Statistical downscaling for daily precipitation in Korea using combined PRISM, RCM, and quantile mapping: Part 1, Methodology and evaluation in historical simulation

Maeng-Ki Kim[1], **Seonae Kim[1]**, Jinuk Kim[2], Jin Heo[3], Jeong-Soo Park[3], Won-Tae Kwon[4], and Myoung-Seok Suh[1]

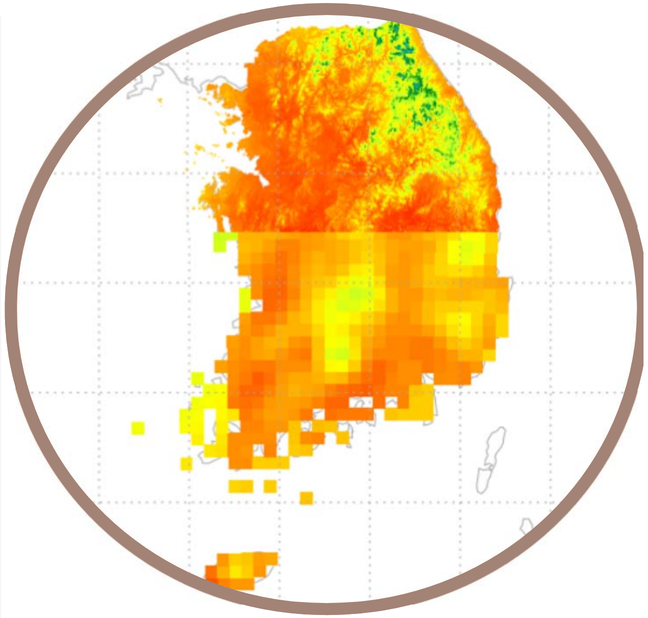
[1] Department of Atmospheric Science, Kongju National University, Gongju, Korea
[2] National Institute of Meteorological Sciences, Jeju, Korea
[3] Department of Statistics, Chonnam National University, Gwangju, Korea
[4] The Korean Society of Climate Change Research, Seoul, Korea`

Kim Maeng-Ki., S. Kim, J. Heo, and J-S. Park, 2016: Statistical downscaling for daily precipitation in Korea using combined PRISM, RCM, and quantile mapping: Part 1, Methodology and evaluation in historical simulation, *Asia-Pacific Journal of Atmospheric Sciences* 52(2),79-89.

INDEX

Global climate models (GCM) and regional climate models (RCM) provide climate change data at global and regional scales, respectively, for practical use, such as climate change adaptation policy-making.

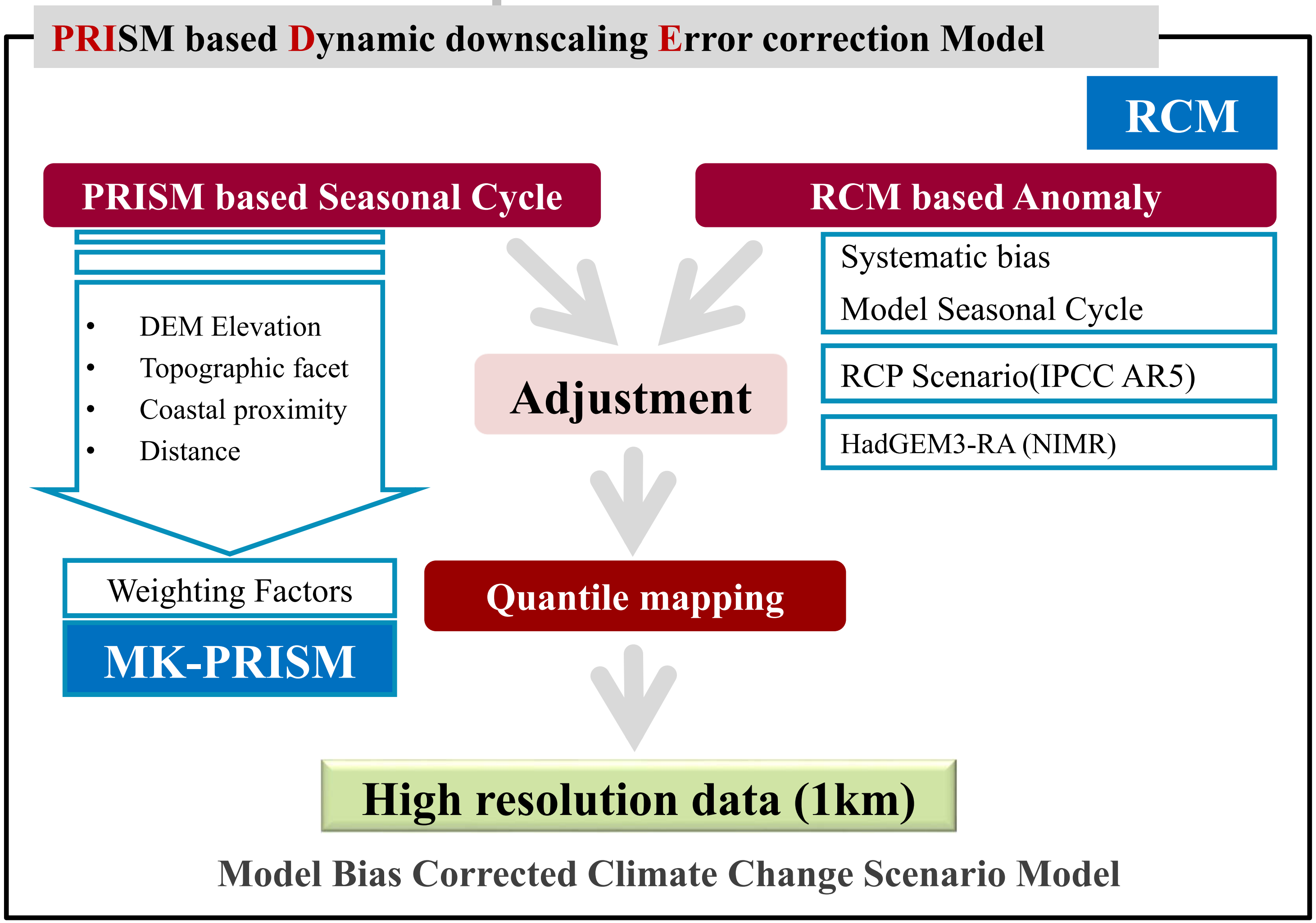
However, dynamic models struggle to provide climate change data with a horizontal resolution of less than 5 km, mainly due to systematic bias (Giorgi and Mearns, 1991; Chen et al., 1999; Hellström et al., 2001; Kim et al., 2004; Oh et al., 2004; Varis et al., 2004; Christensen et al., 2008; Hong and Kanamitsu, 2014) and the computational costs of long-term simulations of various scenario.



In this study, to downscale the climate variables of the RCM to a finer spatial resolution than the observational station scale, we developed the PRISM-based Dynamic downscaling Error correction (PRIDE) model, which is suitable for complex topographies.

Schematic diagram
PRIDE MODEL

Statistical downscaling model for High resolution Climate Change scenario data



DATA – Observed data

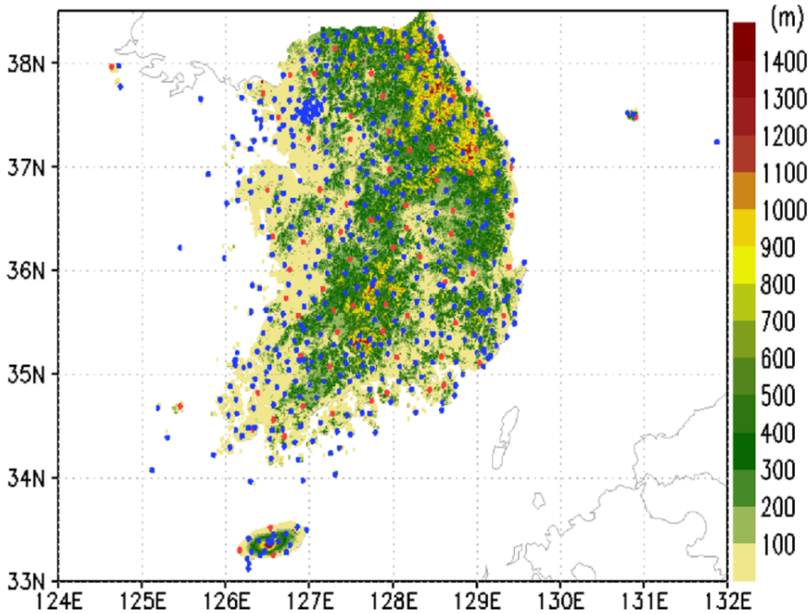


Figure 1 Location of observation stations (circle) and DEM elevation (shading) with 1-km resolution in Korea. Red and blue circles indicate ASOS and AWS, respectively.

- The total number of observation sites : 537 from the Korea Meteorological Agency (KMA)

  * Automated Synoptic Observation System (ASOS) : 75
  * Automated Weather System (AWS) : 462

- Variable : daily precipitation data

- Period : From 2000 to 2010 (11 years).

DATA – Regional Climate Model

HadGEM3-RA (derived by the boundary conditions of the historical run of the HadGEM2-AO(CMIP5), which is a contribution of the National Institute of meteorological Sciences, Korea, (NIMS))

NIMR is participating in the CORDEX with a regional climate model, HadGEM3-RA which is based on the global atmospheric HadGEM3 of the Met Office Hadley Centre (MOHC)

- Grid : the horizontal resolution is selected to 0.11°, about 12.5km
- Historical period : 2000 - 2010 (11years)
  2000 - 2005 (historical data from RCM)
  2006 – 2010 (scenario data of RCP 8.5 from RCM)

GIS Information

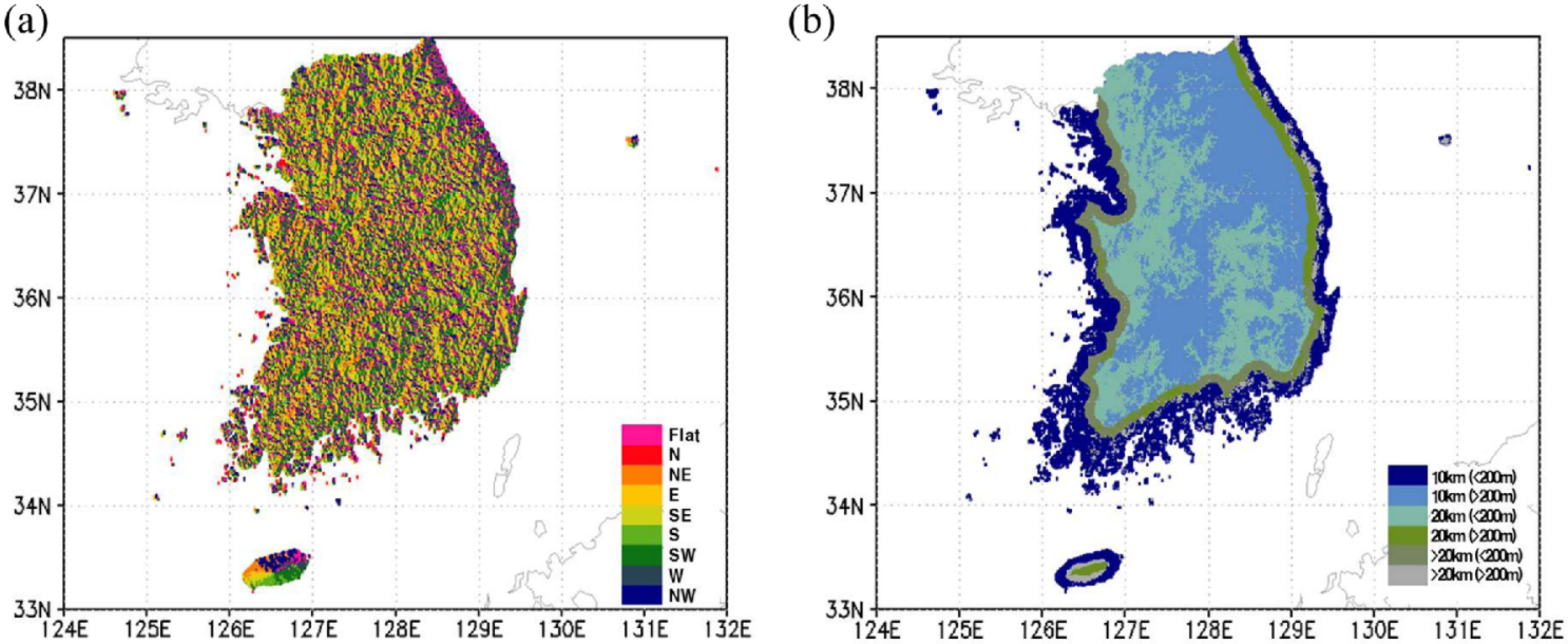


Fig. 2. (a) Topographic facet and (b) coastal proximity for a horizontal resolution of 1 km in Korea. The two values listed in the legend of (b) indicate the distance from the coastal line and altitude from sea level, respectively.

Topographic facets, DEM elevation, and coastal proximity are important parameters with significant influence on climate and weather and thus must be considered in the interpolation to high resolution.

The effect of these parameters is already included in the observation value, especially at monthly or seasonal timescales rather than daily timescales.

a. MK-PRISM module

$$\mathbf{Y} = \boldsymbol{\beta}_1 \mathbf{X} + \boldsymbol{\beta}_0 \qquad (1)$$

$$\mathbf{W} = [\, \mathbf{F}_d\, \mathbf{W}(d)^2 + \mathbf{F}_z\, \mathbf{W}(z)^2 \,]^{0.5}\ \mathbf{W}(f)\, \mathbf{W}(p)$$

[[Eq. (1)]]

Y is the estimated precipitation at the target grid cell.
X is the DEM height at the target grid cell.
$\boldsymbol{\beta}_1$ **and** $\boldsymbol{\beta}_0$ are the regression coefficients.

[[Eq. (2)]]

W is the total weighting function for the weighted linear regression.

W(d) : Weighting factor of distance
W(z) : Weighting factor of elevation
W(f) : Weighting factor of topographic facet
W(p) : Weighting factor of ocean proximity
F_d : the coefficient of the relative importance for the distance (0.8)
F_z : the coefficient of the relative importance for the elevation (0.2)
[we used 0.8(F_{d}) and 0.2(F_z)based on Daly (2006).]

The MK-PRISM module is Modified Korea- PRISM which was changed beta critical value and applied non-precipitation zone.
Please refer to Daly (2006), Daly et al. (2008), Kim et al. (2012) and Kim et al. (2013) for details. Below, a brief summary of the module is presented.

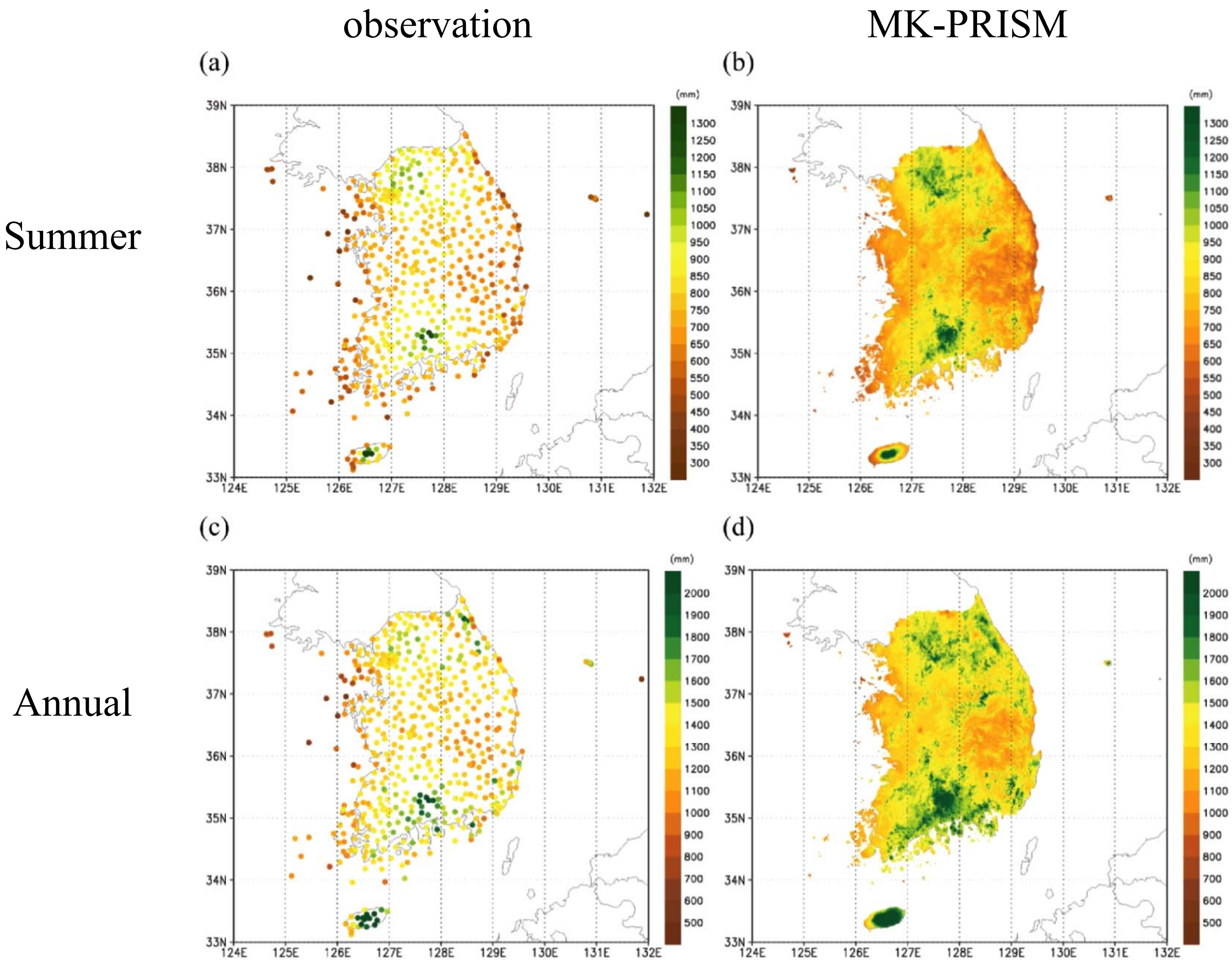


Fig. 4. Spatial distribution of precipitation obtained from observation (left panel) and MK-PRISM (right panel) for 11 years from 2000 to 2010. Upper and lower panels indicate summer and annual precipitation, respectively.

Observation

MK-PRISM

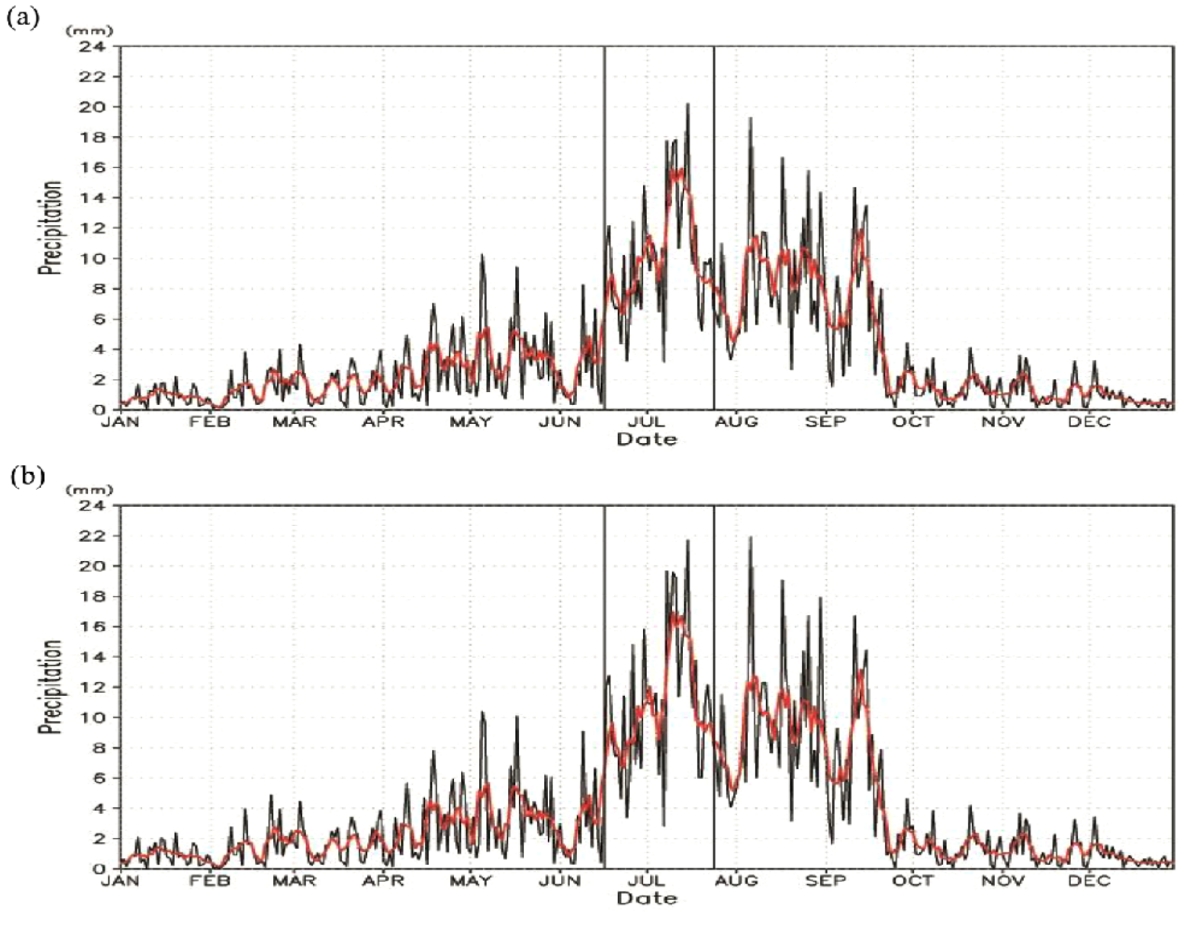


Fig. 5. Time series of 5-d moving average (red line) of daily precipitation (black line) obtained from (a) observation and (b) MKPRISM for 11 years from 2000 to 2010. The two vertical lines represent the starting date (19 June) and ending date (24 July), respectively, of the Changma period defined by KMA (2011).

- The mean bias for daily precipitation showed the range from minimum value of 0.14 mm/day in November to maximum value of 0.67mm/day in August.
- The RMSE showed the range from 0.64mm/day in December to 3.68mm/day in August.

b. RCM and systematic bias

Although the RCM is the best dynamic downscaling tool for climate change data at a regional scale, it faces some limitations when simulating climatology on a daily timescale and climate variability, indicating that systematic bias cannot be avoidable in the current state of the art.

In pure scientific studies, anomalies may be more important than climatology because the response induced by forcing(e.g., greenhouse gas and aerosol forcing) can be evaluated by climate anomalies.

[RCM minus MK-PRISM]

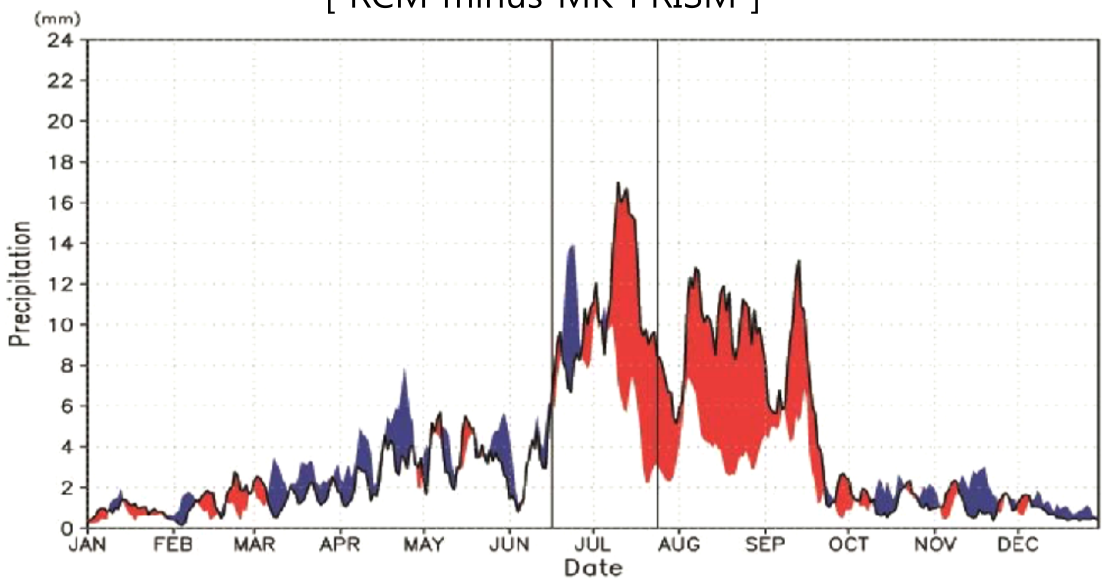
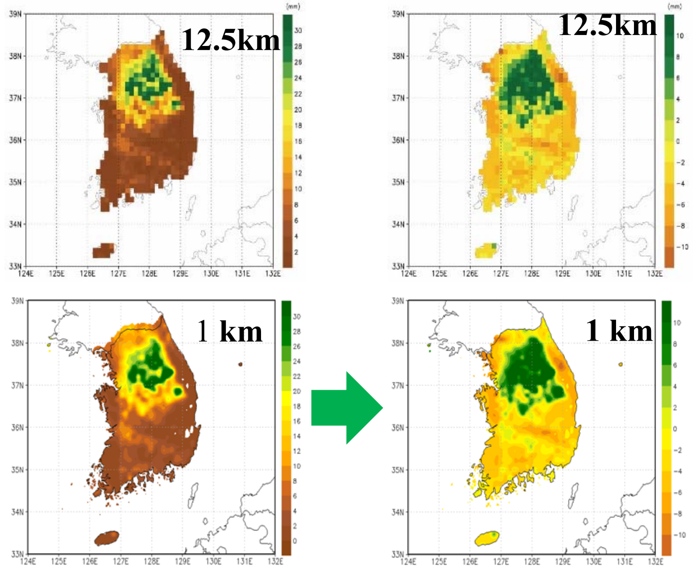


Fig. 6. Time series of 5-d moving average precipitation anomaly (shading) for RCM compared to the climatological annual cycle (solid line). Solid lines indicate the precipitation time series estimated by MK-PRISM. Blue and red shading areas represent overestimation and underestimation, respectively, compared to observation. The two vertical lines represent the starting date (19 June) and ending date (24 July), respectively, of the Changma period defined by KMA.

The seasonal cycle itself is one of the important components of systematic bias in the RCM.

We assumed that the daily precipitation (P^d) about RCM and observation consists of a daily seasonal cycle component and an anomaly component.

Barnes interpolation Method



RCM Daily PRCP (July 10, 2000)

RCM Anomaly (July 10, 2000)

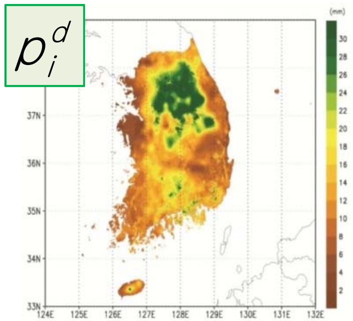
ex) Precipitation on July 10, 2000

m : model , o : observation
d : specific day , s : seasonal cycle , a : anomaly

$$p_m^d = p_m^s + p_m^a \tag{10}$$

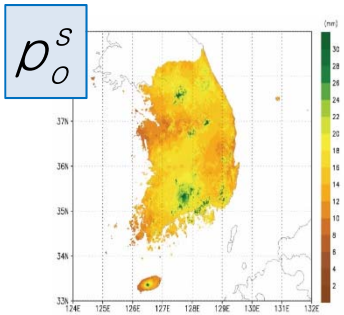
$$p_o^d = p_o^s + p_o^a \tag{11}$$

$$p_i^d = p_o^s + p_m^a \tag{12}$$



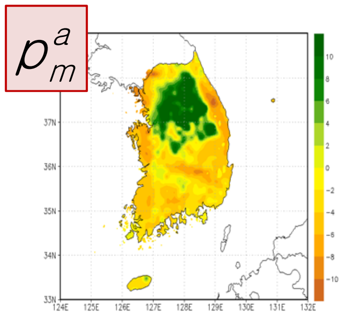
PRIDv2.0 Daily PRCP (7.10.2000)

=



MK-PRISM Climatology (July 10)

+



RCM Anomaly (July 10, 2000)

The RCM significantly underestimates the no precipitation class compared to observation and MK-PRISM but significantly overestimates light precipitation (0 < P ≤10 mm/day).

The PRIDE v2.0 partially removes the bias in the two peaks and one lull, which indicates the improvement of climate change data comparing with RCM.

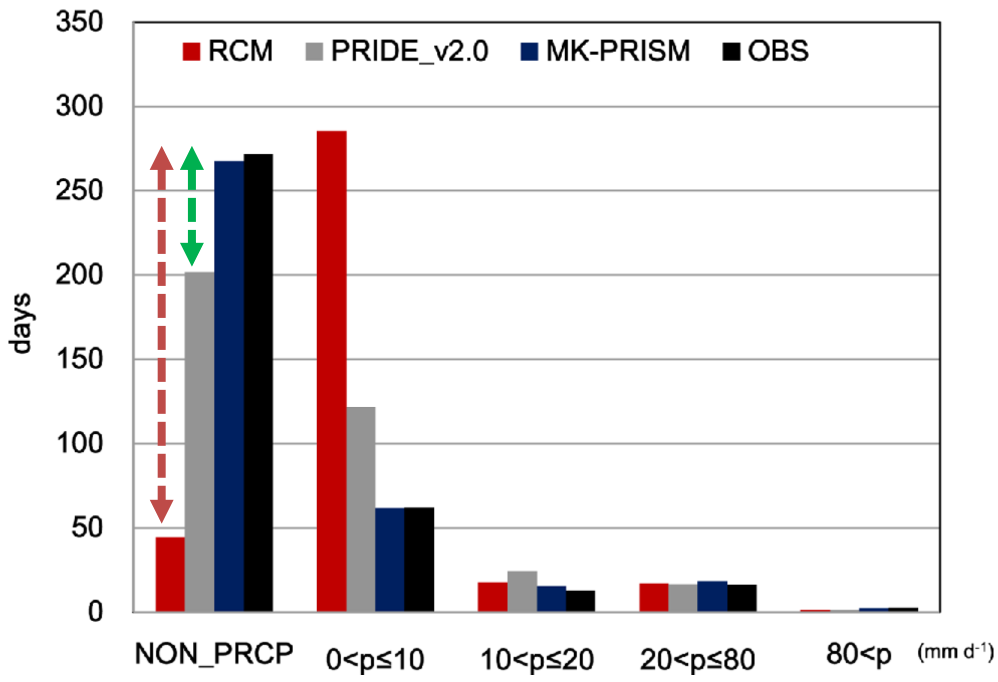
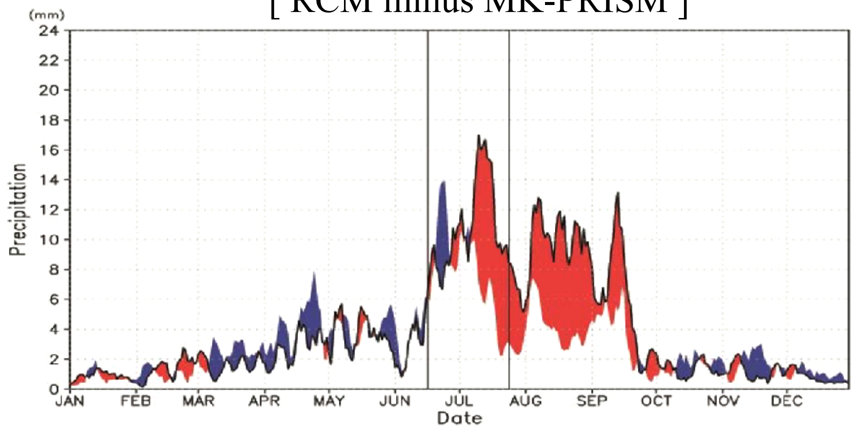
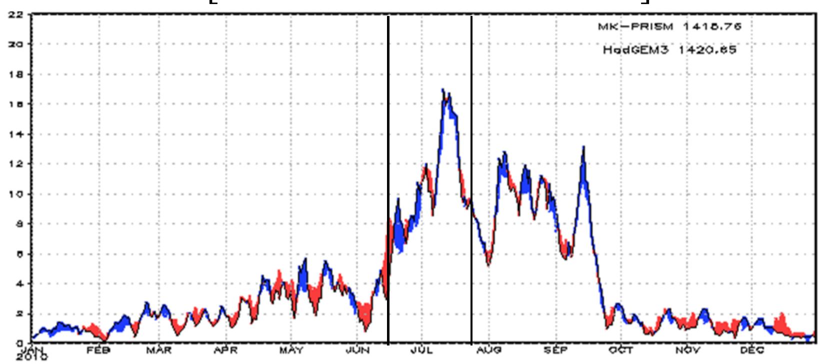


Fig. 7. Histogram of daily precipitation (P) for five classes (P = 0, 0 < P ≤ 10 mm d−1, 10 < P ≤ 80 mm d−1, 80 mm d−1 < P). Red, gray, blue, and black bar indicate RCM, PRIDE v2.0, MK-PRISM, and observation (OBS), respectively, for 11 years from 2000 to 2010.

[RCM minus MK-PRISM]



[PRIDE minus MK-PRISM]



c. Quantile mapping(QM) module and verification

The QM is a popular method for correcting the difference in climate variables between the observation and model using a Probability Density Function (PDF) or Cumulative Density Function (CDF) framework (Piani et al., 2010a; Gudmundsson et al., 2012; Teutschbein and Seibert, 2012).

We have used CDF fitting as the final step to correct systematic bias.

In QM, the first step is to find a transfer function (TF) between observation data and intermediate data for the following training period (Piani et al., 2010a, 2010b) based on the gamma distribution for both the observed and modeled precipitation intensity.

Six Transfer functions (TF)

- **Linear TF**

$$TF1: P_o = b \times P^c{}_m$$

$$TF2: P_o = a + b \times P_m$$

- **Multiple TF**

$$TF3: P_o = (a + b \times P_m) \times (1 - e^{(-\frac{P_m}{\tau})})$$

$$TF4: P_o = b \times P_m$$

- **Exponential TF**

$$TF5: P_o = b \times (P_m - x0)^c$$

$$TF6: P_o = (a + b \times P_m) \times (1 - e^{(-\frac{P_m - x0}{\tau})})$$

Piani et al. (2010b).

$$CDF_{obs}(f(P_m)) = CDF_{model}(P_m) \quad (13)$$

$$P_0 = f(P_m) \quad (14)$$

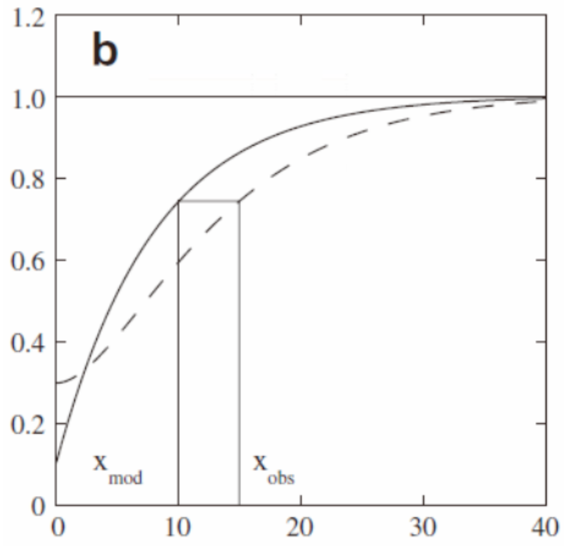
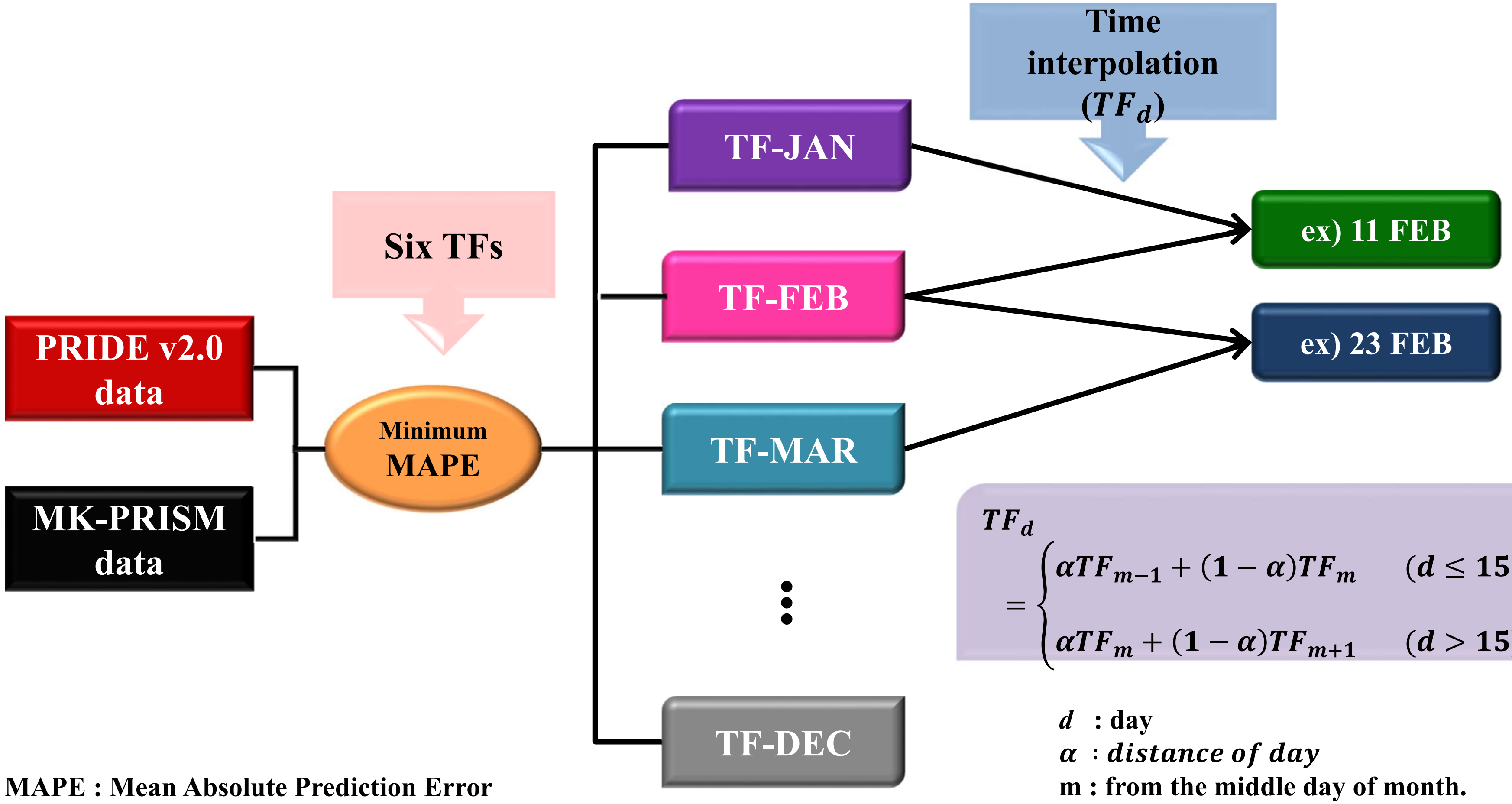


Fig 1(b). Piani et al. (2010b).

The transfer functions are estimated at each grid point and each month.

To avoid discontinuity between two neighboring months, the daily TF is determined by the linear weighting of two TFs based on Piani et al. (2010b) as follows.



$$TF_d = \begin{cases} \alpha TF_{m-1} + (1-\alpha)TF_m & (d \le 15) \\ \alpha TF_m + (1-\alpha)TF_{m+1} & (d > 15) \end{cases}$$

d : day
α : *distance of day*
m : from the middle day of month.

MAPE : Mean Absolute Prediction Error

c. Quantile mapping module and verification

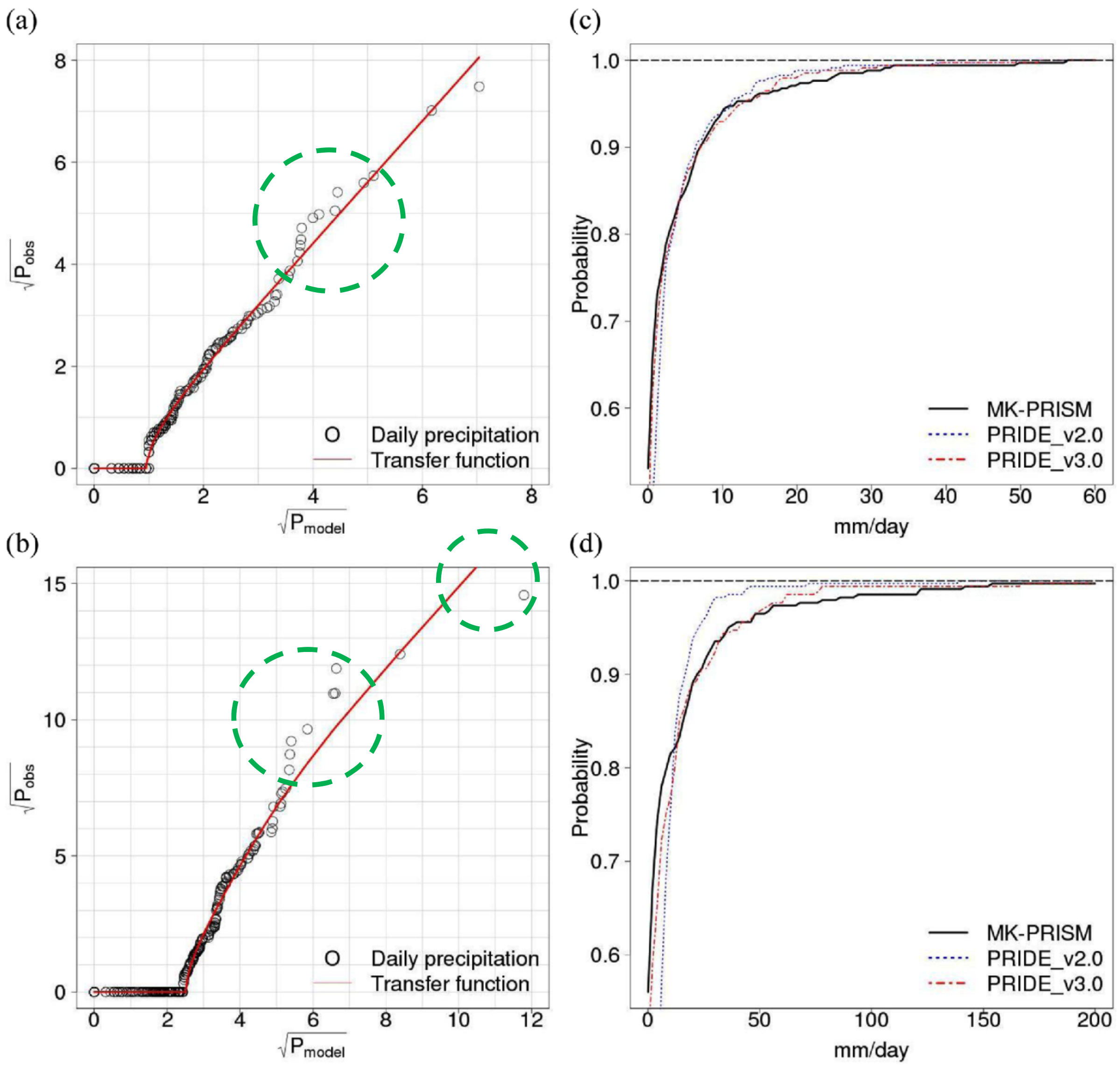


Fig. 8. Scatter-diagram of the square root of precipitation (P) from observation (y axis) and PRIDE v2.0 model (x axis) at the sample target grid during (a) January and (b) July for 11 years from 2000 to 2010. ECDF for daily precipitation (mm d−1) during (c) January and (d) July.

The six transfer functions were independently tested for each grid point to obtain the best transfer function for a given grid point.

We have obtained the spatial distributions from the best transfer function selected for each grid in Jeju (Fig. 9) and the percentage of the transfer function selected (Table 1).

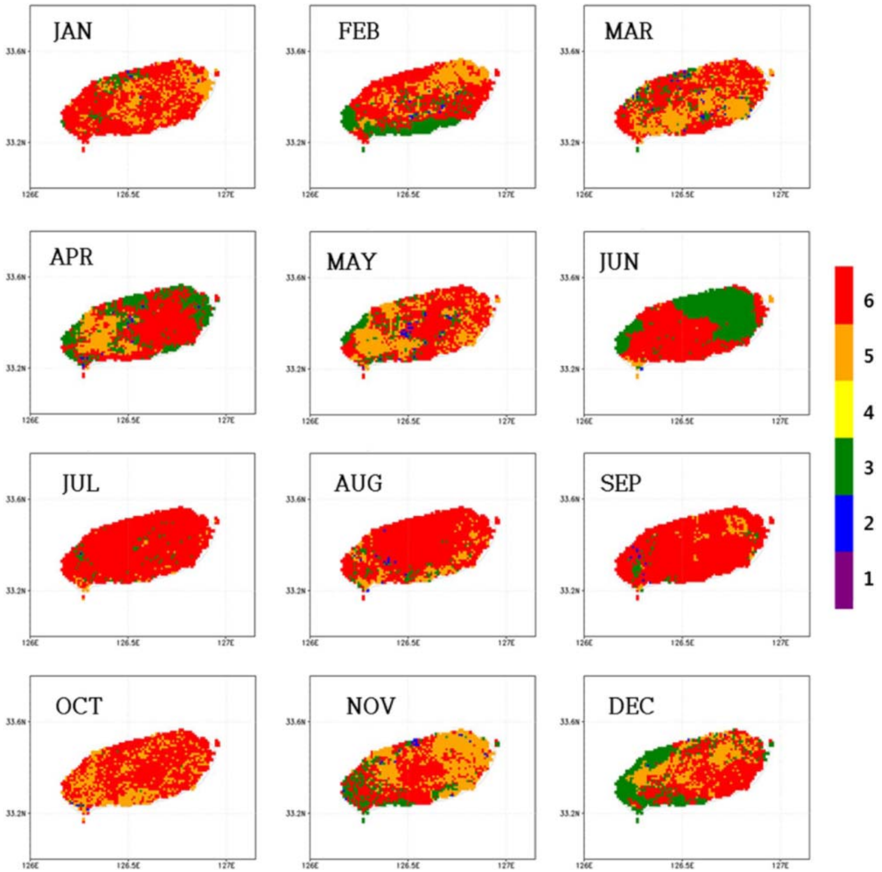


Fig. 9. Spatial distribution of the selected transfer function in Jeju. Numerical values represent the best transfer function for each grid point. See Table 1 for the transfer function.

Table 1 Six transfer function types used in this study and the percentage of the transfer function selected for Jeju for the period 2000-2010. P_o and P_m represent the observation and model, respectively. a, b, c, x0, and τ are the free parameters that are subject to calibration.

Transfer function	Percentage (%)
$TF1$: $P_o = b \times P^c{}_m$	2.0
$TF2$: $P_o = a + b \times P_m$	0.7
$TF3$: $P_o = (a + b \times P_m) \times (1 - e^{(-\frac{P_m}{\tau})})$	13.6
$TF4$: $P_o = b \times P_m$	0.0
$TF5$: $P_o = b \times (P_m - x0)^c$	21.4
$TF6$: $P_o = (a + b \times P_m) \times (1 - e^{(-\frac{P_m - x0}{\tau})})$	62.3

Result of K-fold cross-validation

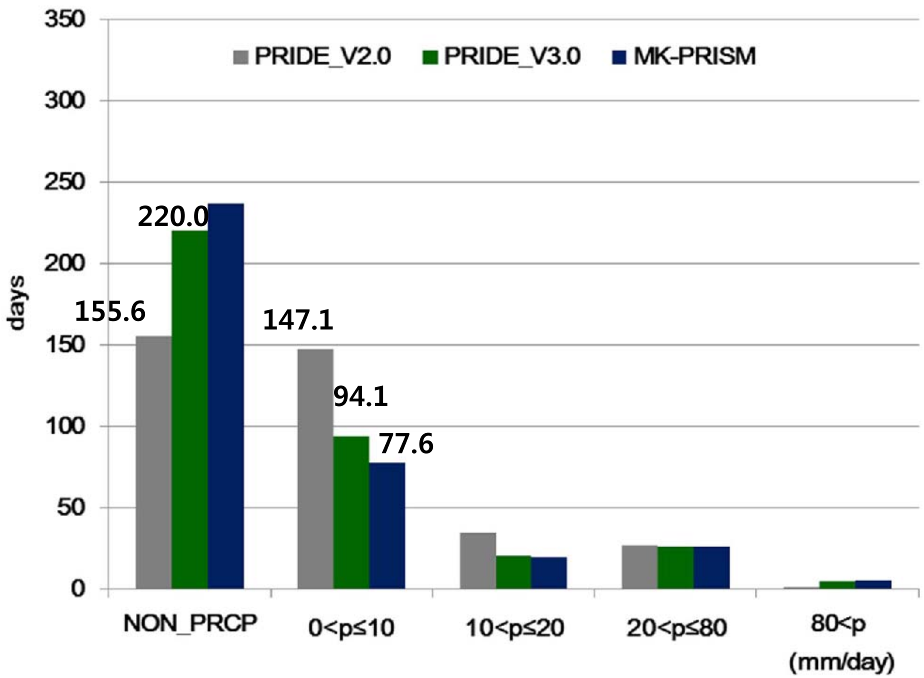


Fig. 10. Histogram verification of daily precipitation (P) for five classes (P = 0, 0 < P ≤ 10 mm/d, 10 < P ≤ 80 mm/d, 80 mm /d < P) in Jeju for 11 years form 2000 to 2010. Gray, green and blue bars indicate PRIDE v2.0, v3.0 and MK-PRISM, respectively.`

Table 2 Relative mean absolute error (MAE) of daily precipitation for PRIDE v2.0 and 3.0 in Jeju. The relative MAE is calculated by $\frac{1}{k}\sum_{i=1}^{k}\left(\frac{|yi-\hat{y}i|}{yi}\right)$ where yi and $\boldsymbol{yi}$ represent the observed and estimated values, respectively.

Month	Relative MAE	
	PRIDE v2.0	PRIDE v3.0
JAN	0.838	**0.620**
FEB	1.023	1.113
MAR	1.891	**1.183**
APR	1.555	1.570
MAY	1.944	**1.640**
JUN	1.158	1.186
JUL	3.479	**2.842**
AUG	3.115	**1.008**
SEP	5.500	**2.703**
OCT	2.090	**1.211**
NOV	1.121	**0.804**
DEC	1.097	1.365
AVERAGE	**2.068**	**1.437**

04 Summary

In this study, we present the Parameter-elevation Relationships on Independent Slopes Model (PRISM)-based Dynamic downscaling Error correction (PRIDE) model, which is suitable for complex topographies, such as the Korean peninsula.

The PRIDE model is constructed by combining the PRISM module, the Regional Climate Model (RCM) anomaly, and quantile mapping (QM) to produce high-resolution (1 km) grid data at a daily time scale.

The results show that the systematic bias of the RCM was significantly reduced by simply substituting the climatological observational seasonal cycle at a daily timescale for each grid point obtained from the PRISM.

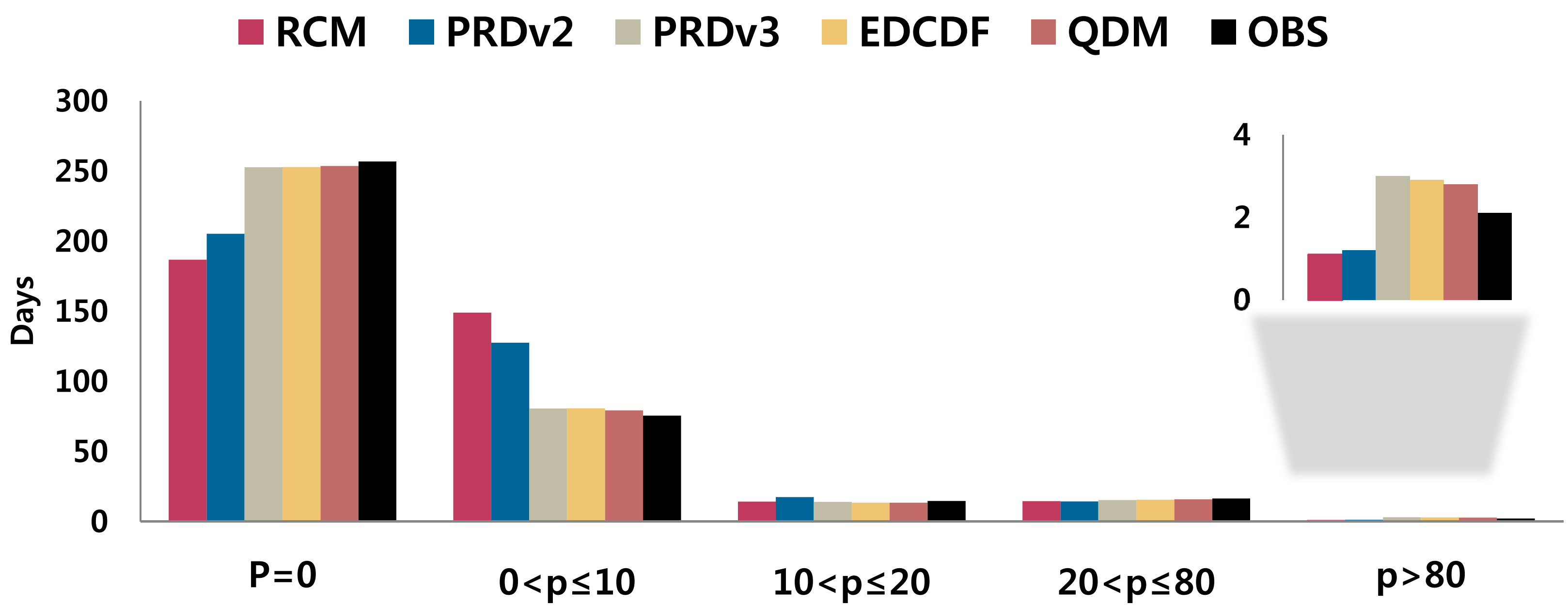
QM was then applied to correct additional systematic bias by constructing the transfer functions under the cumulative density function framework between the model and observation using six types of transfer functions.

K-fold cross-validation of the PRIDE model shows that the number of precipitation days modeled is approximately 90 ~ 121% of the number observed for the five daily precipitation classes, indicating that the PRIDE model reasonably estimates the observational frequency of daily precipitation under a quantile framework.

The relative Mean Absolute Error (MAE) is also discussed in the framework of the intensity of daily precipitation.

Future ..

- We are applying by Using Multi RCM (Ensemble) and are producing Future period scenario.
- We are trying to test other QM method (EDCDF, QDM ,DQM..) for comparing with PRIDE Model.



Baek, H.-J., J. Lee, H.-S. Lee, Y.-K. Hyun, C. Cho, W.-T. Kwon, C. Marzin, S.-Y. Gan, M.-J. Kim, D.-H. Choi, J. Lee, J. Lee, K.-O. Boo, H.-S. Kang, and Y.-H. Byun, 2013: Climate change in the 21st century simulated by HadGEM2-AO under representative concentration pathways. *Asia-Pacific J. Atmos. Sci.*, **49,** 603-618. doi:10.1007/s13143-013-0053-7.

Chen, S., J. O. Roads, H. H. Juang, and M. Kanamitsu, 1999: Global to regional simulations of California wintertime precipitation. *J. Geophys. Res.*, **104**, 31517-31532.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys. Res. Lett.*, **35**, L20709.

Daly, C., W. Gibson, G. Taylor, G. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Clim. Res.*, **22**, 99-113.

Daly, C., E. H. Helmer, and M. Quinones, 2003: Mapping the climate of Puerto Rico, Vieques and Culebra. *Int. J. Climatol.*, **23**, 1359-1381.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031-2064.

Giorgi, F., and L. O. Mearns, 1991: Approaches to the simulation of regional climate change: a review. *Rev. Geophys.*, **29**, 191-216.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical Note: downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390. doi:10.5194/hess-16-3383-2012.

Hellström, C., D. Chen, C. Achberger, and J. Räisänen, 2001: Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. Clim. Res., 19, 45-55.

Hong, K.-O., M. S. Suh, D.-K. Rha, D.-H. Chang, C. Kim, and M.-K. Kim, 2007: Estimation of high resolution gridded temperature using GIS and PRISM. Atmosphere, 17, 255-268.(In Korean)

Hong, S.-Y. and M. Kanamitsu, 2014: Dynamical downscaling: fundamental issues from an NWP point of view and recommendations. Asia-Pac J. Atmos Sci, 50(1), 83-104. doi:10.1007/s13143-014-0029-2.

Kim, M.-K., M.-S. Han, D.-H. Jang, S.-G. Baek, W.-S. Lee, Y.-H. Kim, and S. Kim, 2012: Production technique of observation Grid data of 1km resolution. J. Clim. Research, 7(1), 55-68.(In Korean)

Kim, M.-K., D.-H. Lee, and J. Kim, 2013: Production and validation of daily grid data with 1km resolution in South Korea.. J. Clim. Research, 8(1), 13-25.(In Korean)

Oh, J.-H., T. Kim, M.-K. Kim, S.-H. Lee, S.-K. Min, and W.-T. Kwon, 2004: Regional climate simulation for Korea using dynamic downscaling and statistical adjustment. JMSJ, 82, 1629-1643. doi:10.2151/jmsj.82.1629.

Piani, C., J. O. Haerter, and E. Coppola, 2010a: Statistical bias correction for daily precipitation in regional climate models over Europe. Theor. Appl. Climatol., 99, 187–192. doi: 10.1007/s00704-009-0134-9.

Piani, C., G. P. Weedon, M. Best, S. M. Gomes, P. Viterbo, S. Hagemann, and J. O. Haerter, 2010b: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. J. Hydrol., 395, 199–215. doi:10.1016/j.jhydro.2010.10.024.

Shin, S.-C., M.-K. Kim, M.-S. Suh, D.-H. Jang, C.-S. Kim, W.-S. Lee, and Y.-H. Kim, 2008: Estimation of high resolution gridded precipitation using GIS and PRISM. Atmosphere, 18, 71-81.(In Korean)

Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. J. Hydrol., 456–457, 12-29.

Varis, O., T. Kajander, and R. Lemmelä, 2004: Climate and water: from climate models to water resources management and vice versa. Climatic Change, 66, 321-344.

THE END

THANK YOU

Distance weighting function in the MK-PRISM is defined as follows.

$$W(d) = 1 \ \text{for d} \le 1\text{km} \qquad (3)$$
$$= \frac{1}{\mathrm{d}^{\mathrm{a}}} \ \text{for d} > 1km$$

Where d means the distance between grid point and observation distance weighting exponent *a* of 2 is used.

Elevation weighting are different for three categories as follows.

$$W(z) = 1 \ \text{for } \Delta z \le \Delta z_m \qquad (4)$$
$$= \frac{1}{(\Delta z - \Delta z_m)^b} \ \text{for } \Delta z_m < \Delta z \le \Delta z_x$$
$$= 0 \ \text{for } \Delta z_m \ge \Delta z_x$$

Here Δz represents the difference in distance between target grid cell and station point.

Δz_m and Δz_x represent the critical value of Δz for elevation weight of 1 and 0, respectively.

We have used 100m for Δz_m and 1000m for Δz_x.

Elevation weighting exponent *b* of 1 is used in this study.

Facet weighting was defined for two categories in equation (5).

$$W(f) = 1 \text{ for } \Delta f \le 1 \text{ and } B = 0 \qquad (5)$$
$$= \frac{1}{(\Delta f + B)^{c}} \text{ for } \Delta f > 1 \; or \; B > 0$$

Where Δf indicates the difference in topographic facet between target grid cell and station point and B represents the number of obstacles (between target grid cell and observation point where have different topographic facet.

Coastal proximity weighting is defined in equation (6).

$$W(p) = 1 \text{ for } \Delta p = 0 \qquad (6)$$
$$= 0 \text{ for } \Delta p > p_x$$
$$= \frac{1}{\Delta p^{v}} \text{ for } 0 < \Delta p \le p_x$$

Where Δp represents the difference in ocean proximity between target grid cell and observation point and p_x represents critical value of Δp for coastal proximity weighting of 0.

The coastal proximity exponent (v) of 1.0 is used from sensitivity study in this study.

$$\mathbf{Y = \beta_1 X + \beta_0} \qquad \mathbf{(1)}$$

$$\mathbf{W = [\, F_d\, W(d)^2 + F_z\, W(z)^2\,]^{0.5}\ W(f)\ W(p)}$$

(2)

The regression coefficients are estimated by equation (8) and (9) as follows.

$$\bar{x} = \frac{\sum W_i X_i}{\sum W_i}, \quad \bar{y} = \frac{\sum W_i Y_i}{\sum W_i} \qquad (7)$$

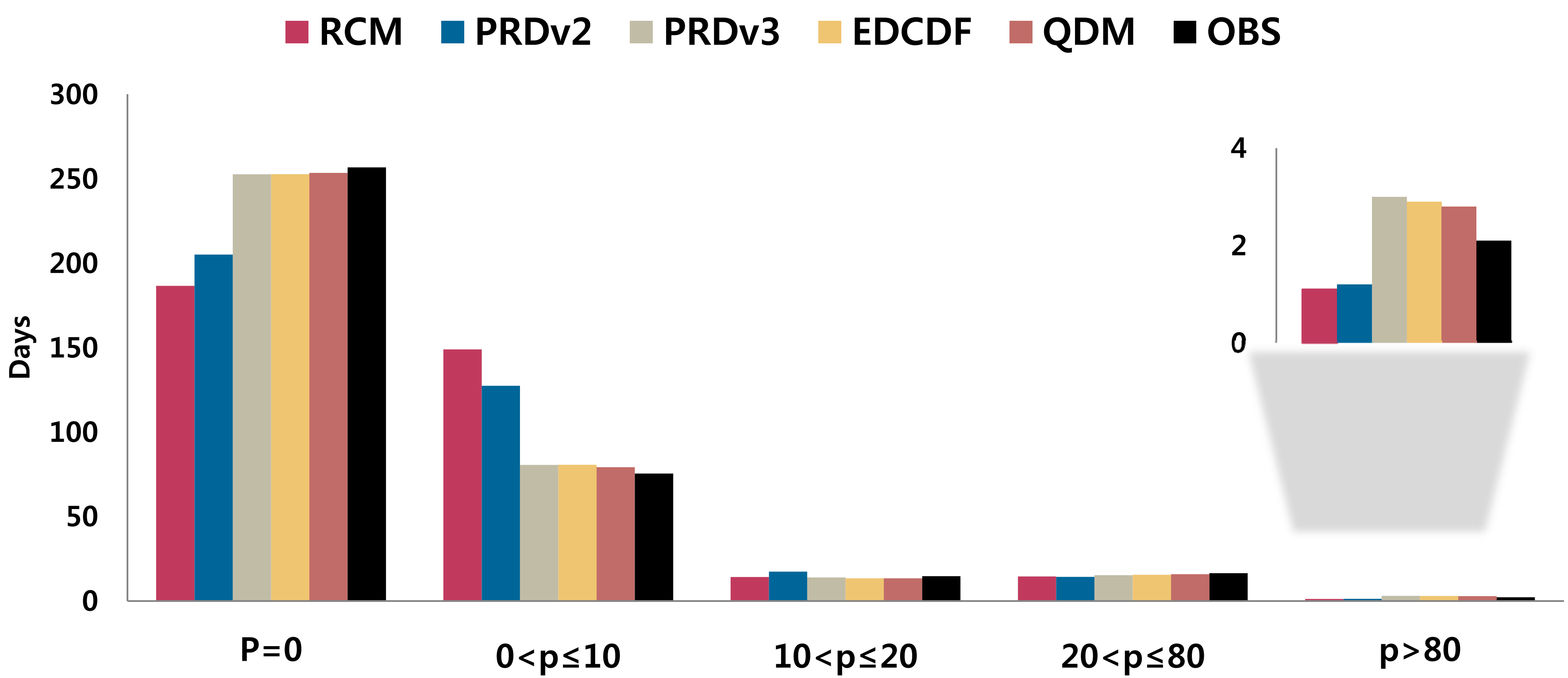
$$\hat{\beta}_1 = \frac{\sum W_i (X_i - \bar{x})(Y_i - \bar{y})}{\sum W_i (X_i - \bar{x})^2} \qquad (8)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (9)$$

$\bar{x}$ and $\bar{y}$ represent the means of variables X and Y, respectively, at the target grid in equation (1).

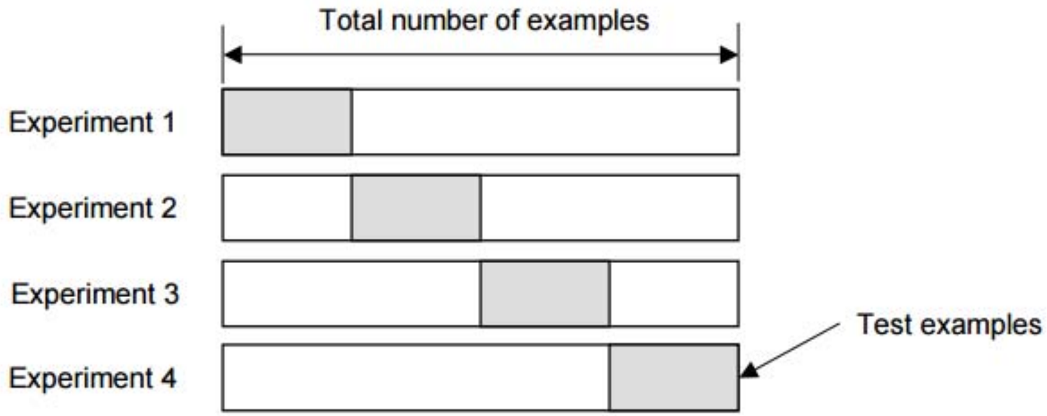
The subscript i indicates a specific station inside the search range around the target grid.

c. Comparing Quantile mapping methods



K-Fold Cross-validation

- **Create a K-fold partition of the the dataset**
  - For each of K experiments, use K-1 folds for training and the remaining one for testing



- **K-Fold Cross validation is similar to Random Subsampling**
  - The advantage of K-Fold Cross validation is that all the examples in the dataset are eventually used for both training and testing
- **As before, the true error is estimated as the average error rate**

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i$$

REF. Intelligent Sensor Systems
Ricardo Gutierrez-Osuna
Wright State University