSE, S, SSW, SW, SWW, W, NWW, NW, and NNW) and a flat for the topographic facets are utilized as used in Ahn et al. (2014), so topographic facets may have 17 directions. Coastal proximity refers to the degree that the same oceanic climate affects the data. The coastal proximity was classified into six categories in this study according to the shortest distances from the coast (less than 10 km, more than 10 km, less than 20 km, and over 20 km) and topographic altitude (less than 200 m and over 200 m).

### 2.2 PRISM\_ORG: original PRISM

The basic precipitation estimation method used in this study was from the PRISM method developed by Daly et al. (1994, 2002) and Daly (2006). PRISM uses differences between the geographical information obtained at observation stations within a radius of influence in order to estimate the precipitation amount at each grid point, thereby assigning a weight to each observation station and estimating a value at each grid point by setting an individual linear regression equation for each grid point.

**Fig. 1** Topography and spatial distribution of the ASOS (blue circles) and AWS (red triangles) stations used for the PRISM input data along with the stations used for validation (pink circles with crosses)



	Mean distance	Maximum distance	Minimum distance
ASOS	33.0km	144.6km	13.7km
AWS	12.5km	53.0km	1.8km

Here, the linear regression equation as a function of altitude is as follows:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \tag{1}$$

where  $\hat{Y}$  denotes the estimated precipitation value at the grid point, and  $\hat{\beta}_0$  and  $\hat{\beta}_1$  refer to the intercept and the slope in the regression equation, respectively.  $X$  refers to the altitude at the target grid point. Here, the weight used during the setup of the  $\hat{\beta}_0$  and  $\hat{\beta}_1$  regression coefficients is determined by the following differences in distance, altitude, topographic facet, and coastal proximity:

$$W = f(W_d, W_z, W_f, W_p) \tag{2}$$

where  $W$  is the total weight, and  $W_d$  and  $W_z$  are the weights due to differences in distance and altitude, while  $W_f$  and  $W_p$  are the weights due to differences in topographic facet and coastal proximity. Here, the selection of the total weight is set as that

used in K-PRISM from a study by Hong et al. (2007) and the equation is as follows:

$$W = [F_d W(d) + F_z W(z)] W(f) W(p) \tag{3}$$

Here,  $F_d$  and  $F_z$  are the weights of the distance and altitude that were set to  $F_d = 0.8$  and  $F_z = 0.2$ , as in Hong et al. (2007). The other weights correspond to the weights of the difference in distance ( $W(d)$ ), in altitude ( $W(z)$ ), in topographic facet ( $W(f)$ ), and in coastal proximity ( $W(p)$ ) (Hong et al. 2007).

The observation stations used to estimate the precipitation at each grid point were those within 30 km of the grid point. The determination of the radius of influence is particularly important for estimating local- and regional-scale precipitation. Increasing the influence radius around the grid points where we want to estimate precipitation has the advantage of increasing the number of observation points included within the radius, but also the disadvantage of weakened local climatological characteristics of precipitation influenced by topography. The 30-km influence

of radius utilized in this study is the one used by K-PRISM, which is optimized in consideration of South Korea's geography and climate, and by many prior studies (e.g., Hong et al. 2007; Shin et al. 2008; Kim et al., 2011). However, with fewer than five observation stations within the radius of influence, the radius of influence is increased by 5 km up to 50 km (Hong et al. 2007; Ahn et al. 2014). The individual weights are calculated at the observation stations within the radius of influence. Thus, the regression coefficient used in Eq. (2) has the following weighted regression coefficients:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n W_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n W_i (x_i - \bar{x})^2} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (4)$$

$$\bar{x} = \frac{\sum_{i=1}^n W_i x_i}{\sum_{i=1}^n W_i}, \quad \bar{y} = \frac{\sum_{i=1}^n W_i y_i}{\sum_{i=1}^n W_i} \quad (5)$$

where  $W_i$  refers to the weight of the  $i$ th observation station among  $n$  observation stations included within the radius of influence. The more similar number of observation stations at which the geographical information is similar to that of the target grid, the larger the weight assigned to the weighted regression coefficient, and hence the greater the influence on the estimation of precipitation at the target grid point.

Among the regression coefficients used in the PRISM grid point estimation, the slope of ( $\hat{\beta}_1$ ) may be abnormally large or small if the altitudes and distances of the observation stations from some of the target grid points within the radius of influence are not variably distributed and this may result in unrealistic estimations (Shin et al. 2008). Ahn et al. (2014) calculated the daily temperature lapse rate and standard deviation

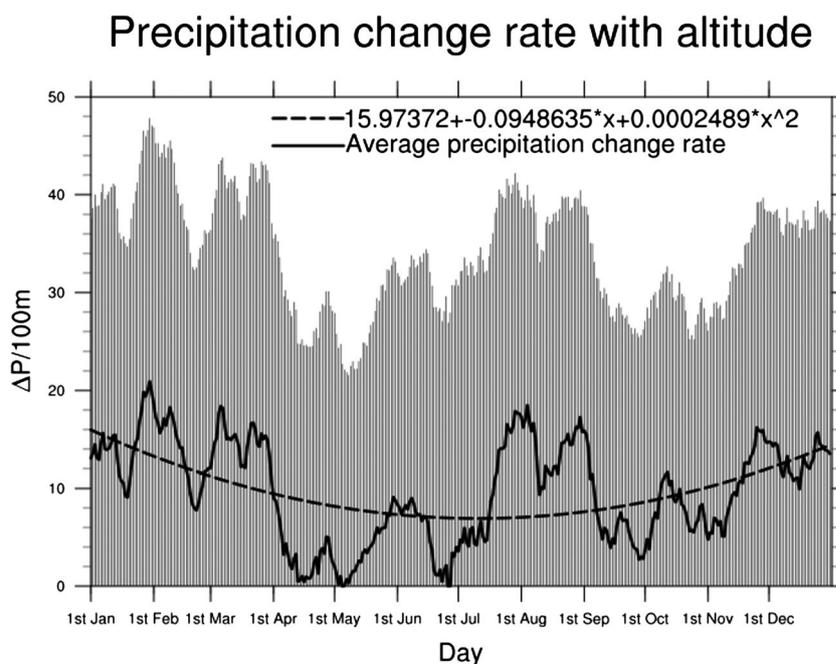
averaged over 40 years from 1973 to 2012, while also limiting the upper and lower limitations of the slope of the regression equation by using the result obtained by calculating a 10-day moving average to remove small variabilities. They also limited the maximum and minimum values of the slope using the reduction rate of the daily mean precipitation according to the altitudes averaged over 40 years by applying the method used in Ahn et al. (2014). We set the maximum and minimum values of the slope to  $+0.5 \sigma$  and  $-0.5 \sigma$ , respectively (Fig. 2). Here,  $\sigma$  refers to the standard deviation of the daily mean precipitation lapse rate over the period of 40 years. The Y-axis of Fig. 2 is the lapse rate in total daily mean precipitation over a total of 40 years from 1973 to 2012, in mm/day. The magnitude of the lapse rate is similar to that shown in various prior studies (e.g., Hong et al. 2007; Sin et al., 2008; Kim et al., 2011). In our PRISM, we used the above input data and weights to estimate the daily precipitation in South Korea from March 1 to August 31, 2012.

## 2.3 Experimental design

### 2.3.1 PRISM\_IDW: estimation of precipitation intensity

PRISM produces more accurate estimates because it considers the topographic characteristics more variously than other objective analysis methods do. Precipitation is in general affected by topography; in other words, the distribution of precipitation depends on the slopes of the mountains. This mountain effect on precipitation has been reported in previous studies (e.g., Henry 1919; Smith 1979). However, in some mountainous regions, PRISM results have overestimated precipitation.

**Fig. 2** Daily mean of the precipitation change rate with altitude ( $\Delta P/\Delta z$ ) over 40 years from 1973 to 2012 in South Korea (bold line) and its 0.5 standard deviations (bar) (a 10-day moving average was applied) (unit: mm/day)



To solve this problem, we used IDW (Shepard 1968), an objective analysis method that does not consider the topographical effect. The IDW method determines precipitation at the interpolating point by assigning larger weights to observation stations closer to the target grid. This method has been widely used in various research areas and practices because it is one of the simplest among the available objective analysis methods (Daly et al. 2002).

The IDW estimation equation defined by Shepard (1968) is as follows:

$$\hat{x} = \left( \frac{\sum_{i=1}^n W_i x_i}{\sum_{i=1}^n W_i} \right) \quad \text{where} \quad W_i = \frac{1}{d^2} \quad (6)$$

where  $\hat{x}$  refers to an estimate at the target grid point,  $W_i$  is the weight of the  $i$ th station, and  $x_i$  is the observed value at the  $i$ th station that is used for the target grid point estimation. The weight at the  $i$ th station is inversely proportional to the square of the distance ( $d$ ) between the target grid point and the  $i$ th observation station. The sum of the weights as shown in Eq. (6) is 1 for observation points within the radius.

In this study, the experiments were designed to detect changes in the precipitation intensity by averaging the precipitation calculated from the IDW and the precipitation

estimated from PRISM\_ORG. The same stations in the PRISM radius of influence were used in the IDW to maintain the objectivity of the estimates. A schematic diagram of the experiment design is shown in Fig. 3. The above experiment is named PRISM\_IDW, which is performed over the same period.

### 2.3.2 PRISM\_MSK: estimation of precipitation area

In contrast with other meteorological variables, daily precipitation can exhibit an uneven spatial distribution when the precipitation occurs locally. If this local precipitation is estimated using PRISM, unrealistic precipitation overestimations within the radius of influence can occur. This problem is not evident in the estimations of monthly precipitation distributed over sufficiently large areas, but it is revealed when the daily precipitation is estimated at the observation stations that are distributed locally.

Thus, a precipitation-masking algorithm (Fig. 4) was used in PRISM\_MSK to correct for the estimation of the precipitation areas. In this study, among the grid estimates obtained from PRISM\_IDW, a precipitation-masking algorithm was applied to the grid points estimated to have precipitation, thereby reducing the overestimated precipitation areas (Fig. 3).

If the distance between the grid point estimated as the precipitation area and the closest precipitation observation point is less than 12.5 km, and if no precipitation occurs at three or more observation stations out of the closest five stations, then the precipitation-masking algorithm masks the precipitation estimated for the corresponding grid points. If the distance is 12.5 km or greater, and if no precipitation occurs at two or more observation stations out of the five closest stations, the algorithm masks the corresponding grid points. Here, 12.5 km, used as the threshold distance for estimating precipitation, is the distance among the average observation points of ASOS + AWS (Fig. 1). The presence of precipitation at a grid point is determined according to its distance from the station where the precipitation is observed as well as whether precipitation occurs or not at the surrounding nearby observation stations, thereby improving the overestimation problem for a precipitation area for a more realistic estimation. PRISM\_MSK also conducted precipitation estimations for the same period, i.e., from March 1 to August 31, 2012.

## 3 Results and validations

### 3.1 Results

PRISM\_ORG and the two improved PRISMs were examined through the experiments, thereby producing three results. These experiments provided the grid-type daily precipitation at a 1-km resolution for 184 days from March 1 to August 31,

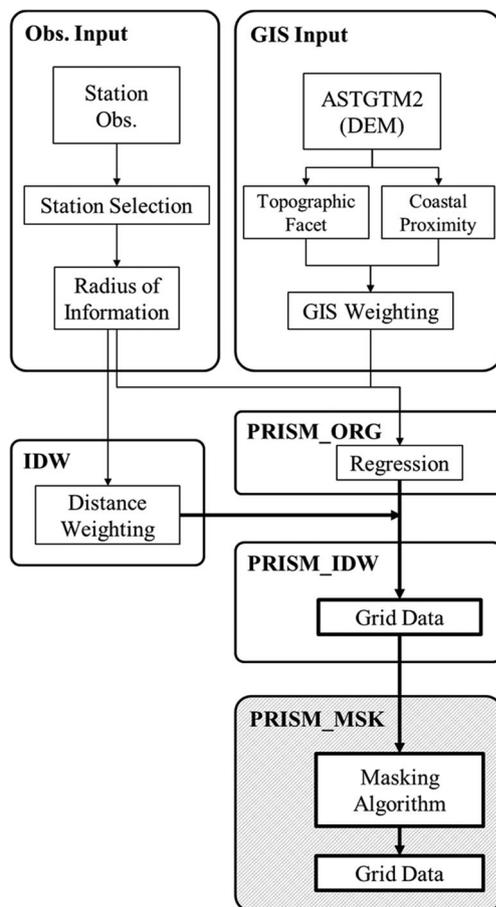
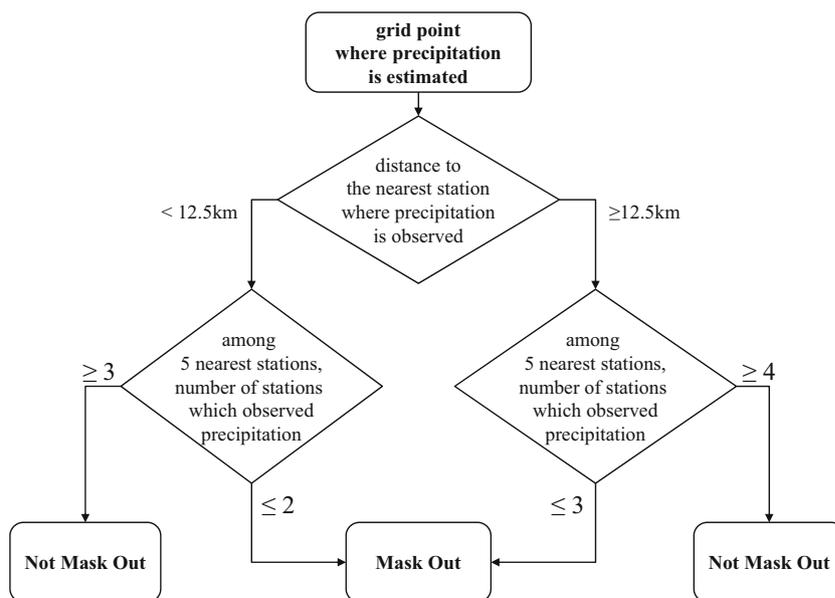


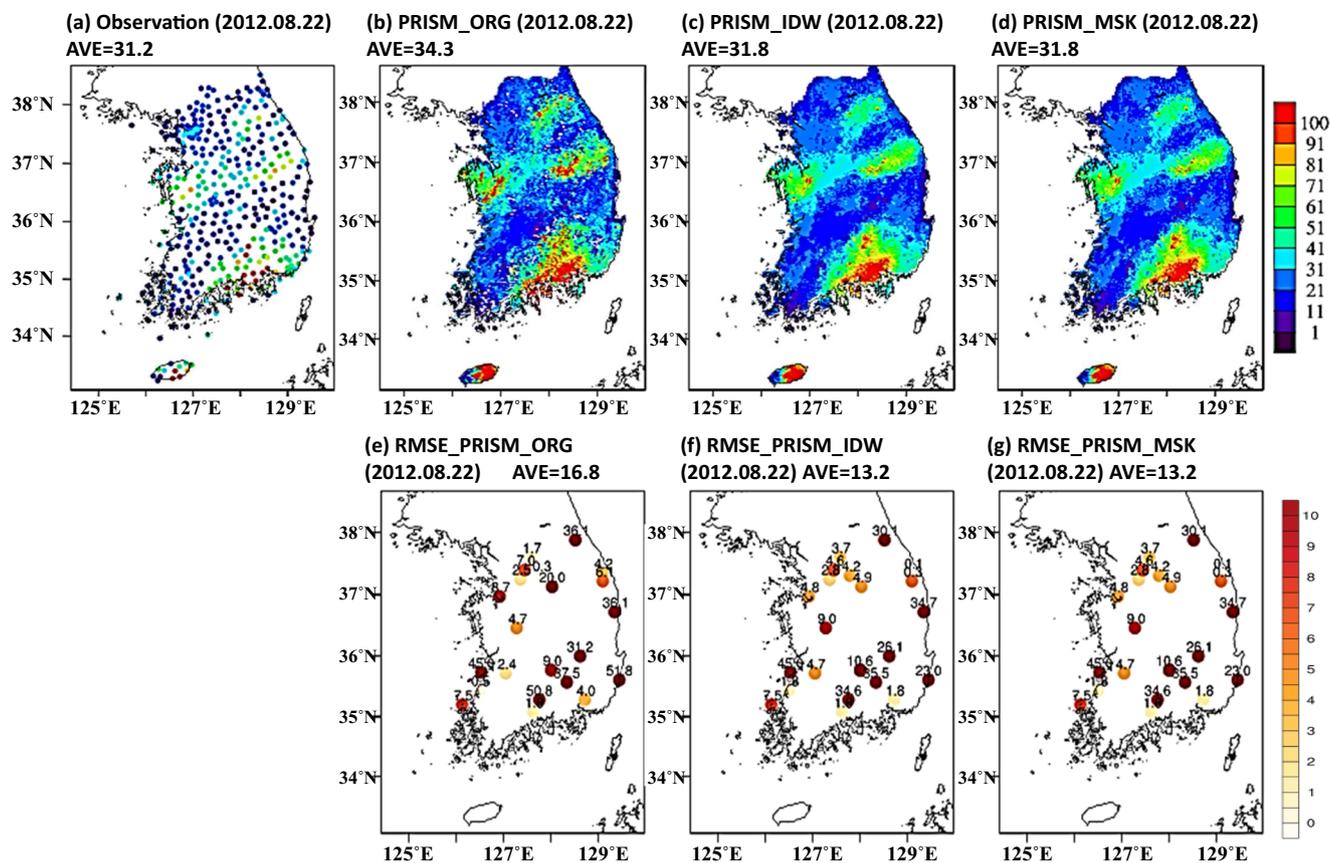
Fig. 3 Process diagram of PRISM and the experiments

**Fig. 4** Precipitation masking algorithm used in PRISM\_MSK



2012. During this period, only less than 6% (12 days) of total cases were not improved by the application of the new method in terms of RMSE (data not shown). The precipitation in the

Korean Peninsula can be clearly divided into two types according to the seasons. That is, the precipitation in summer and that in the other three seasons are mainly caused by the



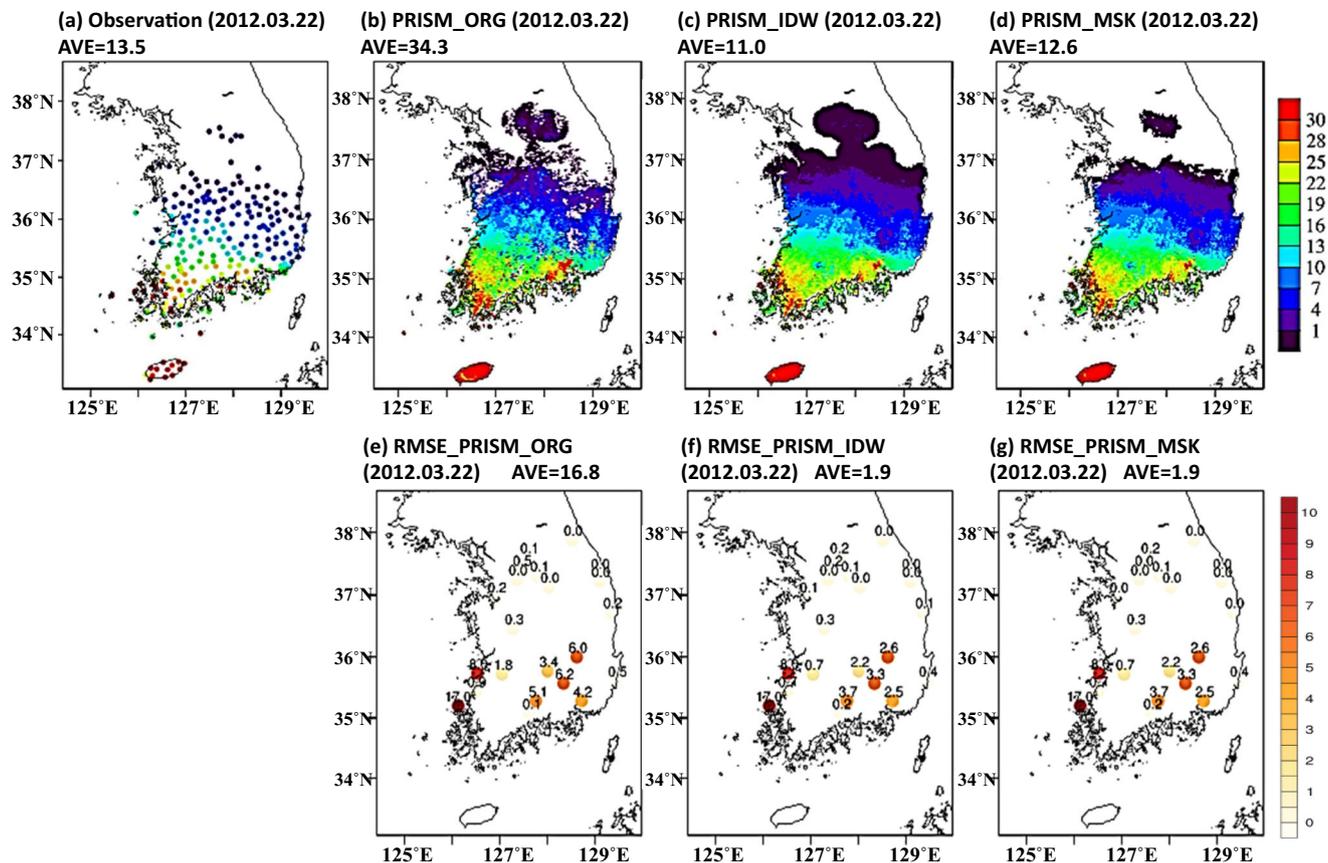
**Fig. 5** Spatial distribution of the daily precipitation on August 22, 2012 for the **a** observed precipitation, **b** precipitation estimated by PRISM\_ORG, **c** precipitation estimated in PRISM\_IDW, and the **d** precipitation estimated in PRISM\_MSK. Root mean-squared errors (RMSEs) are

shown as markers at each validation station for the results of **e** PRISM\_ORG, **f** PRISM\_IDW, and **g** PRISM\_MSK. Spatially averaged values are shown in the top right of each panel (unit: mm/day)

local static instability and the baroclinic instability associated with the synoptic system, respectively. Thus, among the 184 days, for easy analysis of the results, we selected the well-estimated cases for spring (March 22) and summer (August 22) for analyses and discussions. Figure 5 shows three kinds of result for the precipitation estimations (b–d) on August 22, 2012, with the observation data distribution (a) and root mean squared errors (RMSEs) from the validation stations for the three results (e–g). All of the results from PRISM\_ORG, PRISM\_IDW, and PRISM\_MSK showed a similar distribution as the observations. The PRISM\_ORG results (b) exhibited the abovementioned overestimation problem for the precipitation. Figure 5 e shows that mountainous areas such as the Gangwon Province and the Jiri Mountain had high RMSEs of 36.1 mm/day and 50.8 mm/day, respectively. The mountain effect, which is overestimated, decreased in PRISM\_IDW, as shown in Fig. 5c, and the RMSEs in the mountainous areas were reduced to be 30.1 mm/day and 34.6 mm/day, respectively (Fig. 5f). For PRISM\_MSK, shown in Fig. 5d, the result showed no significant differences compared to PRISM\_IDW. The reason is, when the

precipitation covers the entire area, the precipitation-masking algorithm does not change the precipitation area.

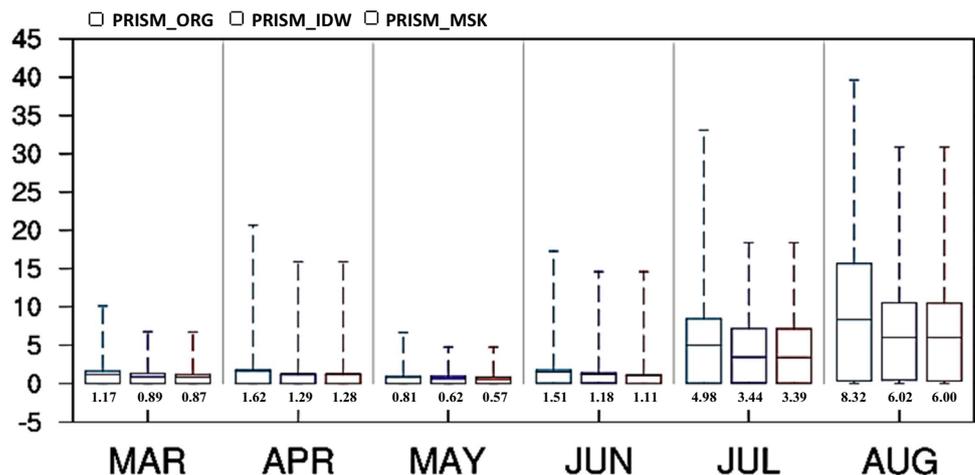
Figure 6 shows the precipitation estimation results from March 22, 2012. Figure 6 a exhibits the observation precipitation distributions, in which overall, the rainfall occurred in the southern region, whereas the local rainfall occurred in the central inland region. PRISM\_ORG, as shown in Fig. 6b, shows the overestimated precipitation areas as the shape of the radius of influence for the local rainfall in the central inland region. PRISM\_IDW, as shown in Fig. 6c, estimates a wider area of precipitation than does PRISM\_ORG, shown in Fig. 6b. This is due to the IDW method that was used to adjust the precipitation intensity. PRISM\_MSK, as shown in Fig. 6d, shows that the overestimated precipitation areas revealed in the previous two results were reduced through the precipitation-masking algorithm. The results of the RMSEs (Fig. 6e–g) show that PRISM\_ORG and PRISM\_IDW, which incorrectly estimated precipitation around the central inland local precipitation area due to overestimation of the precipitation area, had RMSEs of 0.5 mm/day and 0.2 mm/day, respectively. On the other hand, in PRISM\_MSK, the incorrect



**Fig. 6** Spatial distribution of the daily precipitation on March 22, 2012 for the **a** observed precipitation, **b** precipitation estimated by PRISM\_ORG, **c** precipitation estimated in PRISM\_IDW, and the **d** precipitation estimated in PRISM\_MSK. Root mean squared errors (RMSEs) are

shown as markers at each validation station for the results of **e** PRISM\_ORG, **f** PRISM\_IDW, and **g** PRISM\_MSK. Spatially averaged values are shown in the top right of each panel (unit: mm/day)

**Fig. 7** Box plot of the RMSE of PRISM\_ORG, PRISM\_IDW, and PRISM\_MSK for each month. The number under each box denotes the mean of the RMSE for each experiment (unit: mm/day)



precipitation was accurately masked, thereby reducing the RMSE to 0 mm/day.

## 3.2 Validations

### 3.2.1 RMSE

Quantitative comparisons were conducted using a box plot to determine the RMSE distribution of the daily precipitation estimations for the above three results (Fig. 7). Each box refers to the distribution of the average RMSE of the daily precipitation that occurred within a month. The maximum value, the top 25%, the mean value, the bottom 25%, and the minimum value are all expressed in each box plot, which represents the analysis results from March to August 2012 for the three experiments.

The precipitation from March to June was lower than that from July to August, resulting in a low RMSE from March to June. Also, a relatively larger RMSE in July and August was observed due to higher precipitation. The mean values showed considerably improved RMSE results compared to PRISM\_ORG over the entire period with respect to the estimation results of the daily precipitation from PRISM\_IDW and PRISM\_MSK. PRISM\_IDW showed a lower monthly mean RMSE compared to PRISM\_ORG by 0.28 mm (March), 0.34 mm (April), 0.19 mm (May), 0.33 mm (June), 1.54 mm (July), and 2.30 mm (August) per day. PRISM\_MSK showed a lower monthly mean RMSE compared to PRISM\_ORG by 0.30 mm (March), 0.34 mm (April), 0.24 mm (May), 0.40 mm (June), 1.59 mm (July), and 2.32 mm (August) per day. The differences in the RMSE between PRISM\_IDW and PRISM\_MSK were not larger than those between the RMSE of PRISM\_IDW and PRISM\_ORG, but in most months, PRISM\_MSK exhibited slightly lower than or the same RMSE as PRISM\_IDW did. The differences were 0.02 mm (March), 0.00 mm (April),

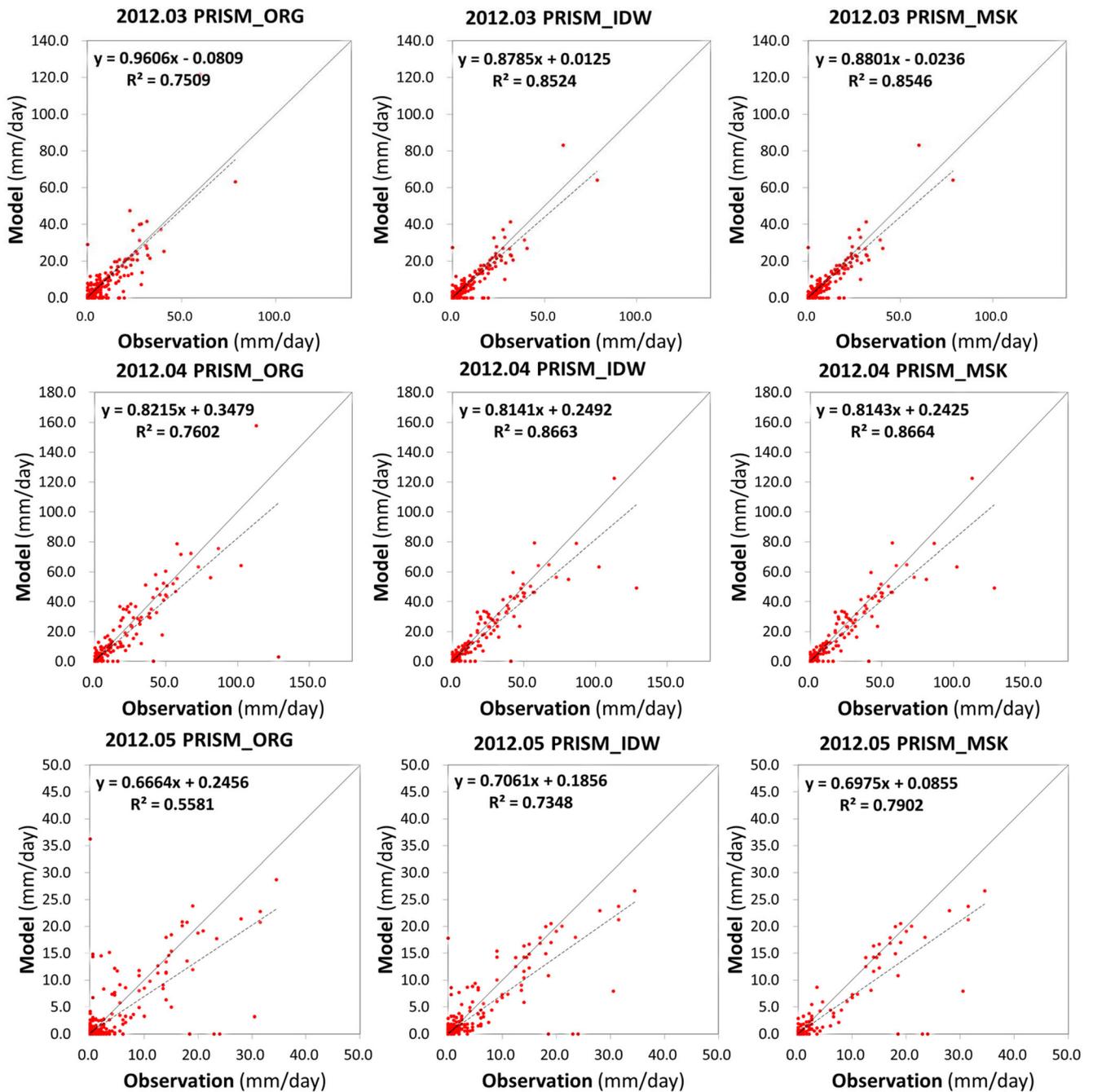
0.04 mm (May), 0.07 mm (June), 0.05 mm (July), and 0.01 mm (August) per day.

Next, the maximum RMSE results of PRISM\_IDW were decreased by 3.37 mm (March), 4.79 mm (April), 1.90 mm (May), 2.68 mm (June), 14.70 mm (July), and 8.77 mm (August) per day compared to PRISM\_ORG. The experimental results from PRISM\_MSK also decreased overall compared to the PRISM\_ORG results, and the decreases were the same as those from PRISM\_IDW, except for March (3.41 mm/day).

### 3.2.2 Scatter plot

Figures 8 and 9 show the scatter plots produced from a one-to-one match between the 22 daily observed validation precipitations (horizontal axis) and the estimated daily precipitation from the three experiments for grid points corresponding to each validation station (vertical axis). Each plot shows the results for each experimental month. The coefficient of determination ( $R^2$ ) that can be interpreted as the proportion of the variation of the estimation that is described or accounted for by the regression (Wilks 2011) is also suggested by the linear regression equation.

First, comparisons of the results between PRISM\_ORG and PRISM\_IDW showed that the estimations from PRISM\_IDW matched the observational distribution more closely overall than those from PRISM\_ORG did, and the  $R^2$  of PRISM\_IDW also increased by 0.11 on average from March to August compared to PRISM\_ORG. In particular, a relatively high proportion of values were skewed from the regression line toward the vertical axis under PRISM\_ORG due to the precipitation overestimations. PRISM\_IDW improved the overestimated precipitation intensities, such that the re-estimated values were closer to the observations. The  $R^2$  values increased in PRISM\_IDW by 0.1015 (March), 0.1061 (April), 0.1767 (May), 0.0885 (June), 0.0978 (July), and 0.1172 (August) compared to PRISM\_ORG.



**Fig. 8** Scatter plots of the precipitation observed at the stations and estimated by PRISM\_ORG (left), PRISM\_IDW (middle), and PRISM\_MSK (right) for March (top), April (middle), and May (bottom). The solid

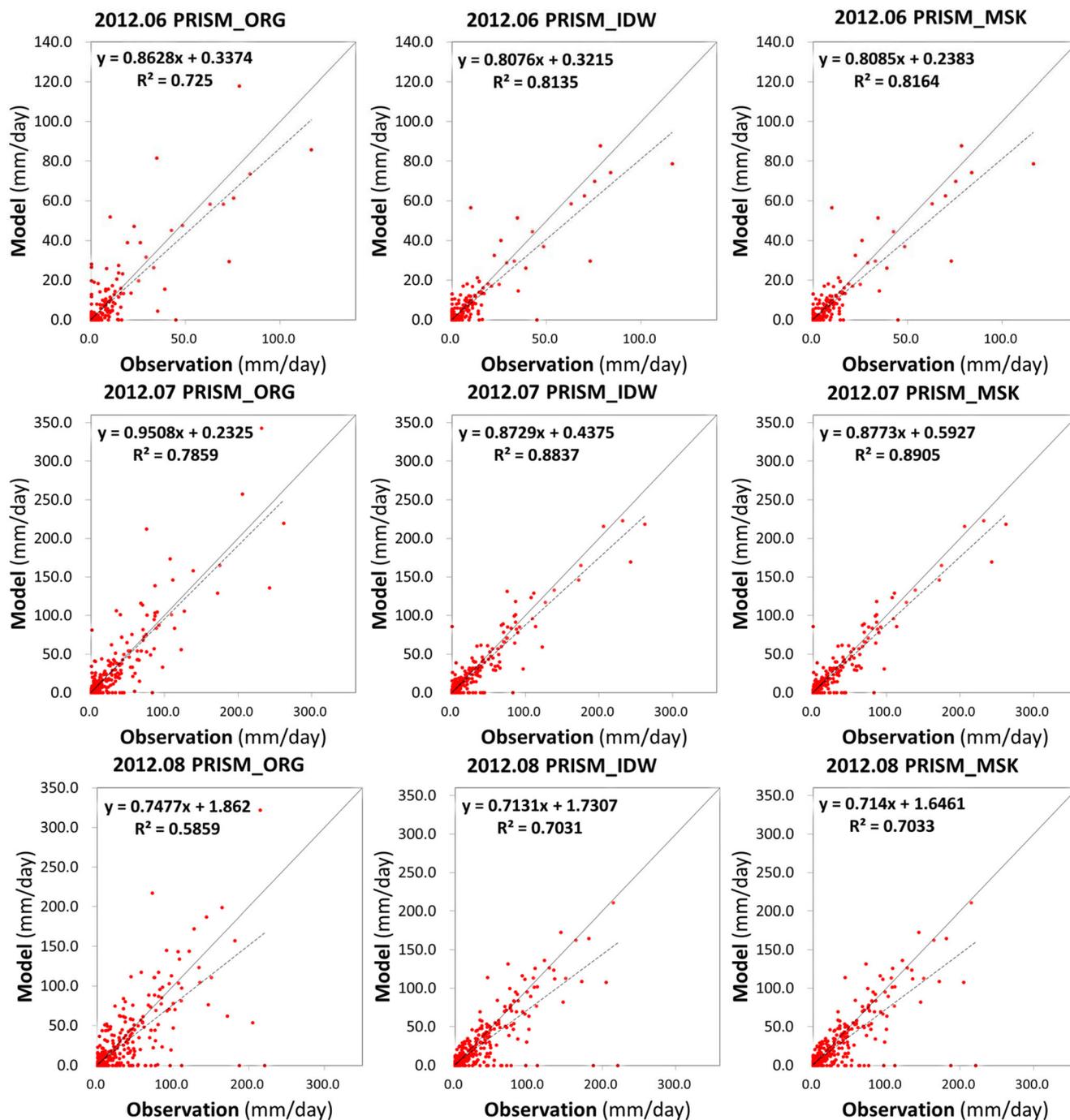
line is the 1:1 line and the dashed line is the linear regression line. The linear regression equation and the  $R^2$  values are shown at the top left in each panel

Next, in terms of the precipitation area, PRISM\_MSK presented fewer estimated errors than PRISM\_IDW did in the area where the local precipitation occurred, thereby slightly increasing the  $R^2$  value over the entire period. The  $R^2$  values in PRISM\_MSK increased by 0.0022 (March), 0.0001 (April), 0.0554 (May), 0.0029 (June), 0.0068 (July), and 0.0002 (August) compared to PRISM\_IDW. Although the numerical increases in the validation values were not

significant, the errors estimated at the precipitation grids where no precipitation occurred (points marked on the vertical axis) were considerably improved (origin point of the graph).

### 3.2.3 Contingency table

The  $2 \times 2$  contingency table between the observed precipitation and the precipitation estimated from PRISM\_ORG,



**Fig. 9** Scatter plots of the precipitation observed at the stations and estimated by PRISM\_ORG (left), PRISM\_IDW (middle), and PRISM\_MSK (right) for June (top), July (middle), and August (bottom). The solid

line is the 1:1 line and the dashed line is the linear regression line. The linear regression equation and the  $R^2$  values are shown at the top left in each panel

PRISM\_IDW, and PRISM\_MSK are shown in Table 1. Each number in the contingency table denotes the percentage of the number of stations where the outputs of each experiment belong to a (1) hit, (2) miss, (3) false alarm, and (4) rejection (Wilks 2011). The table consists of the precipitation or non-precipitation event counts that occurred in the 22 validation stations during the experimental period.

In comparison with the results of PRISM\_ORG and PRISM\_IDW, the results of PRISM\_IDW showed an increased percentage (number) of hits from 28% (1142) to 30% (1227) and a decreased percentage (number) of misses from 5% (209) to 3% (124). However, the percentage (number) of false alarms increased from 9% (371) to 11% (443) and the percentage (number) of correct rejections

**Table 1** Contingency table between the observed precipitation and estimated precipitation at each validation station. *Y* indicates that precipitation occurrence was observed or estimated, and *N* indicates that the precipitation occurrence was not observed or estimated

PRISM_ORG	Observation		PRISM_IDW	Observation		PRISM_MSK	Observation	
	<i>Y</i>	<i>N</i>		<i>Y</i>	<i>N</i>		<i>Y</i>	<i>N</i>
Estimation	<i>Y</i>	<i>N</i>	Estimation	<i>Y</i>	<i>N</i>	Estimation	<i>Y</i>	<i>N</i>
<i>Y</i>	28%	9%	<i>Y</i>	30%	11%	<i>Y</i>	29%	3%
<i>N</i>	5%	58%	<i>N</i>	3%	56%	<i>N</i>	5%	63%
Hit rate	86%		86%		92%			

decreased from 58% (2326) to 56% (2254). This is attributed to the expanded precipitation area resulting from the IDW used in PRISM\_IDW.

The results of PRISM\_MSK compared to PRISM\_IDW showed that the percentage (number) of false alarms decreased from 11% (443) to 3% (133) and the percentage (number) of correct rejections increased from 56% (2254) to 63% (2564), whereas the percentage (number) of hits decreased from 30% (1227) to 29% (1149) and the percentage (number) of misses increased from 3% (124) to 5% (202). Thus, PRISM\_MSK showed improved results compared to PRISM\_ORG. The percentage of hits and correct rejections increased, while the percentage of false alarms and misses decreased. In particular, the percentage of false alarms was reduced by more than half compared with the results of PRISM\_ORG. This means that the overestimated precipitation area was reduced due to the precipitation-masking algorithm applied in PRISM\_MSK.

## 4 Discussion and summary

In this study, experiments were conducted to improve the problems related to daily precipitation estimations using PRISM. The target area was South Korea, and the daily precipitation with a 1-km  $\times$  1-km resolution was estimated for 6 months from March to August 2012 using observational data obtained from 424 stations in South Korea. Two main problems arise when using PRISM\_ORG to estimate the daily precipitation. The first problem was an overestimation of the precipitation intensity in mountainous regions. This problem is caused by the altitude regression coefficient that is set too loosely in areas where the observation stations are distributed unevenly. This problem was solved by using the IDW method, which is an objective analysis method that does not place a weight on altitudes, to average the precipitation intensity with the regressed results of PRISM\_ORG. As a result, the overall RMSEs of the estimation are considerably reduced (Fig. 7), and the RMSEs caused by the overestimation at mountainous regions are significantly decreased (Fig. 5).

The second problem was the overestimation of local precipitation areas. With PRISM\_ORG, if local precipitation occurs in an area that is smaller than the radius of influence, the entire grid points within the radius of influence are estimated as precipitation areas. PRISM\_ORG has a problem that even areas with no observed precipitation are estimated as precipitation areas. This problem could be considerably mitigated by masking the overestimated precipitation grid points by applying the precipitation-masking algorithm proposed in our study, which generated more realistic precipitation areas estimations (Fig. 6). The experimental results showed that numerical improvements were also observed through statistical validation using the RMSE (Fig. 7), and the estimation errors occurring in many grid points in the scatter plots are reduced (Figs. 8 and 9). In addition, the numbers of hits and correct rejections were increased and the numbers of false alarms and misses were reduced, as shown in the  $2 \times 2$  contingency table (Table 1). However, this method may possibly regard the case of very local scale (meso-gamma scale 2–20 km) precipitation as no-precipitation. That is, if the distances among the stations where precipitation is observed are far apart and there is local-scale precipitation among the stations, some grid points located among the stations may be regarded as grids of no-precipitation. Therefore, this method can be more effectively utilized if there is a fine enough observation to detect even a meso-gamma-scale precipitation.

As mentioned in Sect. 3.1, the precipitation in the Korean Peninsula can be clearly divided into two types according to the seasons. That is, the precipitation in summer and that in the other three seasons are mainly caused by the local static instability and the baroclinic instability associated with the synoptic system, respectively. The main purpose of this study is to develop a method that can be more accurately applied to diverse cases throughout the whole year. Therefore, we determined that it would be more appropriate for the study purpose to conduct experiments over a long period of time from spring to summer in a particular year than to look at multiple years for a particular month. Accordingly, the developed method was verified through a total of 184 cases (March 1 to August

31). As discussed above, because the correction is more sensitive to the scale of precipitation and distances among observations than the year of precipitation, it is believed that the application of this method in other years will also produce similar results.

The daily precipitations in South Korea from only March to August 2012 were estimated and used in the experimental study. Although the level of improvement in precipitation estimation is not statistically significant for the entire period and verification points (Fig. 7), analysis of specific dates and points showed statistically significant improvement (data not shown). In addition, as for the spatial pattern and area of precipitation, the proposed method shows better performance in simulating observed precipitation, taking into account geographical features (Figs. 5d, 6d, and 8d). However, it might be necessary to produce longer-term data by expanding the experimental period and by performing verification tests for a more objective verification.

**Funding information** This work was funded by the Korea Meteorological Administration Research and Development Program under Grant KMI2018-01213 and “Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01345205)” Rural Development Administration, Republic of Korea.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Ahn JB, Lee JL (2016) A new multi-model ensemble method using nonlinear genetic algorithm: an application to boreal winter surface air temperature and precipitation prediction. *J Geophys Res* 121(16): 9263–9277
- Ahn JB, Lee JL, Im ES (2012) The reproducibility of surface air temperature over South Korea using dynamical downscaling and statistical correction. *J Meteor Soc Japan* 90(4):493–507
- Ahn JB, Hur J, Lim AY (2014) Estimation of fine-scale daily temperature with 30 m-resolution using PRISM. *Atmos Kor Meteor Soc* 14(1): 101–110
- Ahrens CD (2003) Meteorology today: an introduction to weather, climate and the environment, 7th edn. *Brooks-Cole-Thomson Learning, Pacific Grove*, p 554
- Barnes S (1964) A technique for maximizing details in numerical weather map analysis. *J Appl Meteorol* 9(3):396–409
- Brunetti M, Maugeri M, Nanni T, Simolo C, Spinoni J (2013) High-resolution temperature climatology for Italy: interpolation method intercomparison. *Int J Climatol* 34(4):1278–1296
- Chen H, Xu CY, Guo S (2012) Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *J Hydrol* 434:36–45
- Cressman GP (1959) An operational objective analysis system. *Mon Weather Rev* 87:367–374
- Daly C (2006) Guidelines for assessing the suitability of spatial climate data sets. *Int J Climatol* 26:707–721
- Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J Appl Meteorol* 33:140–158
- Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P (2002) A knowledge-based approach to the statistical mapping of climate. *Clim Res* 22(2):99–113
- Daly C, Helmer EH, Quiñones M (2003) Mapping the climate of PUERTO RICO, VIEQUES and CULEBRA. *Int J Climatol* 23: 1359–1138
- Gerelchuluun B, Ahn JB (2014) Air temperature distribution over Mongolia using dynamical downscaling and statistical correction. *Int J Climatol* 34(7):2464–2476
- Henry AJ (1919) Increase of precipitation with altitude. *Mon Weather Rev* 47:33–41
- Hong KO, Suh MS, Rha DK, Chang DH, Kim C, Kim MK (2007) Estimation of high resolution gridded temperature using GIS and PRISM. *Atmos, Kor Meteor Soc* 17(3):255–268
- Im ES, Ahn JB (2011) On the elevation dependency of present-day climate and future change over Korea from a high resolution regional climate simulation. *J Meteor Soc Japan* 89:89–100
- Jo S, Ahn JB (2014) Improvement of CGCM prediction for wet season precipitation over Maritime Continent using a bias correction method. *Int J Climatol* 35(13):3721–3732
- Johnson GL, Daly C, Taylor GH, Hanson CL (2000) Spatial variability and interpolation of stochastic weather simulation model parameters. *J Appl Meteorol* 39:778–796
- Kim MK, Han MS, Jang DH, Baek SG, Lee WS, Kim YH, Kim S (2012) Production technique of observation grid data of 1km resolution. *J Clim Res* 7(1):55–68
- Krige, D. G. A., 1951: Statistical approach to some mine valuation and allied problems on the Witwatersrand. *University of the Witwatersrand*, 272.
- Lee JL, Ahn JB (2018) A new statistical correction method using a self-organizing method to improve long-term dynamical prediction. *Int J Climatol*
- Perry LB, Konrad CE (2006) Relationships between NW flow snowfall and topography in the southern Appalachians. *USA Clim Res* 32: 35–47
- Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference. ACM*:517–524
- Shin SC, Kim MK, Suh MS, Rha DK, Jang DH, Kim CS, Lee WS, Kim YH (2008) Estimation of high resolution gridded precipitation using GIS and PRISM. *Atmos Kor Meteor Soc* 18(1):71–81
- Smith RB (1979) The influence of mountains on the atmosphere. *Adv in Geophys. Acad Press* 21:87–230
- Tachikawa T, Hato M, Kaku M, Iwasaki A (2011) Characteristics of ASTER GDEM version 2. *Proc. IEEE IGARSS 2011*:3657–3660
- Wilks, D. S., 2011: Statistical methods in the atmospheric sciences. 2nd, *Acad Press*, 306–307.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.